



# 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.

© 2014 Elsevier Ltd. All rights reserved.

## 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 “millimeter” and “good rice” of a female native

\* Corresponding author. Tel.: +31 30 253 5853.

E-mail address: [irischen71@hotmail.com](mailto:irischen71@hotmail.com) (A. Chen).

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

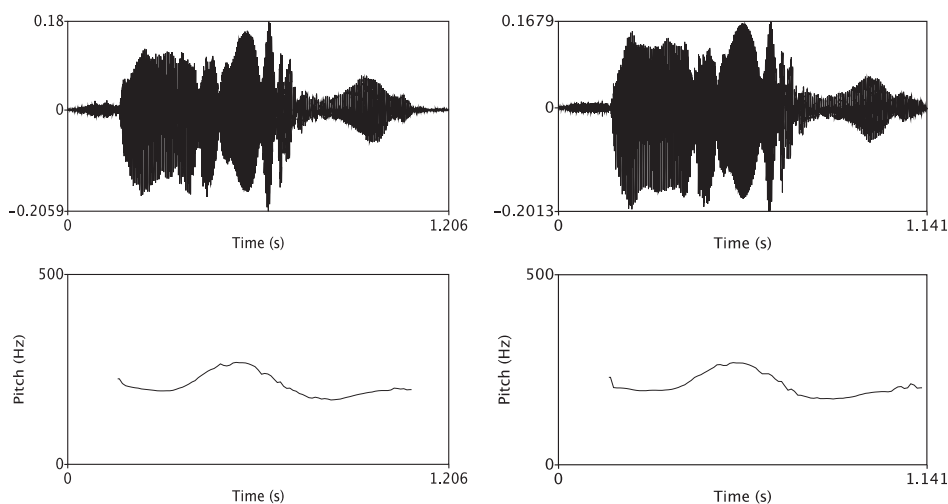


Fig. 1. Natural production of “milimeter” (left panel) and “good rice” (right panel) of a female speaker.

vowel space in such a way that maximal contrastiveness is maintained, whereas contrasts with low perceptibility tend to be either repaired for stronger contrastiveness, or to be merged (Liljencrants and Lindblom, 1972; Lindblom and Engstrand, 1989; Lindblom, 1990; Hume and Johnson, 2001). Referring to the asymmetry, it has been found that [ki] is often mistaken for [ti], but not the other way around. Chang et al. (2001) demonstrated that such confusion asymmetry is specific to a high, front vocalic environment. They suggested that it was the relatively tight constrictions of these vowels that resulted in a slightly longer lag in voice onset timing (VOT) or slight frication, which led the listeners to perceive a change in manner and place of articulation. In one of the few previous studies on perceptual effects of tonal phonology, Hombert et al. (1979) argued that lexical tones derive from a voicing contrast of prevocalic consonants and postvocalic glottal consonants, as the pitch perturbation introduced by the consonantal differences was perceptible for the listeners. These studies speak for a functional basis of the phonological grammar, in that the sounds in languages are organized in a way that suits the human cognitive system. Taking this account, it is plausible that the specific asymmetry observed in T3 sandhi may also reflect listeners' preference.

If we are to understand the possible cognitive constraints that may play a role in the perception of tone sandhi, it is important to disentangle such constraints from perceptual patterns acquired through experience. The performance of native listeners may solely reflect the acquired phonological knowledge. Yet it may also reflect general cognitive constraints that are not specific to T3 sandhi, or a combination of both. In other words, it is not enough to only test speakers whose native language has right-dominant sandhies, such as Mandarin, but also naïve listeners who have no knowledge of such sandhi patterns.

In the current study, we first tested the aforementioned two hypotheses regarding the perceptual knowledge of native listeners. Next, we asked whether the asymmetry in T3 sandhi reflects certain cognitive constraints among the listeners. Specifically, we tested both native Mandarin listeners, who have acquired T3 sandhi, and non-tone language (Dutch) listeners, who are naïve to lexical tones as well as tone sandhi. These two questions are addressed by bisyllabic tone discrimination experiments. In the experiment, we take the T3T3, where both T3s carry the underlying dipping contour as the referent, and the participants faced the question of discriminating T2T3 and T3T2 from it. Corresponding to the two hypotheses, if the native listeners have learned that T3s do not co-occur within a word, then they should perceive a word boundary between the two T3s. If this is the case, we would expect native listeners to discriminate T2T3 and T3T2 equally well from the referent, as they are discriminating between legal and illegal tonal sequences. Alternatively, if native listeners have learned to neutralize between T2T3 and T3T3, then we would expect the discrimination between T2T3 and T3T3 to be of low accuracy. The discrimination between T2T3 and T3T3 should be less accurate compared to that between T3T3 and T3T2. To ascertain whether the perceptual pattern of T3 sandhi is the result of the knowledge of phonological grammar, we also include bisyllabic sequences involving T1 and T4 as a baseline in the same experiment, as the T1–T4 contrast does not require tone sandhi in Mandarin. By including Dutch listeners whose performance cannot be attributed to the learning of T3 sandhi, we aim to find out the possible functionally based cognitive constraints, if any, relating to the asymmetry in T3 sandhi. If the performance of Mandarin listeners is solely determined by their acquired phonological knowledge, then Dutch listeners are not expected to show such knowledge. On the other hand, if T3 sandhi relates to general cognitive constraints, then we expect such constraints to influence Dutch and Chinese listeners in the same way, and the two groups should show some similarities in their discrimination patterns.

## 2. Experiments

### 2.1. Participants

Forty-eight Mandarin listeners and 48 Dutch listeners participated in this task. For the Dutch listeners (aged 19–46 years), at the time of the experiment, three participants were working, and they had finished at least 15 years of education. The remainder of the participants were Utrecht University students. An additional three Dutch participants were tested but excluded for the following reasons: one for not finishing both tasks, two for reporting not to have understood the instructions after being tested. All the Mandarin participants were born and raised in China, in a Mandarin-speaking environment. At the time of the experiment, all the participants were either enrolled in a bachelor or a post graduate program at Utrecht University (age range 21–35 years). Three additional Mandarin participants were tested, but excluded from analysis after reporting to have misunderstood the instructions after the experiment was over. None of the participants reported hearing or speech problems.

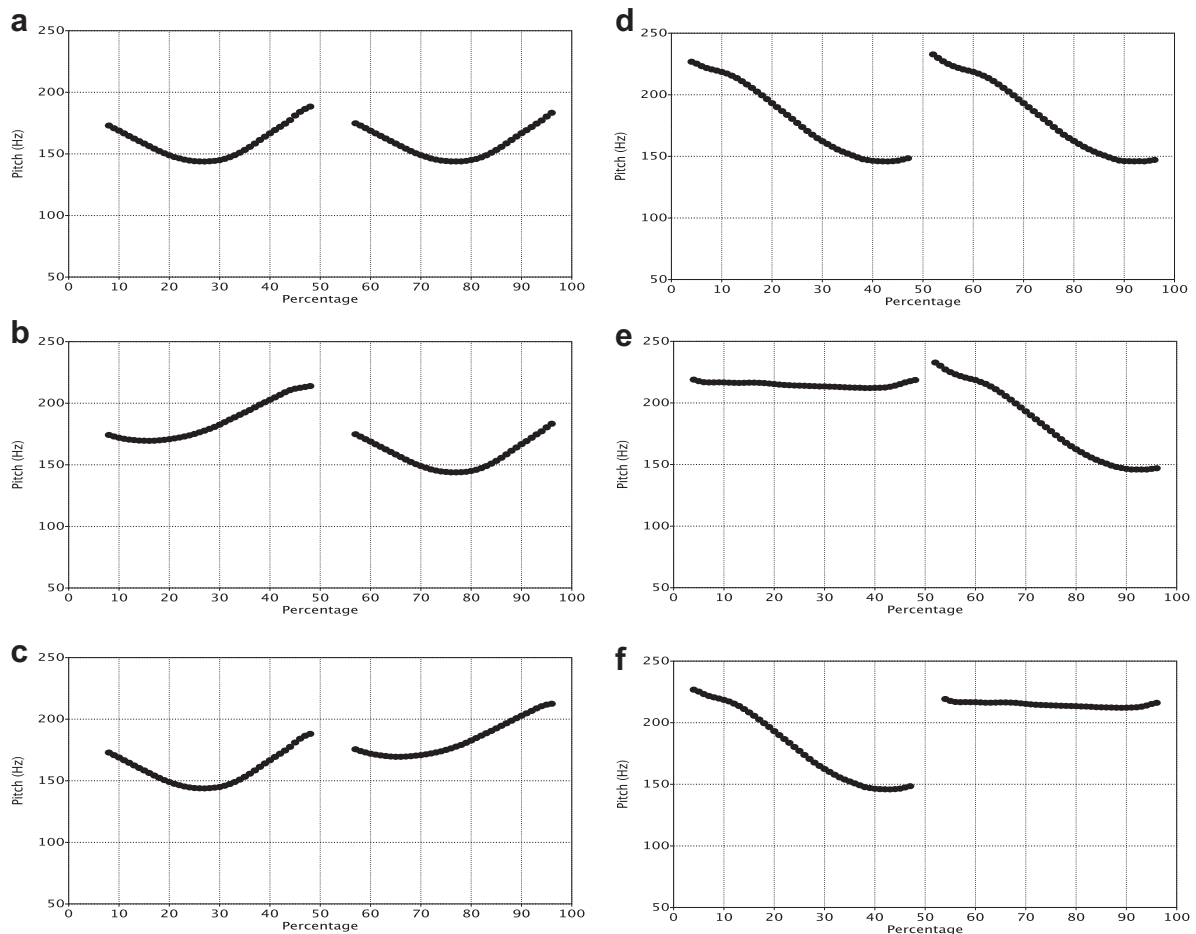
### 2.2. Stimuli

Monosyllables and bisyllabic sequences were generated for a Mandarin lexical tone discrimination task. A female native speaker of Mandarin recorded multiple tokens of monosyllables /ba/, /bou/, /bi/, /da/, /dou/, /di/, /la/, /lou/, /li/, /ma/, /mou/, /mi/, /na/, /nou/, /ni/ with all possible tones in isolation, together with other syllables and sentences. The recording took place in a sound-proofed phonetic lab equipped with a DAT Tascam DA-40 recorder and a Sennheiser ME-64 microphone. The sounds were recorded with PRAAT software (Boersma, 2001). After recording, one token of each syllable with each tone were chosen for further manipulation.

The chosen syllables were divided into two groups—those involving T1 and T4, and those involving T2 and T3. Within each group, in order to ensure that the other acoustical properties such as duration and intensity were identical across different tones, the original syllables with the same segments were first converted into identical duration by using the “lengthening” function in PRAAT. After manipulation, the duration of the syllables had a range between 250 and 300 ms.

Second, to minimize variation in the stimuli carrying T1 and T4, for each syllable, we extracted the pitch contour of T1, and then replaced the original pitch contour of the T4 syllables with the extracted T1 contour. To minimize variation in the stimuli carrying T2 and T3, for each syllable, we extracted the pitch contours of the original T3, and replaced the original pitch contour of the T2 syllable with the extracted T3 contour. By manipulating the syllables in this way, we got a T4 syllable which has exactly the same segmental as well as intensity information as the original T1 syllable, and a T2 syllable which has exactly the same segmental and intensity information as the original T3 syllable. Although T3 tends to have longer duration (Xu, 1997), all the four Mandarin lexical tones can be produced with variable durations. In our pilot study, native Mandarin listeners were able to identify the four lexical tones in their citation forms correctly when duration was equalized, so the equal duration of our stimuli should not be a reason for perceptual confusion between the lexical tones.

Third, to generate the bisyllabic sequences, T1 and T4 realized on the same segment were concatenated to get a T1T4 sequence, and using the same method, we generated T4T1, T1T4 sequences. Similarly, T2 and T3 realized on the same segment were concatenated to get T2T3, T3T2, and T3T3 sequences. These bisyllabic sequences were used as stimuli. The T3 in T3T2 sequence had a full dipping contour rather than a low-falling contour for the following two reasons: first, either with the rising tail or not, the T3 is lower than T2, which ensures that T3 is perceived as a low tone, consistent with its distinctive feature. Second, as the Dutch listeners do not have a phonological representation of Mandarin tones, they perceive the tones acoustically. If we had cut off the rising tail of T3 in the T3T2 sequence, then this T3 would have a low-falling contour, different from the T3 having the dipping contour in T2T3 sequence. In this case, T2T3 and T3T2 would differ from the referent T3T3 not only in terms of “structure”, namely whether the left or the right tone is different from the referent, but also in terms of acoustics. It could be that Dutch listeners would, in that case, pay attention to the acoustical difference between the low-falling T3 and the dipping T3, which would have hindered us from understanding the processing of the asymmetry



**Fig. 2.** (a) The average pitch contour of T3T3 sequences. (b) The average pitch contour of T2T3 sequences. (c) The average pitch contour of T3T2 sequences. (d) The average pitch contour of T4T4 sequences. (e) The average pitch contour of the T1T4 sequences. (f) The average pitch contour of T4T1 sequences.

exhibited in T3 sandhi. By manipulating the original recordings as we did, we ensure that the pitch contour (*i.e.*, the lexical tones) is the only difference between the syllables with the same segments. Therefore, if the listeners discriminate between the sequences, the only information that they will have relied on is the lexical tone itself. The manipulated syllables were judged by multiple native listeners as natural and clear.

Fig. 2a–f plots the time normalized average pitch contour of T3T3, T2T3, T3T2, T4T4, T1T4, T4T4 sequences used in the experiment. The horizontal axis represents the percentage of total duration.

### 2.3. Procedure

The participants participated in a bisyllabic tone discrimination task. They were asked to discriminate between two sequences as quickly as possible by pressing one of two buttons on a button box, labeled “same” and “different”. Within one trial, the two sequences had the same segmental information. Taking T3T3 sequences as referents, participants were asked to indicate whether the T2T3 and T3T2 bisyllabic sequences were identical to the referents. Similarly, taking T4T4 sequences as referents, the participants were asked to indicate whether T1T4 and T4T1 were identical to the referents. For the ease of description, when the referents were composed of two identical tones, we called the referents ‘YY structure’, and depending on whether the first or second syllable was different from the referent, we called them either “XY structure” (*e.g.*, T1T4 and T2T3), or “YX structure” (*e.g.*, T4T1 and T3T2). In the case of T3 sandhi, the YY structure is sandhied to an XY structure, but not a YX structure. In order to observe whether the position (*i.e.*, the first sequence or the last sequence in the to-be-compared pair) of the referents influenced the discrimination, we made the referents occur at each position with equal chance. We also constructed “same” pairs as fillers repeating the afore-mentioned bisyllabic sequences within one trial (*e.g.*, T1T1–T1T1, or T2T3–T2T3). We also introduced T2T2–T3T3 or T1T1–T4T4 as fillers. The stimuli were randomized across the participants. The inter-stimulus interval between the two sequences in a trial was 1500 ms, and the response duration was 1 s (*i.e.*, if the participants did not give their response within 1 s, the next trial began, and they missed one trial).

### 2.4. Results and discussion

A three-way repeated measures ANOVA (structures (YX or XY to be compared with YY)\*tones (sequences involving T1/T4 or T2/T3)\*orders (referent occurred first or last in the to-be-compared pair)) was carried out with the accuracy of the different pairs of each participant. Language background was put into the model as a between-subject variable. A significant main effect was found for all three factors: Fstructures (1, 94) = 60.50,  $p < 0.001$ ; Ftone (1, 94) = 185.82,  $p < 0.01$ ; Forders (1, 94) = 50.53,  $p < 0.01$ . The effect of language group, on the other hand, failed to reach significance: Flanguage (1, 94) = 0.21, *n.s.*. Significant interactions were found between structure and tones, Fstructure\*tones (1, 94) = 28.25,  $p < 0.01$ , and between tones and orders, Ftone\*orders (1, 94) = 33.57,  $p < 0.01$ . Language significantly interacted with tones, Flanguage\*tones (1, 94) = 18.63,  $p < 0.01$ .

Fig. 3 plots the accuracy of the discrimination of Dutch and Mandarin listeners for each tone group. As can be seen from the figure, first, the structure of the tones (*i.e.*, either the first or the second syllable is identical to the referent) significantly influenced the accuracy in discrimination. Overall, both Mandarin and Dutch listeners were more accurate at discriminating the YX structure (compared to the XY structure) from the referents, namely, the YY structure. Specifically, if we look at the discrimination between T2T3 and T3T3 where T3T3 occurs first, the accuracy of Mandarin listeners was not significantly different from chance level,  $T(47) = -0.515$ , *n.s.*. This finding serves as evidence that native listeners’ knowledge of T3 sandhi involves the neutralization between T2T3 and underlying T3T3, rather than treating T3T3 as illegal and perceiving a boundary in between. Second, the discrimination accuracy was significantly higher for those sequences involving T1 and T4 than those involving T2 and T3. The higher accuracy of the discrimination involving T1 and T4 was consistent with previous findings, namely that the T2–T3 contrast is acoustically less salient than the T1–T4 contrast (Hume and Johnson, 2003). Third, the location (*i.e.*, first or last) of the referent in the to-be-compared pairs influenced the discrimination accuracy. Listeners were facilitated when the referent occurred first rather than last. The significant effect of order was not expected *a priori*. We interpret this order effect as the result of memory burden (Cowan & Morse, 1986; Demany and Semal, 2008). The listeners needed to hold both syllables of the first sequence in the to-be-compared pair in memory, and compare the following sequence to the first sequence. When the first sequence had two identical tones, the memory load was lower than when the first sequence had two different tones. Thus, where the two tones were identical, listeners were facilitated when the referent occurred first. Finally, the discrimination accuracy of Dutch listeners was not significantly different from that of Mandarin listeners. It has been shown that Dutch listeners perceived the Mandarin lexical tones acoustically whereas Mandarin listeners perceive tone categorically (Hallé et al., 2004; Xu et al., 2006; Chen, 2012; Liu et al., submitted for publication), here we find that, despite different ways of processing the lexical tones, and though naïve to Mandarin lexical tones, Dutch listeners were fairly accurate at discriminating lexical tones in these tasks.

The significant interactions reveal the specificity of the processing of T2 and T3. First, the significant interaction between tones and structures indicates that the discrimination of the sequences involving T2 and T3 is more influenced by the structure of the to-be-discriminated pairs. In other words, though listeners are less accurate overall when discriminating XY structures from YY structures than when discriminating YX from YY structures, such decline was more evident for the sequences involving T2 and T3. Looking at Fig. 3, it can be seen that both Mandarin and Dutch listeners were less accurate when discriminating T2T3 from T3T3 than when discriminating T3T2 from T3T3. However, for the sequences involving T1 and T4,



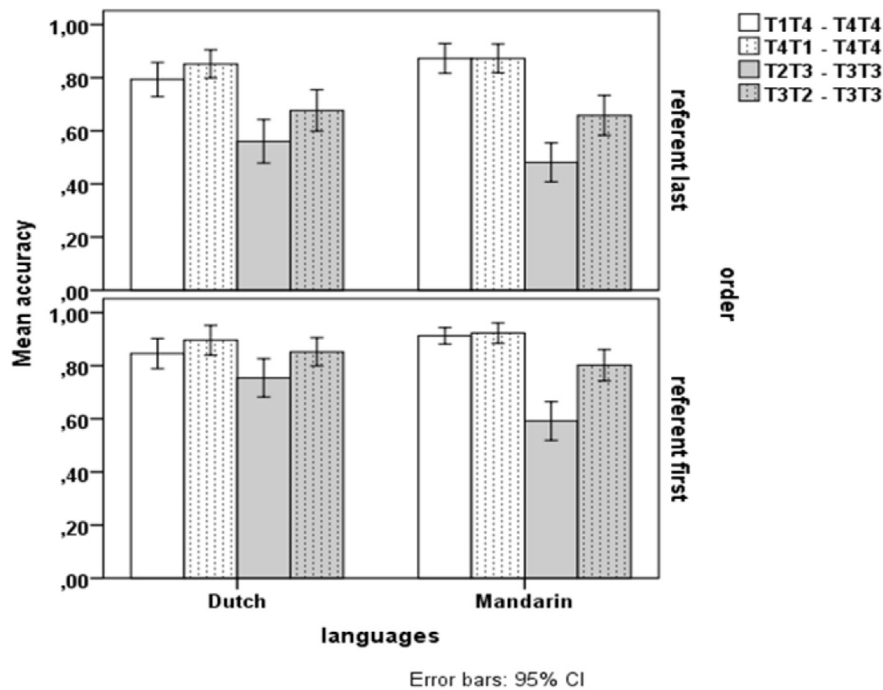


Fig. 3. The accuracy of Dutch and Mandarin listeners when discriminating the sequences involving T2/T3 and T1/T4.

only Dutch listeners showed a decline in discriminating XY from YY, namely, the discrimination between T1T4 and T4T4. Chinese listeners, on the other hand, reached equally high accuracy when discriminating the sequences involving T1 and T4. Dutch listeners showed a consistent tendency in the discrimination of both the T1–T4 group and T2–T3 group, in that the discrimination involving XY is less accurate than those involving YX. However, this effect is more evident for the discrimination involving T2 and T3. A repeated measures ANOVA with only Dutch listeners' responses confirms this point: there is significant interaction between structures (XY or YX to be discriminated from YY) and tone group (i.e., sequences involving T2 and T3, or sequences involving T1 and T4),  $F(1, 94) = 4.11, p < 0.05$ . Second, the interaction between tones and orders demonstrates that the discrimination of sequences involving T2 and T3 was more influenced by the position of the referent. When the referent occurred first in the to-be-compared pairs, the discrimination of the sequences involving T2 and T3 was significantly improved than when the referent occurred last. A similar yet much weaker trend was found for the sequences involving T1 and T4. We have interpreted the main effect of order as the result of memory load. The reason that the sequences involving T2 and T3 were particularly influenced by the location of the referent may be the complexity of the tonal contour of T3. T3 had a dipping contour, where the pitch went first down and then up, and its pitch direction change was more complex compared the other tones. The complexity of T3 makes the influence of memory more pronounced. Also, as T2 and T3 are acoustically similar, when a T2T3 sequence occurred first in the to-be-discriminated pair, to hold the similar-yet-not-identical tones in memory became challenging for the listeners. The fact that, regardless of language background, both the structure effect and the order effect are more evident for the discrimination of the sequences involving T2 and T3 sheds lights on the cognitive basis of the T3 sandhi: there may be an overall tendency to neutralize between the XY and YY structure in perception, and the complexity of T3 makes it more vulnerable to such neutralization. The cognitive constraints also hold for T1 and T4, yet for native speakers, the consolidated representation of these two tonal categories compensates for such constraints.

Importantly, we did not find a main effect of language. This means that Dutch and Mandarin listeners share similar discrimination patterns. For both groups, the discrimination involving T1 and T4 was more accurate than the discrimination involving T2 and T3. Also, for both groups, the discrimination between T2T3 and T3T3 was more difficult compared with the discrimination between T3T2 and T3T3. The Dutch listeners were naïve to Mandarin lexical tones, and their performance cannot be attributed to the knowledge of the phonological grammar. The consistency between Mandarin and Dutch listeners suggests that the asymmetrical pattern of T3 sandhi may well be related to general cognitive constraints. [Gandour and Harshman \(1978\)](#) argued that the “extreme endpoint” of a lexical tone reflects non-linguistic auditory properties of a tone. In our case, the Dutch listeners were discriminating the tonal sequences acoustically rather than linguistically, and they may attend to the non-linguistic auditory property of the sequences. Hence, if the offsets of two sounds are identical, the discrimination is more difficult. Also, the acoustics in speech are transient, and the memory of the onset may decay very quickly, which makes the discrimination between sequences differing in the first syllable more difficult.

At the same time, however, there is significant interaction between language and tones. We should see that, for Mandarin listeners, compared to Dutch listeners, the accuracy difference between the discrimination of T2T3 versus T3T2 with T3T3 is amplified. This means that, though Dutch listeners and Mandarin listeners possibly share the same cognitive constraints,

learning the grammar that neutralizes between T2T3 and T3T3 strengthens this natural preference. Our findings are consistent with [Huang and Johnson \(2011\)](#) who found that T2 and T3 are rated to be more similar by Chinese listeners compared with English listeners, which argued for the influence of phonology on perception. Taking together our study, [Hume and Johnson \(2003\)](#), and [Huang and Johnson \(2011\)](#), it can be seen that phonological grammar and natural cognitive preference interact and influence each other.

### 3. General discussion

In this paper, we examined the perceptual knowledge of T3 sandhi of Mandarin listeners, and whether the processing of T3 sandhi relates to more general cognitive constraints. We tested both native and non-native listeners on their discrimination between the underlying form of T3T3 and the *a priori* possible sandhi candidates, namely T2T3 and T3T2. To determine the specificity in perceiving the tones that are involved in Mandarin phonological grammar, we included sequences involving T1 and T4 as a control condition, as T1 and T4 do not undergo contextual tonal alternations. The results found in the experiments can be discussed from several perspectives. First, Mandarin listeners confused T2T3 and T3T3, which indicates that their perceptual knowledge of T3 sandhi involves a neutralization between these two sequences. The neutralization between T2 and T3 only occurs in the proper context, namely, when they precede a T3. Though T3s do not occur consecutively within a word in Mandarin Chinese, listeners do not necessarily perceive a boundary in between. We included non-native (Dutch) listeners to examine whether the structure of T3 sandhi relates to general cognitive constraints. Indeed, similarities between Mandarin and Dutch listeners were found for the following aspects: first, both groups discriminated between sequences involving T1 and T4 more accurately than those involving T2 and T3. Second, compared with T3T2, T2T3 is perceived to be more similar to T3T3, regardless of whether T3T3 occurred first or last in the to-be-compared pair. These findings add evidence to the claim that the T1–T4 contrast is acoustically more salient compared with the T2–T3 contrast ([Hume and Johnson, 2003](#)). Also, it demonstrates that T2T3 is perceived to be more similar to T3T3 than T3T2, and such perceptual asymmetry is not driven by learning the phonological grammar of T3 sandhi. Importantly, T3 sandhi occurs in such a way that it neutralizes the naturally perceptually similar sequences, namely T2T3 and T3T3, rather than between the perceptually less similar T3T2 and T3T3 sequences. Our findings provide more evidence for the view that phonological grammar reflects, and may be shaped, by perceptual constraints of the listeners ([Steriade, 1995, 1997](#); [Ohala, 1981, 1993, 2012](#); [Blevins, 2006](#)). Not only is there a confusion asymmetry of segments reflecting ease of perception ([Chang et al., 2001](#)), on the suprasegmental level, listeners also naturally confuse some tonal sequences more easily than the others. Nevertheless, regardless of the similarities found between Dutch and Mandarin listeners, language-specific patterns are still observed. The accuracy difference between T2T3–T3T3 discrimination and T3T2–T3T3 discrimination is amplified among Mandarin listeners. Such amplification reveals that the knowledge of phonological grammar, in return, shapes perception. Though the discrimination between T2T3 and T3T3 structure is hindered overall, learning a grammar that requires the neutralization between T2T3 and T3T3 strengthens the cognitive bias.

One question still remains unanswered; namely, why T2T3 is particularly confusable with T3T3. It should be noted that Dutch listeners discriminate between T4T1 and T4T4 more easily than between T1T4 and T4T4. Hence, we suspect that, generally, the discrimination between YX and YY is easier than between XY and YY. [Gandour and Harshman \(1978\)](#) pointed out that the offset of a lexical tone reflects the auditory non-linguistic property of a tone. Hence, it might be that, in auditory discrimination, listeners attend more to, or better memorize, the offset, whereas the memory of onset may decay quickly ([Cowan and Morse, 1986](#); [Demany and Semal, 2008](#)). As T2 and T3 have a similar tonal contour ([Hume and Johnson, 2003](#)), these two tones are particularly vulnerable to these cognitive constraints in discrimination. By this account, one possible interpretation of T3 sandhi is that T2T3 and T3T3 are perceptually similar, and the more complex T3T3 sequence is merged to the less complex T2T3 sequence. It should be seen that Mandarin listeners discriminate T1T4 and T4T1 equally well from T4T4, and this suggests that the cognitive constraints can be overcome through learning. It might also be that once learned the lexical tones, native Mandarin listeners reach ceiling accuracy when discriminating the acoustically salient T1 and T4, and the XY/YX–YY discrimination asymmetry is masked. For better understanding of the offset biases as assumed in the current study, discrimination of lexical tones that share offset but differ at onset, such as T1 and T2, should be tested in future studies.

The Chinese tone sandhies have been a thorny issue in the study of phonology for decades. We took Mandarin Chinese T3 sandhi as an example here, and found convincing evidence that there is a functional basis for the preference to neutralize between T2 and T3 in a specific context, namely before another T3. Such basis may also hold for the other right-dominant tone sandhies found in other Chinese dialects ([Zhang, 2007](#)). We speculate that transient and fast decaying acoustic information carried by the tones is influenced by constraints such as memory limit, especially in certain acoustic contexts. Thus, the various tone sandhi patterns may occur or even evolve in specific ways as a result of human cognitive constraints. We have studied one case in the current study, and more cross-linguistic research on how cognitive constraints interact with tonal grammar is needed. In addition, the left and right dominant sandhies may relate to different functional bases. To explore what drives the different realization of left and right dominant sandhies will extend our understanding of cognitive predispositions and speech prosody.

### References

- Beckman, M.E., Edwards, J., 1990. Lengthening and Shortening and the Nature of Prosodic Constituency. In: Kingston, J., Beckman, M.E. (Eds.), *Papers in the Laboratory Phonology I: Between the Grammar and the Physics of Speech*. Cambridge University Press, Cambridge, UK, pp. 152–178.

- Blevins, J., 2006. A theoretical synopsis of evolutionary phonology. *Theor. Linguist.* 32 (2), 117–166.
- Boersma, P., 2001. Praat, a system for doing phonetics by computer. *Glott Int.* 5 (9/10), 341–345.
- Chan, M.K.M., Ren, H.-M., 1989. Wuxi tone sandhi. *Acta Linguist. Hafniensia* 21 (2), 35–64.
- Chan, M.K.M., 1995. An autosegmental analysis of Danyang tone sandhi: some historical and theoretical issues. In: Zee, E. (Ed.), *Wuyu yanjiu* (Studies of the Wu Dialects). New Asia Academic Publishing, Hong Kong, pp. 145–184.
- Chang, S., Plauché, M., Ohala, J., 2001. Markedness and consonant confusion asymmetries. In: Hume, E., Johnson, K. (Eds.), *The Role of Speech Perception in Phonology*. Academic Press, pp. 79–101.
- Chen, A., 2012. Universal Biases in the Perception of Mandarin Tones, from Infancy to Adulthood (LOT dissertation series), p. 315.
- Chen, M.Y., 2000. *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge University Press, Berkeley, CA.
- Chen, Y., Yuan, J., 2007. A corpus study of the 3rd tone sandhi in standard Chinese. In: *Proceedings of Interspeech*, 2007, pp. 2749–2752.
- Cheng, R.L., 1968. Tone sandhi in Taiwanese. *Linguistics* 41, 19–42.
- Cowan, N., Morse, P., 1986. The use of auditory and phonetic memory in vowel discrimination. *J. Acoust. Soc. Am.* 79 (2), 500–507.
- Demany, L., Semal, C., 2008. The role of Memory in auditory perception. *Springer Handb. Audit. Res.* 29, 77–113.
- Edwards, J., Beckman, M.E., Fletcher, J., 1991. The articulatory kinematics of final lengthening. *J. Acoust. Soc. Am.* 89, 369–382.
- Gandour, J., Harshman, R., 1978. Cross-language difference in tone perception: a multidimensional scaling investigation. *Lang. Speech* 21, 1–33.
- Hallé, P.A., Chang, Y., Best, C.T., 2004. Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *J. Phonet.* 32 (3), 395–421.
- Ho, A.T., 1976. The acoustic variation of Mandarin tones. *Phonetica* 33, 353–367.
- Hombert, J.M., Ohala, J.J., Ewan, W.G., 1979. Phonetic explanations for the development of tones. *Language* 55, 37–58.
- Huang, T., 2004. *Language-specificity in Auditory Perception of Chinese Tones* (Doctoral dissertation). Ohio State University. Retrieved from: [http://rave.ohiolink.edu/etdc/view?acc\\_num=osu1092856661](http://rave.ohiolink.edu/etdc/view?acc_num=osu1092856661).
- Huang, T., Johnson, K., 2011. Language specificity in speech perception: perception of Mandarin tones by native and nonnative listeners. *Phonetica* 67, 243–267.
- Hume, E., Johnson, K., 2001. A model of the interplay of speech perception and phonology. In: Hume, E., Johnson, K. (Eds.), *The Role of Speech Perception in Phonology*. Academic press, pp. 3–26.
- Hume, E., Johnson, K., 2003. *The Impact of Partial Phonological Contrast on Speech Perception*. Retrieved from: [http://corpus.linguistics.berkeley.edu/~kjohnson/papers/Hume\\_Johnson2003.pdf](http://corpus.linguistics.berkeley.edu/~kjohnson/papers/Hume_Johnson2003.pdf).
- Johnson, K., Martin, J., 2001. Acoustic vowel reduction in Creek: effects of distinctive length and position in the word. *Phonetica* 58, 81–102.
- Klatt, D.H., 1975. Vowel lengthening is syntactically determined in connected discourse. *J. Phonet.* 3, 129–140.
- Lehiste, I., 1960. An acoustic phonetic study of internal open juncture. *Suppl. Phon.* 5.
- Liljencrants, J., Lindblom, B., 1972. Numerical simulations of vowel quality systems: the role of perceptual contrast. *Language* 48, 439–462.
- Lindblom, B., 1968. Temporal organization of syllable production. *Speech Transm. Lab. Q. Prog. Status Rep.* 2–3, 1–5.
- Lindblom, B., 1990. Explaining phonetic variation: a sketch of the H&H theory. In: Hardcastle, W., Marchal, A. (Eds.), *Speech Production and Speech Modeling*. Kluwer Academic Publishers, Dordrecht, the Netherlands, pp. 403–440.
- Lindblom, B., Engstrand, O., 1989. In what sense is speech quantal? *J. Phonet.* 17, 107–121.
- Liu, L., Chen, A., Kager, R., 2014. Tone perception in Mandarin and Dutch adult listeners. *Lang. Cognit. Neurosci.* (submitted for publication).
- Ohala, J., 1981. The listener as a source of sound change. In: Masek, Hendrik, Miller (Eds.), *Proceedings of the Chicago Linguistics Society*, 17, pp. 178–203. Chicago.
- Ohala, J., 1993. The perceptual basis of some sound patterns. In: Connell, D., Arvaniti, A. (Eds.), *Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence*. Cambridge University Press, Cambridge, pp. 87–94.
- Ohala, J., 2012. The listener as a source of sound change, an update. In: Sole, M., Recasens, D. (Eds.), *The Initiation of Sound Change, Perception, Production, and Social Factors*. John Benjamins, Philadelphia, pp. 21–35.
- Oller, D.K., 1973. The effect of position in utterance on speech segment duration in English. *J. Acoust. Soc. Am.* 54, 1235–1247.
- Peng, S., 1996. *Phonetic Implementation and Perception of Place Coarticulation and Tone Sandhi* (Unpublished PhD degree). Ohio State University, Department of Linguistics.
- Shih, C., 1986. *The Prosodic Domain of Tone Sandhi in Mandarin Chinese* (Doctoral dissertation). University of California at San Diego.
- Shih, C., 1997. Mandarin third tone sandhi and prosodic structure. In: Wang, J., Smith, N. (Eds.), *Studies in Chinese Phonology*. Mouton de Gruyter, New York, pp. 82–123.
- Speer, S., Shih, C., Slowiaczek, M., 1989. Prosodic structure in language understanding: evidence from tone sandhi in Mandarin. *Lang. Speech* 32, 337–354.
- Steriade, D., 1995. Neutralization and the Expression of Contrast.
- Steriade, D., 1997. Phonetics in phonology: the case of laryngeal neutralization. *UCLA Work. Pap. Linguist.* 3, 25–146.
- Wang, W.S., Li, K.P., Brotzman, R.L., 1963. *Research on Mandarin Phonology*. Project on Linguistic Analysis. Ohio State Research Foundation, pp. 1–63, 6r.
- Wang, W.S., Li, K., 1967. Tone 3 in pekinese. *J. Speech Hear. Res.* 10, 629–636.
- Wightman, C.W., Shattuck-Hufnagel, S., Ostendorf, M., Price, P.J., 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *J. Acoust. Soc. Am.* 91, 1707–1717.
- Wu, Z., 1982. Tonal variations in Mandarin sentences. *Chin. Linguist.* 6, 439–450.
- Xu, Y., Gandour, J.T., Francis, A.L., 2006. Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *J. Acoust. Soc. Am.* 120 (2), 1063–1074. Retrieved from: <http://link.aip.org/link/?JAS/120/1063/1>.
- Xu, Y., 1997. Contextual tonal variations in mandarin. *J. Phonet.* 25 (1), 61–83.
- Yue-Hashimoto, A.O., 1987. Tone sandhi across Chinese dialects. In: *Chinese Language Society of Hong Kong. Wang Li memorial volumes*, English, pp. 445–474.
- Zhang, J., 2007. A directional asymmetry in Chinese tone sandhi systems. *J. East Asian Linguist.* 16, 259–302.