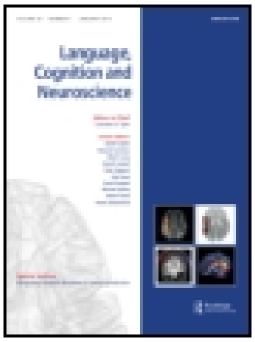

Language, Cognition and Neuroscience

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

Representation of “T3 sandhi” in mandarin: significance of context

Yaxuan Meng, Hilary Wynne & Aditi Lahiri

LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2021.1893769

REGULAR ARTICLE

Representation of "T3 sandhi" in mandarin: significance of context

Yaxuan Meng, Hilary Wynne and Aditi Lahiri

Faculty of Linguistics, Philology & Phonetics, Clarendon Institute, University of Oxford, Oxford, UK

ABSTRACT

T3 sandhi in Mandarin is a phonological alternation where two adjacent T3-T3 tones are not allowed and become T[2]-T3. Although this phenomenon has been extensively studied in previous work, concerns about how T3 sandhi is represented in the mental lexicon remain controversial. We approached the issue of representation in a series of cross-modal priming experiments, including two semantic priming experiments with and without context. Our results indicated that T2 failed to prime a target beginning with T3 in a non-sandhi context. However, both T2 and T3 were able to induce semantic activation when appropriate contextual information was present. Conversely, when such information was absent, only T3 activated the lexical entry. Thus, our study suggests that T3 is specified, as expected for dissimilation, and T2 is resolved by a re-writing rule only under appropriate contextual information.

ARTICLE HISTORY
Received 29 May 2019
Accepted 13 5.Add supplemental link February 2021

KEYWORDS
Context; dissimilation; T3 sandhi; Mandarin Chinese

1. Introduction

Contextual phonological alternations across morphemes, both segmental as well as tonal, are part and parcel of phonology. In a sequence of consonants, features of one can affect the other such that they can either become more similar (assimilation) or move apart (dissimilation). Assimilations can be either progressive or regressive: in progressive assimilation (e.g. *cat[s], dog[z]*), the plural marker <s> absorbs the voicing of the stem-final consonant while in regressive assimilation (e.g. English *in-possible* > *i[m]possible, in-coherent* > *i[ŋ]coherent*), the place of articulation of the first consonant of the stem (e.g. *possible, coherent*) affects the preceding nasal consonant. Tonal assimilations (also described as tonal spreading) are equally well established (cf. numerous examples from Hyman & Schuh, 1974, Gwari /òkpá/ > [òkpǎ] "length"; i.e. the low tone on the first vowel spreads to the second vowel /o^{L}kpa^H > o^{L}kpa^{LH}). Dissimilation, on the other hand, disallows items of the same category in a sequence, as in Latin where in some instances two laterals are not permitted and /l/ is replaced by /r/; e.g. *nav-alis* "naval", but *sol-aris* "solar" (Steriade, 1987, also reported in Gussenhoven & Jacobs, 2017). One notable example of dissimilation is Grassmann's Law in Sanskrit and Greek where two aspirated consonants are not allowed and therefore the first one becomes de-aspirated: e.g. Sanskrit *b^{h}a-b^{h}uva* > *ba-b^{h}uva*. De-aspiration as dissimilation is still prevalent in many Indo-Aryan languages such as Bengali /d̪ud̪^{h} b^{h}aṯ/ > [d̪ud̪ b^{h}aṯ] "milk rice".

This study focuses on tonal alternation, where the underlying mechanism concerning how the tone sequences are represented and processed during speech recognition has been widely debated. Tonal alternation is pervasive in Chinese dialects and perhaps the best-known example in Mandarin is the T3 "tone sandhi", where the first T3 in a sequence of T3-T3 words obligatorily changes from a low "dipping" (i.e. falling-rising) tone (T3) to a rising tone (T2), e.g. 采取, "to adopt". Thus, the T3 sandhi phonological rule would be characterised as T3 ➡ [T2] / _T3. The other two tones in Mandarin are T1 (high level) and T4 (which is falling). Although T3 is a falling-rising tone in isolation, elsewhere it can be considered to be an underlying L(ow) tone (cf. Chao,1968; Duanmu, 1999; Shih, 1997; Yip, 1980, and many others). What prevails is that T3 changes to T2 when another T3 follows.

In languages such as Mandarin, lexical tone must play a significant role in spoken word recognition and the obvious questions that arise are how T3-T3 sandhi sequences are represented in the mental lexicon and how the surface T2-T3 forms are mapped onto these representations. In other words, if lexical tone is used to constrain the activation of candidates that mismatch in

CONTACT Yaxuan Meng yaxuan.meng@ling-phil.ox.ac.uk Faculty of Linguistics, Philology & Phonetics, Clarendon Institute, University of Oxford, Walton Street, Oxford OX1 2HG, UK
Supplemental data for this article can be accessed at https://doi.org/10.1080/23273798.2021.1893769

tone, then how do Mandarin speakers process T3-sandhi words and how should such words be represented? Before we discuss our views, we will briefly consider the diverging literature.

2. Mandarin tone sandhi

Mandarin Chinese is a tonal language in which both segmental and suprasegmental (i.e. lexical tone) cues are lexically relevant: tonal information is used to differentiate meanings of words that share the same segmental properties (cf. Chao, 1968; Yip, 1980). Tonal alternations in Chinese dialects are described as sandhi. Mandarin T3-sandhi involves changing a sequence of T3-T3 to [T2]-T3. Following Shih (1997), we will mark the derived Tone 2 as [T2]; underlying tones will not have any brackets. The term "sandhi" means "coming together" and in the context of T3-sandhi (which prohibits the occurrence of identical tones), the term appears to be a misnomer. Although dissimilation is not "sandhi" in the strictest sense, we will continue to use the term in this paper due to its extensive application in previous literature. Four lexical tones exist in Mandarin (cf. Xu, 1993) and, according to the shape of the F0 contours, the four tones can be described as a high level tone (tone1 or T1), a rising tone (tone2 or T2), a dipping-rising tone (tone3, or T3), and a high falling tone (tone4, or T4) (see Figure 1). When combined with lexical tones, the base syllable consisting of identical segments can represent different word meanings. The classic example is *ma* with its four distinct meanings: "mother" (妈) with T1, "numb" (麻) with T2, "horse" (马) with T3, and "to abuse" (骂) with T4.

Tonal alternation in Mandarin is widespread and automatic in disyllabic compounds. For example, the lexical tone of the first morpheme in the disyllabic compound 采取, "to adopt", transforms from T3 *cai3* to T2 *cai2* when followed by another T3 取*qu3*. Therefore, the realisation of T3 in Mandarin is particularly sensitive to the context. In isolation, T3 first dips and then rises, forming a complex contour tone; when this occurs in connection to a following syllable with T3, the falling tone is not realised (cf. Shih, 1997) (see Figure 2). The experimental literature on the role of Mandarin tone in spoken word recognition has primarily focused on the notion that perception of a particular tone reduces the activation of potential candidates that mismatch in tonal information (Zhou, Qu, Shu, Gaskell, & Marslen-Wilson, 2004; Lee, 2007; Sereno & Lee, 2015). In an auditory form priming task, Lee (2007) examined the role of Chinese lexical tone in constraining lexical access. Four types of monosyllabic prime-target pairs were presented: pairs that share both segments and tones (e.g. *heng1-heng1*), pairs that shared only segments (e.g. *heng1-heng4*), pairs that shared only tones (*heng1-tai1*), and pairs that shared neither segments nor tones (e.g. *heng1-wa2*). Reliable facilitation was only found in identical pairs (i.e. *heng1-heng1*), but the targets failed to be activated when prime-target pairs only shared segments and not lexical tones; (i.e. prime *heng1* did not activate target *heng4*). These results indicated that listeners were using tonal information to reduce the number of competitors and restrict unintended activation (see also Zhou et al., 2004).

A number of studies have examined this phenomenon in order to identify the underlying mechanism that contributes to the recognition of T3 sandhi words (Chien, Sereno, & Zhang, 2016; Zhou & Marslen-Wilson, 1997). This process was described as dissimilation by researchers including Yip (1980) and Duanmu (1999). Under Yip's theory of tonal registers (p. 195), phonological tone is made up of two complex features which

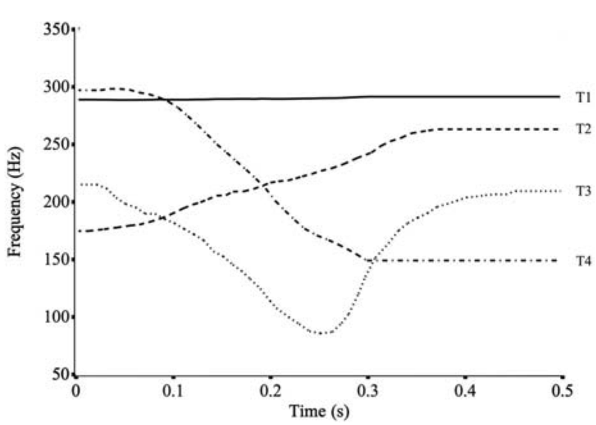

Figure 1. Illustration of F0 contours of the four Standard Mandarin lexical tones spoken in isolation with the segment /ma/ by a native Mandarin speaker. X-axis was normalised to the length of T3.

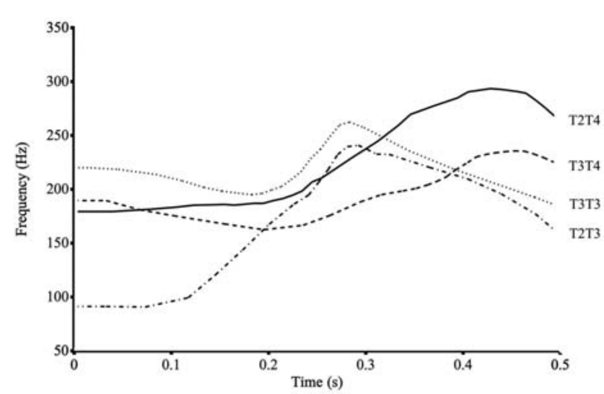

Figure 2. F0 contours of words with T2T3, 财产/cai2 chan3/ "property"; T2T4, 裁判/cai2 pan4/ "referee"; T3T3, 采取 /cai3 qu3/ "to adopt"; T3T4 采纳 /cai3 na4/ "to accept".

include Register [± Upper] and the melody or Tone [± High] "usually abbreviated to H and L". In her analysis, Register and Tone interact to account for the four Mandarin tones and the traditional number system is translated as the following: T1 (55) = [+ Upper] HH, T2 (35) = [+ Upper] LH, T3 (21) = [-Upper] LL, T4 (51) [+ Upper] HL. T3 sandhi is treated as a dissimilation process on the register level, "which dissimilates the feature [-Upper]" (p. 289). A similar analysis within an OT framework is provided by Duanmu (1999) who states that given an input /L-L/ (T3-T3), the preferred output would be LHL where "T3S can be seen as a dissimilation between two low tones" (p. 29). However, different views concerning how the T3 sandhi words are represented in the mental lexicon were proposed by a number of experimental studies (Chien, Sereno, & Zhang, 2016; Zhou & Marslen-Wilson, 1997).

In Zhou and Marslen-Wilson's seminal paper (1997), three hypotheses were investigated concerning the representation of T3 sandhi words in the mental lexicon: the "canonical representation", "surface representation", and "abstract representation". The "canonical representation" view assumed that all T3 sandhi words are represented in the citation form (i.e. *T3-T3*), despite the tone changing on the surface. In contrast, the "surface representation" view postulated that T3 tone sandhi words are always represented in their surface form, that is, *T3-T3 > T[2]-T3*. Finally, the "abstract representation" view assumed that the initial morphemes involved in the T3 sandhi have two surface forms, i.e. both T2 and T3, causing the lexical representation to be "abstract" and thus compatible to both surface forms.

To test these three hypotheses, Zhou and Marslen-Wilson conducted two auditory-auditory priming lexical decision tasks in which both primes and targets were disyllabic words (e.g. 才华 *cai2hua2, "*talent" ~ 采取 *cai3qu3, "*to adopt"). In Experiment 1, participants were asked to make lexical decisions on disyllabic tone sandhi *T3-T3* targets (pronounced as *T[2]-T3*) which had tonal alternation (e.g. 采取 *cai[2]qu3*, "to adopt" from the citation form *cai3qu3*). The target words were preceded by three types of disyllabic primes: T2-TX primes (where TX = any tone) with the same initial segments and surface tone of targets (e.g. 才华 *cai2hua2*, "talent"), T3-TX primes with the same initial segments and underlying tone of the targets (e.g. 彩虹 *cai3hong2*, "rainbow"), and control primes that were neither morphologically, semantically, nor phonologically related to the targets (e.g. T1-T2 天鹅 *tian1e2, "*swan").

Under the "canonical representation" hypothesis, it was predicted that both T2 and T3 initial primes would inhibit the processing of T3 sandhi targets due to phonological interference caused by the application of a re-writing rule to the initial syllable of T3 sandhi words. If the "surface representation" hypothesis was valid, it was predicted that the processing of T3-sandhi words would be inhibited by T2-initial primes but not T3-initial primes, due to cohort competition between T2 primes and T3 tone sandhi words sharing the same initial homophonic morphemes. Finally, under the "abstract representation" hypothesis, it was predicted that both T2 and T3 primes would inhibit the processing of targets because a T3 sandhi word is a cohort competitor of both T2 and T3 initial primes.

The results of Experiment 1 showed that the T2-TX primes significantly inhibited the processing of the T3-T3 sandhi targets as compared to controls, while T3-TX primes significantly facilitated the processing of targets (see Table 1 for a summary of the results). The authors proposed that these results best supported the "surface representation" view (i.e. T2 words cause inhibition), but also acknowledged that the "canonical representation" view was applicable on the basis that there may be two parallel processes competing in spoken word recognition – one process mapping the speech input (i.e. T2-T3) directly onto representations beginning with a T2, and a second process mapping onto representations beginning with T3 syllables using a re-writing rule. That is, when the second T3 in a sandhi word is heard by listeners, the lexical access system automatically interprets the first T2 as T3 and the represented canonical form is corrected through phonological inference. The garden-path will lead to the inhibition of T2 primes on T3 sandhi targets since T2 initial words would be activated when hearing the initial syllable of T3 sandhi words, resulting in re-activation. However, it remained unclear why the interference did not occur when T3 sandhi targets were preceded by T3 primes.

To verify the possibility of the "canonical representation" view, a second auditory-auditory priming experiment was conducted to investigate specifically whether the priming effect of T3-T3 sandhi words was similar to that of a T2 initial prime (i.e. T2-T3) or any other T3

Table 1. Summary of the results in Zhou and Marslen-Wilson's and Chien et al.'s studies.

Study	Auditory PRIME	Auditory TARGET	Result
Chien et al. (2016)	T2	T3-T3 > T[2]-T3	No priming
	T3	T3-T3 > T[2]-T3	Priming
Zhou & Marslen-Wilson (1997) (Exp1)	T2-TX	T3-T3 > T[2]-T3	Inhibition
	T3-TX (≠T3)	T3-T3 > T[2]-T3	Priming
(Exp2)	T3-T3 > T[2]-T3	T2-TX	Inhibition
	T3-T4	T2-TX	Inhibition
	T2-T3	T2-TX	Inhibition

initial prime (e.g. T3-T4). In contrast to the first experiment, the T3 sandhi words (e.g. 采取 *cai[2]qu3*, "to adopt" derived from *cai3 qu3*) served as primes rather than targets, and the targets were disyllabic words consisting of genuine initial T2 morphemes without tone alternation (e.g. 裁判 *cai2pan4 "*referee"). Four types of primes were used: words containing initial homophonic T2 morphemes of targets (e.g. 财产 *cai2chan3*, "property"), T3 initial words containing the same initial segments of targets but a different tone (e.g. 彩虹 *cai3hong2,* "rainbow"), T3 tone sandhi words with the same initial segments and surface tone of targets (e.g. 采访 *cai[2]fang3*, "to interview"), and controls bearing no semantic, morphological, or phonological relationship with targets (e.g. 预料 *yu4liao4*, "to predict"). Zhou and Marslen Wilson reasoned that, if the "surface" and "abstract representation" views were valid, then both the T2 and T3 initial sandhi primes would inhibit the processing of T2 targets. Alternatively, if the "canonical representation" view was the correct one, T2-initial primes were predicted to facilitate processing while T3-initial primes and T3 sandhi words were predicted to inhibit processing based on the assumption of strategic control of direct access and access via phonological interference. Unfortunately, the results of Experiment 2 showed an overall inhibitory effect of T3 sandhi, T2, and T3 initial primes on the processing of targets compared to the controls; i.e. **all primes** triggered inhibition. These findings were consequently inconsistent with all of the predictions based on the three representational hypotheses. Table 1 provides a summary of their results.

Although none of the three views could fully explain the results of both experiments, it was proposed that the "abstract representation" hypothesis held the least advantage due to the practical role of lexical tone in reducing ambiguity in Mandarin. That is, the authors reasoned that underspecification of tonal information in the lexicon would introduce inefficiency and complication to the process of lexical access and therefore was inappropriate for the processing of tonal sandhi. As will become apparent below, we believe that Zhou and Marslen-Wilson were correct in their interpretation.

The somewhat indeterminate results above prompted further investigations into the representation of T3 sandhi words, and various avenues have been examined in more recent studies (cf. Chien et al., 2016; Zhang & Lai, 2010). Chien et al. (2016) adopted an auditory-auditory priming task similar to that in Zhou and Marslen-Wilson (1997). They also tested two hypotheses: surface and underlying, the latter being the same as the "canonical representation" view in Zhou and Marslen-Wilson (1997). Unlike Zhou and Marslen-Wilson (1997), Chien et al. (2016) used three types of monosyllabic primes: T2 primes that overlapped with the targets in the surface form, T3 primes that overlapped with the targets in the underlying form, and an unrelated control prime (always with T1). The targets were all disyllabic T3-T3 words which underwent tonal alternation; i.e. the participants heard T2-T3 from underlying T3-T3. They found that the T3 primes (e.g. *fu3* 辅, "to guide") significantly facilitated the processing of the T3 sandhi words (e.g. *fu3dao3* 辅导, "to counsel") while T2 primes (e.g. *fu2* 服, "to assist") did not, even though the T2 primes overlapped with the targets in the surface tone. Thus, a monosyllabic T3 caused priming of T3-T3 targets, while T2 did not. On the basis of this finding, the authors argue that T3 sandhi words are represented as T3-T3 in their underlying forms in the mental lexicon.

The two studies (Chien et al., 2016, Zhou & Marslen-Wilson, 1997) notably differed in terms of the auditory primes: Chien et al. used monosyllabic primes while Zhou and Marslen-Wilson used disyllabic primes (an overview of the results of Chien et al. 2016, and Zhou & Marslen-Wilson, 1997 is provided in Table 1). Thus, it is only possible to really compare Zhou and Marslen-Wilson's Experiment 1 with Chien et al.'s study since Experiment 2 employed different non-sandhi targets. In Table 1, we can see that Zhou and Marslen-Wilson found activation of T3-T3 targets with both T2-T3 and T3-TX primes, except that for the former, the result was inhibition, while for the latter the priming was positive. Recall that Zhou and Marslen-Wilson (1997) argued that inhibition for the T2-T3 primes was due to word cohort competition. If we now compare the initial morpheme of these disyllabic primes, viz. syllables with T3 and T2, then Zhou and Marslen-Wilson's results are not comparable with Chien et al. who only found priming with T3 primes, and no priming or inhibition with T2 primes.

Given the polysemous nature of Mandarin tones, a monosyllabic word with a specific tone can lead to multiple lexical meanings, and thereby competition; cf. the well-known example of /ma/ which can occur with any of the four tones, each with a different meaning (see Figure 1 above). Thus, for a syllable heard in isolation, a very large semantic cohort must presumably be accessed and activated; cf. Lee (2007) discussed earlier, who obtained only identity priming where both segments and tones of the prime and target matched. Competition among a large semantic cohort may lead to the failure of T2 primes to activate the targets. This issue aside, however, it is crucial to note that both studies elicited evidence in support of underlying/canonical T3 representations of T3-T3 sandhi words.

Different views about the possibility of representation of T3 tones have also been raised in other studies (Li &

Chen, 2015; Politzer-Ahles et al., 2016) and support for these views came from electrophysiological evidence where asymmetric mismatch negativity (MMN) responses were found using an oddball paradigm. In this paradigm, participants listen to the repetition of auditory stimuli (standards) followed by an unexpected deviant. The assumption is that repetitive presentation of a standard stimulus activates the phonological representation of a certain sound; when a deviant sound mismatches with this representation, an MMN is elicited. The amplitude of MMN is positively correlated with the difference between standards and deviants, that is, a larger MMN is induced by larger difference between standards and deviants (Näätänen, Paavilainen, Rinne, & Alho, 2007). A prototypical standard with an atypical deviant exhibits smaller MMNs than vice versa (Ikeda, Hyashi, Hashimoto, Otomo, & Kanno, 2002). Correspondingly, MMNs are often smaller when standards have underspecified phonological representations with fully specified deviants and vice versa (Eulitz & Lahiri, 2004; Cornell, Lahiri, & Eulitz, 2013). Prototypicality effects have also been reported for stress, where illegally stressed standards with legal deviants demonstrate less MMN effects than the reverse (Honbolygó & Csépe, 2013). Where the Mandarin tone sandhi is concerned, the question is whether T3 is prototypical and/ or underspecified.

Li and Chen (2015) examined the issue of whether a T3 allophonic variant ("T3V") could be activated automatically when processing isolated T3 words without tone sandhi context. Stimuli with the same segment (e.g. /ma/) but different lexical tones (T1, T2, and T3) were chosen and in total, four conditions were constructed (T1/T3, T3/T1, T2/T3, T3/T2; standard/deviant). The authors found that, compared to between-category tonal variation, the activation elicited by within-category tonal variation was weaker within the left hemisphere, but stronger in the right. Both T1/T3 and T3/T1 conditions led to similar MMNs in terms of amplitude and latency; i.e. the effects were symmetric. However, asymmetric effects were found for the T2-T3 pair: higher and earlier MMNs were observed in the T2/T3 tone pair compared to the reversed T3/T2 condition. In addition, the MMN detected in the T3/T2 condition showed an earlier peak than in both the T2/T3 and T1/T3 conditions. To explain the results, the authors proposed that the lexical T3 has two representations in long-term memory, the canonical T3 and the variant T3V, whose pitch contour is similar to that of T2. Consequently, when T3 serves as the standard, both the canonical T3 and its variant T3V representation are activated from long-term memory. Consequently, little conflict between T2 and T3V is generated when listeners encounter the deviant T2 in continuous T3 speech streams. A similar result was also reported by Nixon, Chen, and Schiller (2015) for the T3 sandhi context. In their study, two picture-word interference experiments were carried out and T3 sandhi words were presented either as targets or distractors. During the tasks, participants were instructed to ignore a visual distractor while naming a picture as quickly and accurately as possible. The results in both two experiments provided evidence for priming effects of T2 and T3 monosyllabic or disyllabic words with initial T2 or T3, suggesting multiple levels of representation of T3 in sandhi words. To clarify, T2 and T3 sandhi words share the similar pitch contour while belonging to different tone categories; T3 and T3 sandhi words belong to the same tone category but have different tone realisation. Thus, the conclusion was that the priming effect of T3 words reflected the activation of the tone category representation, whilst the facilitation of T2 words suggested a context-specific representation of the actual pitch contour.

In contrast to Li and Chen's explanation, Politzer-Ahles et al. (2016) proposed that T3 is always underspecified. The rationale behind the assumption of underspecification is as follows. If a standard is underspecified for a specific feature which the deviant has, then the MMN will be smaller; if the standard, in contrast, is specified for the feature opposing to that of the deviant, then the MMN will be larger. Earlier work on segmental features has provided evidence in support of these assumptions (cf. Eulitz & Lahiri 2004; Cornell, Lahiri, & Eulitz, 2013). It has been argued that phonemes are made up of a number of features, some of which can be underspecified (Archangeli, 1988). Phonemes /g/ and /d/ share many features such as [VOICE] and [OBSTRUENT] but differ in their place of articulation, /g/ being [DORSAL] and /d/ being [CORONAL]. It has been argued that [CORONAL] is underspecified in the mental representation while [DORSAL] is not. Under this assumption, for the pair /g/ ~/d/, if /g/ is the standard and [d] is the deviant, then $/g/_{[d]}$ triggers higher MMN since the [CORONAL] extracted from the signal of [d] mismatches with the representation of [DORSAL] of /g/. Conversely, for $/d/_{[g]}$ the MMN peak will be significantly lower because the extracted feature [DORSAL] does not conflict with the underspecified feature [CORONAL] of /d/. Translating this rationale for tonal sandhi, Politzer-Ahles et al. (2016), maintained that their results for T3-T2 showed a similar pattern asymmetry as has been observed for [CORONAL] underspecification. They first used three tone pairs combining with different vowels *yi([i])* (Experiment 1), *wu([u])* (Experiment 2): T2-T3, T2-T4, T4-T3. Each tone pair could be either a standard or a deviant. Their results showed that whenever T3 was used as a standard, the

MMNs were lower. Thus, for both T2/T3 and T4/T3 (standard/deviant) pairs, the MMNs were always asymmetric with higher MMNs when T3 was the deviant. To minimise the low-level perceptual features and maximise the phonological processing, all four tones (T1, T2, T3, and T4) were produced on five vowel carriers *yi([i])*, *wu ([u])*, *yu([y])*, *e([ɤ])*, and *a([a]) in* Experiment 3. Additionally, three artificial tones were created by extracting and modifying portions of English speech. All possible combinations of the four tones (regardless of carrier vowels) in two directions, as well as a control block including the tokens from each of the four tones and three artificial tones, were presented. The results remained unchanged even when the T3 tokens were reduced with half the contour, and such asymmetry held in a contrast between T1 and T3, where T1 is a level tone. These asymmetric MMNs led the authors to argue that T3 is always underspecified and not just in the context of T3-T3.

Thus, the views on the representation of T3 and particularly how T3 sandhi words are represented in the mental lexicon remain controversial. To further investigate the nature of the representation of Mandarin T3 sandhi words, we approached the problem via a series of cross-modal priming experiments. Experiment 1 employed a lexical decision task, where the disyllabic targets were always non-sandhi words beginning with T3; i.e. there were no T3-T3 targets. The monosyllabic primes were T2, T3, or T4. The goal was to investigate whether only identical tones triggered priming or whether T2 behaved differently. Experiments 2 and 3 built on the results of Experiment 1, using direct and indirect semantic priming to tease apart the relationship between the surface T2 which resulted from a sandhi context and underlying T2.

3. Experiment 1: cross-modal priming

The aim of the current study was to investigate whether primes sharing the same syllable with initial target but different tone could trigger priming of the targets, providing a baseline of how T3 words are represented in the mental lexicon. As mentioned above, tone sandhi in Mandarin Chinese involves the first tone of the sequence T3-T3 becoming more dissimilar compared to the second tone. Although listeners cannot anticipate the upcoming tone when they perceive the first tone as T3, while they could negatively predict that the upcoming tone cannot be another T3. Furthermore, the activated representation when hearing T3 has to be T3 since it itself cannot be derived from another source. In contrast, T2 can be derived from T3 in the right context; thus, it can be represented as itself but also be contextually considered as underlyingly T3.

In Experiment 1, we investigated the extent to which monosyllabic T2, T3, T4 primes would activate disyllabic targets beginning with T3 where the second member could be T1, T2 or T4. That is, there were no sandhi T3-T3 targets. Response time (RT) was measured for a target like 反射 *fan3she4* "reflection" when primed by T2, T3, or T4 syllables such as *fan2*, *fan3* or *fan4*. If the activation found is due to the surface acoustic similarity between the two tones, then the expectation would be that the *fan3* prime will lead to significantly faster RT compared to T4 controls. On the other hand, if T3 is underspecified, then T2 would also activate the target. Our expectation was that only identical tones will lead to activation.

Table 2. Example of primes and targets in Experiment 1.

Auditory prime	Example	Visual target
T2	*fan2* "annoying"	反射
T3	*fan3* "opposite"	*fan3she4*
T4 (control)	*fan4* "meal"	"reflection"

3.1. Method

3.1.1. Participants

The first two experiments were conducted in a quiet room at Tsinghua University in Beijing. 52 native speakers (female = 27; average age = 21.4 years) of Mandarin took part in Experiment 1. Two participants who later reported that they also spoke a different dialect were excluded. All other participants were raised in Beijing and only spoke Standard Mandarin. All participants had normal or corrected-to-normal vision and no hearing impairments.

3.1.2. Materials

Thirty-six disyllabic target words were selected. As mentioned above, all targets began with a T3 syllable, while the second syllable could be T1, T2 or T4 (e.g. 反射 *fan3she4* "reflection"; see Table 2 and Appendix A). Three sets of monosyllabic primes were constructed which were segmentally identical with the first syllable of the target (e.g. *fan2* "annoying", *fan3* "opposite" or *fan4* "meal" ~ 反射 *fan3she4* "reflection"), but had different tones (either T2, T3, or T4). None of the primes was directly semantically related to the target.

All monosyllabic T3 primes were real words. When constructing the corresponding T2 primes, half of these items were real words (e.g. *niu2* "cow") and half were nonwords (e.g. *zong2*). The T2 nonwords were chosen as primes given that the homophones exist pervasively in Chinese. In other words, each syllable can correspond to multiple morphemes and each morpheme can make up many disyllabic compound words. The existence of homophones could elevate the strength of lexical competition and render the priming effect even though an auditory prime and target can be very distant in meaning. To circumvent the homophone problem and its potential effect on priming, the nonword T2 primes were adopted because segment-tone combinations could exist that are phonologically plausible yet meaningless without a corresponding morpheme. To avoid the influence of lexicality of T2 on the final analysis, the lexicality of T2 primes was well-controlled as a factor. The corresponding T4 control primes were real words. Thirty-six additional disyllabic real-word targets were chosen as fillers and an additional 72 nonword targets were chosen to counterbalance the number of *yes/no* responses. No auditory stimuli or characters were repeated twice. Additionally, all characters of the primes and targets were matched for frequency defined by the *List of Frequently Used Characters in Modern Chinese* (1988).

3.1.3. Recording of stimuli

All auditory stimuli were recorded by a female native Standard Mandarin speaker from Beijing. Stimuli were written in *pinyin* so that the pseudo-word primes could be pronounced naturally. Recording was conducted with the software Audacity in a sound-attenuated room with a Roland R-26 WAV recorder at a sampling rate of 44.1 kHz. The auditory stimuli were then extracted using the acoustic analysis software PRAAT (Boersma & Weenink, 2015), and the level of all auditory items was equalised.

3.1.4. Procedure

The stimuli list consisted of three trial types: critical trials, fillers, and nonwords. The list was pseudo-randomised with the constraint that there were not more than four responses in a row that required the same response key. For critical trials, a Latin-square design was used to make three versions of the list so that a third of the target words were preceded by T3 primes, a third were preceded by T2 primes, and a third were preceded by T4 control primes. Visual targets in the three versions were the same and occurred in the same sequence. The only difference across the three versions was the tone of the auditory primes on critical trials. This ensured that every participant saw each target only once and heard all three prime conditions. Primes on noncritical trials (fillers and nonwords) were chosen so that each auditory prime was heard only once and all tones occurred with equal likelihood. No auditory stimulus was repeated.

The stimuli were presented with experimental software developed by Reetz and Kleinmann (2003). Each trial started with a "beep" tone. The auditory primes were played through the headphone 300 ms after the offset of the "beep." Visual targets (in simplified Chinese characters) were then displayed for 300 ms immediately at the offset of the auditory primes. The inter-trial interval was 1600 ms. Participants were required to make a lexical decision on the visual target as quickly and as accurately as possible.

3.1.5. Data Screening

Raw data were trimmed using the following procedures which remained constant across all three experiments. Participants as well as items whose overall accuracy was lower than 2 standard deviations (*SD*) across the board were excluded. This resulted in a loss of 14.2% of the data (254 data points). In addition, responses with reaction times (RT) above 2*SD* of all responses were also excluded, which led to a further loss of 5.2% of the data (80 data points). Goodness of fit was established by model comparison and normality of residuals.

3.2. Results

3.2.1. Model Fitting

Reaction times were analysed using linear mixed effect models (LMMs) using the R package *lme4* (Bates et al., 2015). The dependent variable was RT with two fixed factors: *Prime Tone* (T2, T3 or the control T4) and *Lexicality* (of T2 primes: word or pseudo-word), and two random factors: *Subject* and *Item*. Since the accuracy rate was very high, to avoid ceiling effects, it was not taken into consideration in the final analysis. The analysis started from the maximal model with two fixed factors: *Lexicality* and *Prime Tone*, two random intercept terms of *Subject* and *Item*, as well as random slopes of *Subject* indexed by *Lexicality* and *Prime Tone*, and random slope of *Item* indexed by *Prime Tone* (no random slope of *Item* by *Lexicality* was included because *Lexicality* was a between-item factor).[1]

A series of likelihood ratio tests were conducted and a linear model with fixed factors *Prime Tne* and random intercepts for *Subject* and *Item* provided the best model fit.[2] In this model, *Prime Tone* had a significant effect ($\chi^2(2) = 16.03$, p=0.0003*) upon *RT*: T3 primes significantly facilitated the processing of targets compared to T4 controls (Est. = −14.47, SE = 4.85, t = −3.01, p = .007*), while the response time between T2 and T4 did not differ significantly (Est. = 3.82, SE = 4.85, t = 0.79, p = .71). Therefore, T3 primes facilitated the processing of targets whereas T2 primes did not (see Table 3 and Figure 3 below).

Table 3. Mean Response Times (in ms) for Experiment 1.

Condition	Mean	SD	Priming effect (compared to T4)
T2	582	102	4
T3	565	102	−13**
T4 (control)	578	100	N/A

Note. ** *p*< .001. The negative value of priming effect represents faster response.

3.3. Discussion

Experiment 1 was conducted to explore whether primes sharing the same initial syllable with the targets but different tone could trigger the facilitation of the disyllabic targets beginning with T3. The results showed a significant priming effect of T3 primes when compared to T4 controls. Recall that T3 primes and targets shared the same initial tone and morpheme (e.g. *fan3 "*opposite" ~ 反射 *fan3she4 "*reflection"). However, no facilitation was observed for T2 primes. This finding suggests that an isolated T2 syllable cannot prime a target beginning with T3 in a non-sandhi context. To further explore the underlying mechanism of the faciliatory effect of surface tones on T3 tone sandhi words, we next conducted a semantic priming experiment to test whether the activation of T2 primes is universal to all T3 initial words.

4. Experiment 2: non-mediated semantic priming

After the base-line study in Experiment 1 indicated that a surface T2 syllable does not activate a T3 initial word in a cross-modal priming task, a cross-modal semantic priming task (Experiment 2) was planned to investigate whether T2 primes could activate visual targets which are semantically related to the T3 primes. The primes in this experiment were monosyllabic words consisting

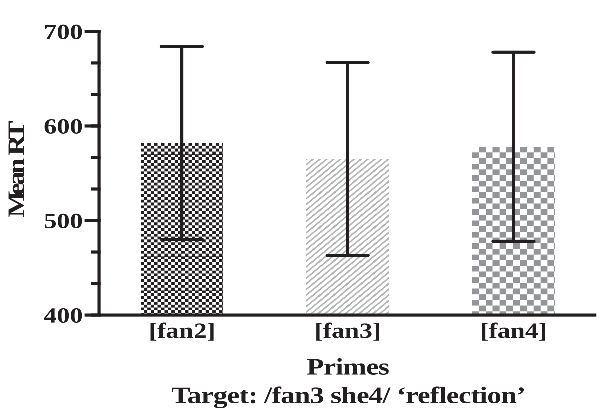

Figure 3. Mean response times for the three priming conditions in Experiment 1. For example, when the target word was /fan3 she4/ "reflection", the primes were: [fan2] "annoying", [fan3] "opposite" and [fan4] "meal".

Table 4. Example of primes and targets in Experiment 2.

Auditory prime	Example	Visual target
semantically related T3	*nao3* "brain"	头部
T2	*nao2* "to scratch"	*tou2 bu4*
T4 (control)	*nao4* "noisy"	"head"

of T2, T3, and T4. The disyllabic targets were semantically related only to the T3 prime. Again, like Experiment 1, the disyllabic targets could be anything other than T3 sandhi words: e.g. *nao2 "*to scratch, *nao3"*'brain, *nao4* "noisy" for the target: 头部*tou2bu4* "head" (Table 4). It was predicted that only T3 primes would activate the visual targets.

4.1. Method

4.1.1. Participants

The experimental conditions were the same as in Experiment 1. For this experiment, 52 native speakers (female = 35; age = 23.8 yrs.) of Mandarin who did not participate in Experiment 1 took part. Five participants who reported speaking another dialect were excluded. The remaining participants were raised in Beijing and only spoke Standard Mandarin. They all stated that they had normal or corrected-to-normal vision and no hearing impairments.

4.1.2. Materials

All primes were selected on the basis of the SubtLex (Cai & Brysbaert, 2010) database, as well as the *Modern Chinese Word Frequency List by Part of Speech* and could be identified as monosyllabic words or free morphemes. The primes were matched monosyllables, bearing the three tones T2, T3, and T4, generating 108 primes in total. Among the primes, half of the T2 primes were words and half were pseudo-words. As for T4 controls, all were real words. The disyllabic targets were selected from the definitions of each prime word provided in the *Xinhua Dictionary* (2004*)* to ensure semantic relatedness. For example, *xiang3* "to think" is chosen as a base word from the dictionary and allowed us to construct a suitable target *si1kao3* "consideration" (Appendix B).

Subjective ratings were acquired to ensure the semantic relationship between prime and target (see the semantic relatedness rating below). Additionally, we ensured that no T2 or T4 primes were synonyms or semantically related to the target. To match the percentage of word/nonword targets, 36 real word fillers and 72 nonword targets were chosen. All words were verified for character frequency against the *List of Frequently Used Characters in Modern Chinese* released by the State Language Commission (1988).

4.1.3. Recording and procedure

Recordings and procedures were the same as in Experiment 1.

4.1.4. Semantic Relatedness Rating

To ensure that the monosyllabic base and the disyllabic target were semantically related, subjective ratings were acquired. Items were rated by a separate group of 52 participants from Tsinghua University using a Likert scale from 1 to 7, where 7 was considered to be highly semantically related and 1 was considered to be not at all semantically related. Participants were asked to circle the number that best fit the relatedness between the primes and targets on a sheet of paper. The average rating score was 5.51, with no item being rated below 4. Therefore, all prime-target pairs were considered suitable items and included in further analyses.

4.1.5. Data Screening

Items, participants, and reaction time results were subjected to the same screening procedure as detailed in Experiment 1 and this resulted in a loss of 11.94% of the data (198 data points).

4.2. Results

4.2.1. Model Fitting

The optimal model contained the fixed factor *Prime Tone* and random intercepts of *Subject* and *Item* ($\chi2$ (2) = 9.40, p = .009*).[3] According to this model, T3 primes led to significantly faster reaction times compared to the T4 controls (Est. = −14.90, SE = 5.14, t= −2.90, p=.01*); whereas response latencies in the T2 prime condition did not differ significantly from those of the T4 control (Est. = −3.05, SE = 5.13, t = −0.60, p=.83) (see Table 5 and Figure 4).

4.3. Discussion

The results in Experiment 2 revealed that responses to disyllabic targets which were synonymous to preceding T3 monosyllabic primes were significantly faster than the corresponding T4 control primes. However, T2 primes had no effect on the targets: when compared to the controls, the reaction times were not significantly different.

Table 5. Mean Response Times (in ms) for Experiment 2.

Condition	Mean	SD	Priming effect (compared to T4)
T2	584	99	−3
T3	573	106	−14**
T4 (control)	587	106	N/A

Note. ** p< .01. The negative value of priming effect represents faster response.

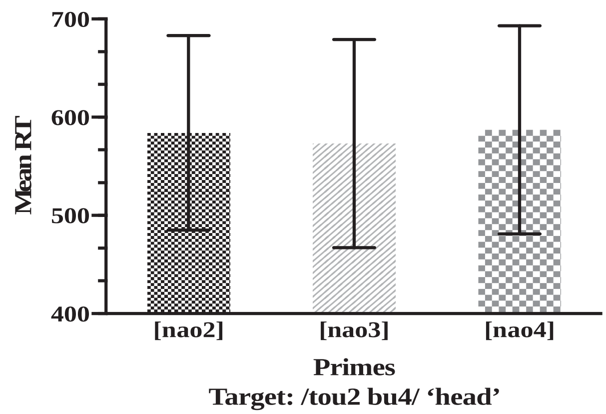

Figure 4. Mean response times for the three priming conditions in Experiment 2. For example, target word /tou2 bu4/ “head” and primes [nao2] “to scratch”, semantically related [nao3] “brain” and [nao4] “noisy”.

Therefore, it appears that the surface T2 primes did not activate targets that were semantically related to the T3 primes. In contrast, the priming effect by T3 primes which were semantically related to the targets was substantial and straightforward.

However, there is another factor that needs to be considered. Although our monosyllabic T2 primes were not semantically related to the target in isolation, they could be if presented in context. For instance, while *nao2* “to scratch” is not by itself synonymous with *tou2bu4* “head”, it could be if combined with other contextually-related morphemes, e.g. *nao3sun3 > nao[2]sun3* “brain damage”. In other words, the monosyllabic T2 could be semantically related to the target through the underlying mediated T3 sandhi word. If this contextually-driven relationship is not strong enough to facilitate the processing of targets, transfer from T2 to T3 will be blocked. Thus, it is clear that the semantic relationship between the underlying T3 sandhi words (e.g. /nao 3 sun3/ “brain damage”) and targets (e.g. /tou2 bu4/ “head”) must be tightly controlled. To further explore the relevance of context in the priming of tone sandhi words, we constructed a mediated semantic priming task.

5. Experiment 3: sandhi mediated semantic priming

Recall that in Experiment 1, T3 primes were found to activate non-sandhi targets beginning with T3 followed by any other tone (i.e. T3-TX). T3 primes were also found to activate semantically related targets which did not have T3 (i.e. T3 *nao3* “brain” primes *tou2 bu4* “head”; Experiment 2). Due to the effect of T3 sandhi, a T3 in the context of another T3 surfaces as T2. Consequently,

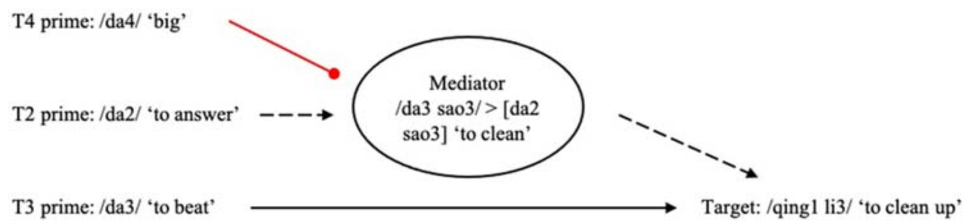

Figure 5. The priming effect of three different primes through the mediator in Experiment 3.

in earlier experiments, researchers investigated whether both T2 and T3 could activate a T3 sandhi word and Chien et al. (2016) found that in an intramodal auditory-auditory priming task, only T3 facilitated the sandhi *T3-T3 > [T2]* -T3, but T2 did not. Since we assume that the sandhi process is one of dissimilation, we also assume that the T3 in the sandhi context must be specified as T3; otherwise it cannot be changed in the context of a similar tone. That is, both tones have to match so that the first one can be replaced. Thus, contrary to Chien et al.'s study, we would expect that both T3 and T2 would activate a T3-T3 sandhi word, but only in the correct sandhi context. That is, T3 activation is predicted because T3 in sandhi words is represented as T3 itself, while T2 would activate the T3 only in the sandhi context due to an inferencing rule.

Experiment 3 was thus planned to extend the results of Experiment 2. Recall that monosyllabic T3 activated its semantically related target while the effect of the T2 prime was no different from the control T4 in Experiment 2, where the targets were not in a sandhi context. In Experiment 3 we tested the following question: will both T3 and T2 activate a *T3-T3 > [T2] -T3* word and in turn facilitate a target semantically related to the tone sandhi word? Figure 5 provides a diagram to illustrate the possible activation patterns. We expect that T3 *da3* would quite naturally activate *da3sao3* (since the first tone matches) and in turn facilitate *qing1li3* since these would be semantically related. The question is whether T2 would do the same. Since this is a sandhi context, we hypothesised that if *da2* could activate *da3sao3* (in a sandhi context), it too would activate a semantically related counterpart. The activation of *da3 sao3* by T2 would be an instance of "inferencing" or back-tracking as hypothesised by Gaskell and Marslen-Wilson (1998).

Experiment 3 has two parts – a direct semantic priming task (Experiment 3a) and a mediated semantic priming task (Experiment 3b) – outlined in Table 6 with a further diagram in Figure 5 to clarify the expected facilitation. In (3a), the primes are monosyllabic T2, T3, T4 with targets beginning with sandhi T3-T3 words. Thus, monosyllabic primes *da3* "to beat", *da2* "to answer", and *da4* "big" are paired with a target such as *da3sao3 > [da2]sao3* "to clean". Note that in this instance the primes are not semantically related to the target. The second part (Experiment 3b) includes the same primes, but the targets (e.g. *qing1li3*) are semantically related to the previous targets (e.g. *da3 sao3*). The nature of the prime-target relationship in the second part of the experiment was, thus, covert. The target was unrelated to the primes both in form and in direct meaning; rather, the target was chosen such that it was semantically related to a T3 sandhi word whose first tone was related in form to the T3 prime but not to the others. We describe the proposed activation in more detail with reference to Figure 5.

The predictions in Experiment 3a are as follows. We expect T3 primes like *da3* to facilitate the related sandhi target *da3sao3*. However, if T2 words like *da2* also facilitate the target, then it would be due to the fact that it is in the sandhi context. The predictions of Experiment 3b are more indirect. If the T3 primes like *da3* do succeed in activating *da3sao3,* then one might expect that a word semantically related to the target would also be activated; thus, *da3* should facilitate *qing1li3*. The same logic follows for T2 primes. If *da2* has successfully primed the sandhi word *da3sao3,* then it also could prime a semantically related word like *qing1li3*. If T2 words indeed do succeed in this indirect (mediated) priming, we can be very sure that these primes have actually facilitated a sandhi word; that is, *da2* will also have activated *da3sao3* since it surfaces as *da[2]sao3*. A question that follows is, will a rewriting rule along the lines of Gaskell and Marslen-Wilson (1998) allow the mediated facilitation if the T3 is in the correct sandhi context? If T2 primes do indeed activate visual targets which are neither semantically nor tonally related, then the activation can only be

Table 6. Example of task design in Experiment 3. The "mediator" condition is never presented but the first morpheme is homophonous to the T3 condition and it is semantically related to the target.

Auditory Prime	Example	Mediator	Visual Target
T3	*da3* "to beat"	打扫 *da3sao3* "to clean"	清理 *qing1li3* "to clean up"
T2	*da2* "to answer"		
T4 (control)	*da4* "big"		

realised through the underlying mediator. To ensure that the activation of visual targets was passed on from the activation of mediators, Experiment 3a was conducted to first explore the priming effect of auditory primes on the mediator and Experiment 3b was conducted to further investigate the priming effect of primes on visual targets.

5.1. Experiment 3a

5.1.1. Method

5.1.1.1 Participants. Forty-two native Mandarin speakers from the University of Oxford (female = 29; mean age = 24.2 yrs.) were recruited to participate in this experiment. All the participants were native Mandarin speakers and reported that they did not speak any other dialect other than Standard Mandarin. Participants either had normal or corrected-to-normal vision and none had hearing impairments.

5.1.1.2 Materials. All sandhi words were selected from the *Modern Chinese Word Frequency List by Part of Speech* and were ranked by frequency (see Table 6, Appendix C). The initial morpheme of all the sandhi words was the most frequent amongst its homophones. Also, no T2 or T4 words were synonyms or semantically related to the target. The lexical status (word or pseudo-word) of T2 critical primes was controlled in the analysis as a random factor (e.g. *zou3* is a word whereas *zou2* is a pseudo-word, although phonologically plausible). In total, 27.8% (10 out of 36) of the T2 critical primes were pseudo-words due to the restrictions mentioned above. The frequencies of both mediating sandhi words and target words were matched for T2 word and T2 pseudo-word primes, respectively. To match the percentage of word and nonword targets, 36 filler words and 72 nonword targets were chosen. All characters were verified for character frequency against the *List of Frequently Used Characters in Modern Chinese* (State Language Commission, 1988). In Experiment 3a, only the auditory primes with the corresponding T3 sandhi (mediated) target word pairs were presented to participants.

5.1.1.3 Recording and procedure. The recordings and procedures were the same as detailed in Experiment 1.

5.1.1.4 Semantic Relatedness Rating. Two sets of synonym ratings (valid for both Experiments 3a and 3b) were taken. The first goal was to ensure that the two sets of visual targets were semantically related and the second goal was to determine that the monosyllabic primes were not semantically related to the targets in (3b). Thus, we ensured that the sandhi targets like *da3sao3* "to clean" 打扫 were related to the second set of targets such as *qing1li3* "to clean up" 清理. Furthermore, the primes, particularly the first morpheme of the "sandhi" word (*da3* "to beat" 打), were not semantically related to the second set of targets *qing1li3* "to clean up" 清理). Two groups of people (in total 200) who did not participate in Experiment 3 completed the semantic ratings by filling in an online questionnaire. They were instructed to circle the number that best fit the relatedness of two items on a scale from 1 to 7, where 7 was considered to be highly semantically related and 1 was considered to be not related at all. A score of 5 or higher was accepted as being semantically related. Anything less than 4 was considered to be unrelated. We also avoided any spurious relationship between the prime and the target. For example, the target *jian3qing1* "to attenuate", which is mediated through its synonym *huan3jie3*, was discarded because *huan3* "slowly" had a high synonymous rating with the target (5.25 on a scale of 7), meaning *huan3* alone could be argued to be sufficient to activate the semantic target. In contrast, *qing1li3* "to clean up", mediated through *da3sao3* "to clean", was considered an acceptable item because *da3* "beat" had a low synonymous rating to the target (2.25). In all, six items were excluded and not considered in further analyses.

5.1.1.5 Data Screening. The data screening was the same as explained in Experiment 1 and led to a loss of 12.9% of the data (162 data points).

5.1.2. Results

5.1.2.1 Model Fitting. The LMMs analysis was performed with the fixed factor *Prime Tone* (T2, T3, T4) and *Lexicality* (word, nonword), as well as the random factors *Subject* and *Item*. Through step-wise likelihood ratio tests of the maximal random structure, a model with random slopes and intercepts between *Prime Tone* and *Item*, and random intercept for *Subject* was found to provide the best model fit .[4] Through model comparison with and without the relevant fixed factors, the main effect of *Prime Tone* (χ^2 (1) = 9.91, p = .0016*) was found to be significant.[5] According to this model, both T2 and T3 primes led to significantly faster reaction times compared to the T4 controls (Est. = −19.15, SE = 7.81, t= −2.45; p=.014*; Est. = −21.88, SE = 6.09, t= −3.59, p=.0003*). The response latencies in the T2 prime condition did not differ significantly from those of the T3 ones (Est. = 2.73, SE = 7.97, t = 0.34, p = .73). Therefore, both T2 and T3 primes could activate the mediators whose initial syllables were identical with the T3 primes (see

Table 7. Mean Response Times (in ms) for the prime-mediator word pairs in Experiment 3a.

Condition	Mean	SD	Priming effect (compared to T4)
T2	545	103	−22*
T3	542	98	−25***
T4 (control)	567	108	N/A

Note. * $p< .05$, *** $p< .001$. The negative value of priming effect represents slower response.

Table 7 and Figure 6). That is, both *da2* "to answer" and *da*3 "to beat" primed words like *da3sao3 > da [2]sao3* "to clean".

5.2. Experiment 3b

5.2.1. Method

5.2.1.1 Participants. The participants in Experiment 3b were the same as those of Experiment 3a; however, no participant was exposed to identical visual targets across the two experiments. The two experiments were counterbalanced across participants.

5.2.1.2 Materials. The primes in Experiment 3b were identical to those of 3a, but the targets while semantically related to those of Experiment 3a, were entirely different (see Figure 5).

5.2.1.3 Recording and procedure. The recordings and procedures were the same as detailed in Experiment 1.

5.2.1.4 Semantic Relatedness Rating. The semantic rating was the same as that in Experiment 3a.

5.2.1.5 Data Screening. The data screening was the same as detailed in Experiment 1 and this resulted in a loss of 13.5% of the data (170 data points).

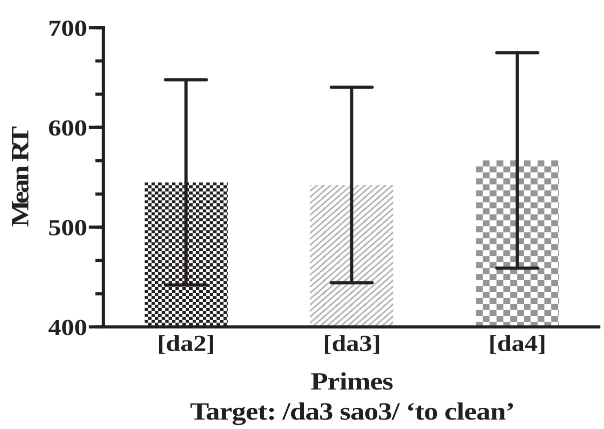

Figure 6. Mean response times for the three prime-mediator priming conditions in Experiment 3a. For example, when the mediator word was /da3 sao3/ "to clean", whose onset is identical with the T3 prime, the primes were: [da2] "to answer", [da3] "to beat" and [da4] "big".

Table 8. Mean Response Times (in ms) for the prime-target word pairs in Experiment 3b.

Condition	Mean	SD	Priming effect (compared to T4)
T2	522	96	−15*
T3	521	98	−16*
T4 (control)	537	104	N/A

Note. *$p< .05$. The negative value of priming effect represents faster response.

5.2.2. Results

5.2.2.1. *Model Fitting*. The optimal model contained *Prime Tone* as a significant fixed effect and random intercepts for *Subject* and *Item* (χ^2 (1) = 8.27, $p = 0.004$*).[6] The response time for T4 controls was significantly longer when compared with T3 primes, (Est. = 15.38, $SE = 5.54$, $t = 2.78$, p=.006*) and equally importantly, also significantly different compared to T2 primes (Est. = 13.22, $SE = 5.56$, $t = 2.38$, p=.017*) (see Table 8 and Figure 7). Thus, both T3 and T2 primes successfully facilitated the targets mediated via the semantically related sandhi words in Experiment 3a; e.g. *da2* and *da3* primed *qing1 li3* via the mediating of *da3 sao3* (see Figure 5). In addition, T2 and T3 primes did not differ significantly in facilitating the targets through T3 sandhi mediators (Est. = 2.16, $SE = 5.53$, $t = 0.39$, p=.70).

5.4. Discussion

The results of Experiments 3a and 3b indicated that the mediated semantic priming task was successful, since both T2 and T3 successfully primed the semantically mediated visual target. In Experiment 3b, although the

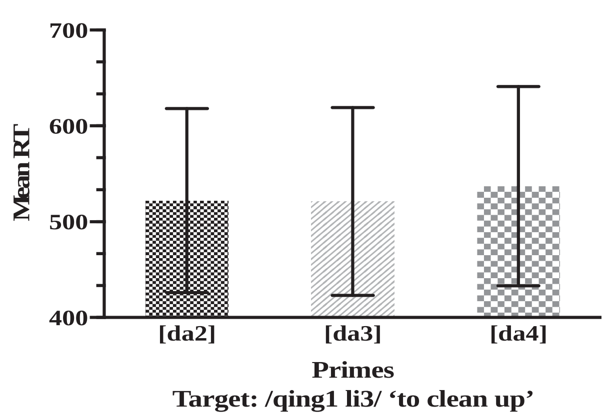

Figure 7. Mean response times for the three prime-target priming conditions in Experiment 3b. For example, when the target word was /qing1 li3/ "to clean up", which is semantically related to the mediating agent "sandhi" word /da3 sao3/ "to clean", the primes were: [da2] "to answer", [da3] "to beat" and [da4] "big".

prime-target relationship was not semantically synonymous, the T3 primes were still expected to activate a large number of compounds beginning with T3, which in turn would semantically activate other words. This design showed that, not only did T3 activate T3-T3 words and further facilitate other semantically related words, T2 also did the same (cf. Figure 5).

Thus, the T2 primes facilitated the processing of targets (when compared with T4 controls), suggesting that they could activate the mediating T3 tone sandhi words. In the correct sandhi context, T2 primes were able to facilitate the processing of targets when they semantically related to the mediating T3-T3. This is in contrast to the lack of facilitation effects when the necessary contextual information was absent (Experiment 2); i.e. T2 failed to facilitate disyllabic non-sandhi T3-TX words.

6. General discussion

The current series of studies was designed to investigate the phonological representation involving the well-known T3 tone sandhi in Mandarin where a sequence of two T3 words is disallowed and transforms the first one into T2; i.e. *T3-T3 > T[2]-T2* (cf. Yip, 1980). Although this phenomenon has been extensively studied in previous work, concerns about how T3 sandhi is represented in the mental lexicon remain controversial. While a number of psycholinguistic studies have approached this issue, two behavioural studies in particular have discussed the representational issues concerning tone sandhi: first, how are tone sandhi words represented and what is the processing mechanism involved in accessing and recognising the correct forms? Second, how should Mandarin tone sandhi words be represented in their underlying form: as *T3-T3*, as the surface form *T[2]-T3*, or with both representations?

Based on a series of experiments using auditory-auditory priming which required lexical decision on targets, Zhou and Marslen-Wilson (1997) concluded that there were two viable options for the representations of words undergoing tone sandhi – the "canonical" (i.e. underlying or *T3-T3*) and the "surface" (i.e. *T[2]-T3*)view. The "canonical" view proposed that all T3 sandhi words are represented in the citation form (i.e. *T3-T3*), despite the tone changing on the surface. In the "surface representation" view, T3 tone sandhi words are represented in their surface form, that is, *T3-T3 > T[2]-T3*. Notably, the primes and targets employed in these experiments were always disyllabic (e.g. 才华 *cai2hua2, "*talent" ~ 采取 *cai3qu3, "*to adopt"). In the second major behavioural study, Chien et al. (2016) concluded that representations of sandhi words in Mandarin are most likely stored as T3-T3, that is in their underlying forms. This series of auditory-auditory experiments, in contrast, used monosyllabic primes and disyllabic targets (e.g. *fu3* 辅, "to guide" ~ *fu3dao3* 辅导, "to counsel"). Therefore, it remains a matter of concern that Zhou and Marslen-Wilson's (1997) results covered a wide range which persuaded them to suggest that the representation of T3 should be both canonical (underlying T3) as well as including its variant T2. To further probe into the nature of the representation of T3 in the sandhi context in a behavioural study, we approached the issue by using a cross-modal paradigm both with and without overt semantic priming.

Along with other phonologists such as Yip (1980) and Duanmu (1999), our results support the view that Mandarin tone sandhi is a dissimilatory process whereby when two tones are similar, the first tone is altered. The sandhi process only affects T3 words where *T3-T3 > T[2]-T3*. Such a process suggests that for dissimilation to produce the correct result, the initial tone in the underlying representation must be specified. If the initial T3 is not specified, then it cannot change and become another tone, viz., T2. Native speakers frequently encounter and recognise such sequences and therefore must be able to access the *T3-T3* words even if they hear *T[2]-T3*. Thus, we hypothesised that the mechanism that would allow them to do so would be an instance of inferencing, by which native listeners would allow a surface *T[2]* to activate a *T3* but only in the correct context. Our experiments were geared towards investigating this issue.

Experiment 1 was a form-priming experiment, which investigated to what extent T3 and T2 monosyllabic words would prime T3-TX non-sandhi targets. Three types of auditory monosyllabic primes T2, T3, T4 (cf. *fan2 "*annoying", *fan3 "*opposite", *fan4 "*meal") were chosen for non-sandhi disyllabic targets beginning with a T3 syllable (e.g. *fan3she4 "*reflection"). The segmental content of the primes and the initial word of the target were the same. Our results revealed that, in using T4 primes as controls, T3 primes facilitated the processing of T3 non-sandhi targets while T2 did not. Thus, it appears that without the contextual information of tone sandhi, T2 cannot prime T3 if it is not the surface form derived from T3.

Next, we turned to semantic overlap between T3 primes and targets with different tones. Experiment 2 investigated whether a disyllabic (non-sandhi) target would be facilitated only by a synonymous T3 prime or whether it also be activated by a T2 competitor? Again, the primes were monosyllabic T2, T3, T4 sharing the same segments: (e.g. *nao2*, "to scratch"), T3 (*nao3,* "brain"), T4 (*nao4,* "noisy"). Disyllabic targets (which did not contain an initial T3) were chosen such that they were overtly synonymous only with the T3 primes; e.g. prime *nao3* "brain", target *tou2bu4*, 头部, "head". The targets were also chosen such that the primary meaning of the T3 primes was related to the most prominent meaning of the target. Our results revealed semantic facilitation of these targets when they were primed by synonymous T3 items (e.g. (*nao3,* "brain" ~ *tou2bu4*, 头部, "head"), but not by non-synonymous T2 primes (e.g. *nao2*, "to scratch" ~ *tou2bu4*, 头部, "head"). These findings suggest that transfer from T2 to T3 is blocked when no contextual information can be accessed. Nevertheless, although the T3 primes were chosen such that they were semantically related to the target, the fact remians that Mandarin is very polysemous and T2 initial disyllabic compounds derived from a sandhi word could have a semantic connection as well. Thus, for example, the target *nao[2]sun3* from *nao3sun3* would mean "brain damage" and could be related to the prime "brain". However, in our experiment, it was clear that only the underlying T3 monosyllabic primes facilitated its semantically related targets. To further tease apart the relationship between derived T2 and underlying T3, we designed our third experiment, this time with sandhi mediation.

Our third experiment consisted of two components: a direct semantic priming task and the main experiment involving mediated sandhi. To demonstrate that tonal sandhi is indeed dissimilation, we reasoned as follows: if both T3 and T2 activate a *T3-T3 > T[2]-T3* sandhi word, this should in turn facilitate a target semantically related to the tone sandhi word (see Figure 5 & Table 6). Thus, we expected that T3 *da3* would quite naturally activate *da3sao3* (since the first tone matches) and in turn facilitate *qing1li3*, since these two disyllabic words are semantically related. The question remained whether T2 would do the same. Since this is a sandhi context, we hypothesised that if *da2* would also activate *da3sao3* (in a sandhi context), it too would activate a semantically related counterpart. The activation of *da3sao3* by T2 would be an instance of "inferencing" or back-tracking, as hypothesised by Gaskell and Marslen-Wilson (1998). The primes in this experiment were again monosyllabic words with T2, T3, and T4. Crucially, none of the primes (e.g. *da3 "*to beat", *da2 "*to answer", *da4 "*big") were related in form or meaning to the targets (e.g. *qing1li3 "*to clean up"). However, these targets were semantically-linked via mediating T3 sandhi words (e.g. *da3sao3 "*to clean") which shared the same initial morpheme with the primes. The results in Experiment 3b showed that T2 primes indeed facilitated this processing of targets through the mediators and the processing was significantly faster when compared to the corresponding T4 controls.

It is worth mentioning that both T2 and T3 primes facilitated the processing of targets in Exp3a, which contradict the results reported by Chien et al. (2016). There are two possible reasons for the contradictory findings. Firstly, Chien et al. used an auditory-auditory form priming task and thus an inhibition could have occurred when the primes and the initial syllable of the targets were homophonous. Therefore, the failure to find a priming effect of T2 primes could be due to inhibition rather than a lack of facilitation. Secondly, the auditory targets were presented 250 ms after the offset of the primes in Chien et al.'s study. In our study, the interval was 0 ms because we used a cross-modal paradigm. The phonological encoding of tonal information takes place rapidly (compared to segmental information) and then continues to the next stage of processing (Zhou & Zhuang, 2000). Thus, the tonal activation was larger with shorter SOA while the activation of segmental information was stable across different SOAs (Nixon et al., 2015). Consequently, the contradictory results here may also be due to the different intervals used in the two studies.

Taken together, the results of all three experiments suggested that surface T2 activates T3 words **only** when appropriate sandhi context is provided. Such results cannot be fully explained by either the "surface" or the "abstract representation" hypotheses considered by Zhou and Marslen-Wilson (1997). Recall that the "surface representation" view postulates that T3 sandhi words are represented as the surface form *T[2]-T3,* which ought to allow T2 primes to facilitate lexical decision while T3 primes should not. Our results do not support this.

In Zhou and Marslen-Wilson's (1997) "abstract representation" view, it was hypothesised that if T3 is underspecified then both T2 and T3 primes would map onto the representations of sandhi words, and thus both T2 and T3 primes were expected to enhance the lexical access of sandhi words compared with T4 controls and facilitate the processing of T3 sandhi words. Their results did not meet their expectations and as we mentioned earlier, they concluded that underspecification was not an explanation for T3 sandhi. MMN studies such as Politzer-Ahles et al.'s (2016) MMN study also claimed to provide support for the underspecification view, largely because of observed asymmetries between T3 and all the other tones. That is, when T3 was the standard, smaller MMNs were found while larger MMNs were observed for T3 deviants. However, there is no a priori theoretical reason as to why one would expect T3 to be underspecified. Underspecified features often lead to asymmetry in assimilation while specified features do not. Words like *rai<u>n</u>bow* or *ha<u>n</u>d-kerchief* can be pronounced as *raimbow* and *haŋkerchief* but *gumdrop* or *kingdom* do not change to *gundrop* or *kindom*. Where the Mandarin tones are concerned, T3 is described in the traditional literature as the most complex tone: 314 compared to T1 (55), T2(35), T4 (51). Thus, T1 is a high level tone, T2 is rising, T4 is falling while T3 is a falling-rising tone. Indeed, one of the reasons for "reducing" T3 in Politzer-Ahles's study was to make it simpler with a less rising contour while maintaining the falling part of the complex tone (HL_H).

Then why the observed asymmetry in the MMN results? One hypothesis is that when a sequence of T3s are used as the standard stimuli, native listeners perceive it as atypical rather than prototypical. Under normal circumstances, disyllabic sequences of T3-T3 would become T[2]-T3 and in running speech T3-T3-T3-T3 could be produced as T[2]-T3-T[2]-T3, or perhaps if cyclic, then T[2]-T3-T[2]-T3 (Cheng, 1987; Shih, 1997, and many others). Furthermore, the question arises: if the T3 is underspecified that how were the sequence of T3s perceived by native speakers? If the array of T3s were perceived as alternating T3 and T2s, then in the pair T3/T2, the deviant T2 would presumably match on ocasion with the standard thereby reducing the MMN. The reverse does not hold: a sequence of T2s as the standard remains unchanged. If, on the other hand, the standard T3 sequences were perceived as intended (i.e. a sequence of T3), the stimuli may in fact be perceived as an atypical, possibly illegal sequence, given that the T3 sandhi process is extremely productive. We believe that the anomalous quality of the T3 standard stimuli would lead to a lower MMN no matter what the deviant was. In contrast, when other tones (i.e. T1, T2, or T4) were used as standards, the mismatch with the deviant T3 would lead to a relatively lower MMN.

Notably, the results of Politzer-Ahles et al. (2016)'s study deviated somewhat from those in Li and Chen (2016), such that in Politzer-Ahles' study, asymmetries were found when T3 was contrasted with all the other tones. In contrast, asymmetric MMNs were only found in T2 and T3 word pair in Li and Chen's study. Politzer-Ahles et al. attributed the difference to the linguistic background that Li and Chen (2015) tested a homogeneous group of speakers who came from Beijing and only spoke Mandarin, while their participants were more heterogeneous and came from various dialect backgrounds. Furthermore, the stimuli used in Li and Chen's study consisted of natural speech, while Politzer-Ahles et al.'s stimuli were manipulated to control the acoustic properties. Therefore, both the group and stimuli difference could result in the diverse results by previous studies.

We have noted that the iterative nature of T3 sandhi would be atypical for a Mandarin listener since not only is the sandhi process very common, but there can also be variations depending on the syntactic and prosodic branching (cf. Kuo, Xu & Yip, 2007; Yip, 1980; Chen, 2000). The application of T3 sandhi has been argued to be cyclic and dependent on prosodic and syntactic branching (cf. Shih, 1997). A frequently quoted example from Cheng (1987) is "Old Li buys good book" which shows that not all sequences of sandhi are permitted. In this particular sentence, only two (c) and (d) are allowed.

(1) Lao3 Li3 mai3 hao3 shu1

Old Li buy good book

a. **Lao[2] Li3 mai3 hao3 shu1*

b. *?Lao[2] Li[2] mai3 hao3 shu1*

c. *Lao[2] Li3 mai[2] hao3 shu1*

d. *Lao[2] Li[2] mai[2] hao3 shu1*

Crucially, it is possible that such a sequence of monosyllabic T3s is atypical or possibly even illegal for native listener as the standard. In fact, most studies examining atypical forms in MMN employ such sounds as deviants (cf. Näätänen et al., 1997 for Finnish/Estonian vowels, Kazanina et al., 2006 for Turkish/Russian intervocalic voiced consonants). Additionally, research on a different type of suprasegmental phenomena such as stress comes closer to that of tone; incorrect stress also is more difficult to process. For example, Friedrich et al. (2004) show that syllable fragments with incorrect stress in German hinder lexical integration. For both Finnish (Ylinen et al., 2009) and Hungarian (Hobolygó & Csepé, 2013), it has been noted that illegal stress patterns can cause substantial effects, such as eliciting double MMNs. A crucial point to note here is that, in Hungarian, two consecutive MMN components were elicited **only** when the deviant had an illegal stress pattern; when the deviant had a legal stress pattern, no such MMNs were elicited. Thus, when the standard is legal and the deviant atypical, there is greater MMN activity compared to when the standard is illegal and the deviant typical. This is reminiscent of the asymmetric MMN results in the tone sandhi research: a sequence of T3s as the standard would certainly be atypical and would render lower MMN compared to a typical standard consisting of T2s or T4s.

6.1 An excursus on "underspecification"

Although we claim that Mandarin tone sandhi does not support the notion that T3 is underspecified, it does not mean that tones cannot be underspecified.

The concept of underspecification in the psycholinguistic literature was first investigated with respect to features. A segmental phoneme is made up of features and contrastive features are usually assumed to be represented. This was the nature of the argument used in Lahiri & Marslen-Wilson (1991), who argued that nasal vowels in Bengali contrasted with oral vowels and thus the feature [NASAL] was present in the underlying representation for vowels (cf. /kãʧ/ *glass,* /kaʧ/ *wash*-IMPERATIVE). Additionally, Bengali underlying oral vowels can also become nasalised in the context of a following nasal consonant (B: /kɑn/ >[kɑ̃n] *ear*). Thus, in Bengali, vowels are not specified for nasality in the context of a nasal consonant and surface forms (e.g. [kɑ̃n]) would be derived when the nasality from the consonant spreads to the *underspecified* vowel. In Bengali, the process of nasal spreading neutralises the contrast between oral and nasal vowels, which means that on hearing a nasal vowel [ɑ̃] it is not possible (if all else remains the same) for the listener to identify whether it is an underlying nasal vowel as in (/kɑ̃ʧ/) or a derived vowel in the context of a nasal consonant ([kɑ̃n] < /kan/). Results from Lahiri and Marslen-Wilson showed that, irrespective of whether Bengali listeners heard a nasalised vowel or a real nasal vowel, they preferred words with an underlying nasal vowel /Ṽ/ over words with /VN/. Thus, in a majority of instances perceived [kɑ̃] facilitated words begin with an underlying nasal /kɑ̃/ not the surface nasalised [kɑ̃n], suggesting that the surface nasalisation was not stored in the mental lexicon.

The same logic works for place of articulation features such as [LABIAL], [CORONAL], [DORSAL] where the assumption has been that [CORONAL] is often considered to be underspecified. This gives rise to asymmetric assimilations. As mentioned above, *gunboat* can become *gumboat,* but *gumdrop* would not be pronounced as **gundrop.* Thus, universally, [CORONAL] sounds are more likely to take on the place of articulation of the following consonant but not vice versa. In terms of lexical access, it has been shown that perceiving [m] can access both /m/ and underspecified /n/ but not vice versa (cf. Lahiri & Kotzor, 2017 for a review). Thus, access follows representation, making it asymmetric.

Tones are not part of segments. Standard assumptions in autosegmental phonology (cf. Goldsmith, 1990, Leben, 1973 and etc.) allow us to adopt hierarchical structures such that tones are represented on a separate tier from phonemes. In many languages, not all morphemes are tone bearing and indeed, there can be underspecified morphemes which may acquire tones from other morphemes. A standard example of a tone spreading to an unspecified morpheme comes from Mende (Leben, 1973):

(2) HL HL H L H L

ˇ | |

/mbu/ > [mbu] /mbu + ma/ > [mbu ma]

(i) owl (ii) 'on' owl

Thus, for the bare noun *mbu* ("owl"), the floating HL tone is associated with the stem vowel /u/ leading to a falling tone HL, while in the second example *mbu-ma*, the morpheme /ma/ is underspecified for tone. Consequently, the floating tones HL are now divided between the two morphemes: the H is associated with the stem while the L tone spreads to /ma/.

Politzer-Ahles et al. (2016) also refers to another example of tonal underspecification found in languages where there is a bitonal contrast such as Norwegian and Swedish (Wetterlin & Lahiri, 2015). In these languages, it has been assumed that one of the two surface accents is unmarked and the specified accent dominates the accent of the word. For example, in Standard Norwegian, the L* is the marked specified accent which can be carried by stems and affixes. If a prefix is specified for Accent 1, then the entire word will be Accent 1. In contrast, Accent2 (H*) is assigned as a default to an unspecified trochee. These assignment rules lead to alternations such as *tál -e* > $[\text{tale}]_2$ "to talk", ^{L}*óm*$_1$ *- tal - e* > $[^{L}\text{omtale}]_1$ "to discuss"; the prefix $/\text{om}/_1$ is specified for Accent 1 and the entire word is assigned the same accent while *tale* carries the default H.

In Mandarin Chinese, however, all tones appear to be specified. If we assume the classical descriptions of the various tones in Mandarin and provide an autosegmental notation with H and L, we would have the following configurations: T1(55) = HH (level high), T2 (35) = LH (rising), T3(314) = HLH, T4(51)=HL (falling). Tonal sandhi could then be described as T3+T3, HLH + HLH > LH + HLH, leading to the account of sandhi as dissimilation. If like Yip (1980), Duanmu (1999), and Kuo et al. (2007), we assume that T3 is L, then T3 sandhi could be described as before: L + L > L HL. Either way, this reveals T3-sandhi to be instances of dissimilation. Had T3 been underspecified and thus subject to assimilation, what would it be assimilating to? Furthermore, if T3 were unspecified, then why would the first member of a sequence of two T3s be altered? An underspecified T3 could only mean that all morphemes carrying this tone would not have a tonal representation. Mandarin has a neutral tone which could be assumed to be unspecified, since its pronunciation varies considerably depending on its neighbours. Thus, a neutral tone would be lower when following T4 (HL) but it would be higher when preceded by T5 (HH), taking on the tone closest to it. This is not what happens in a sequence of T3 words.

Another option claimed in Li and Chen (2016) is that words undergoing tonal sandhi are multivariant, carrying T3 and T3 V. However, since all T3 morphemes may potentially undergo tonal sandhi, the implication is that all T3 words would be stored as having both T3 and T3 V. Such a surplus of representations certainly misses the point that a morpheme originally carrying a T3 (HLH) becomes T3 (LH) **only** under one circumstance: when another identical tonal sequence follows. Thus, the LH tonal variant is entirely context dependent. Results from our mediated priming task confirm that both T3 and T2 syllables can activate and facilitate T3 words **but only in a sandhi context**. T2 does not and cannot prime a T3 word in another context.

Thus, what appears to hold systematically for the previous studies (as well as our own) is a view similar to that of phonological inferencing, where auditory inputs are analysed in the context of phonological rules or constraints that specify the phonological contexts in which the alternation can take place (Gaskell & Marslen-Wilson, 1998). The phonological interference applies that input T2 activates the underlying representation of T3 **only** in the context of another following T3, and the re-writing rule is justified by context: hearing a T2 activates a cohort of morphemes and words with initial T2 and meanwhile, the mapping from speech signal to the underlying representation undergoes the phonological interference which re-writes the initial T2 to T3. The re-writing is only permitted when the following syllable immediately follows is also a T3. Thus all the morphemes or words with initial T3 will be activated. This also maintains the economical operation of the system: if hearing T2 activates morphemes with T3 as well as words starting with T3, the number of items that are activated would increase significantly due to homophones and thus radically reduce the efficiency of the system.

In sum, the current study was conducted to investigate the processing of Mandarin T3 sandhi where a sequence of T3-T3 surfaces as T2-T3. The results in our study suggest that the T3 must be fully specified since it itself triggers a dissimilation process. When such information is present, both T2 and T3 activate the lexical entry, whereas when relevant contextual information is absent, only T3 activates the lexical entry and T2 does not. In summation, our findings are in line with the "canonical representation" view of Zhou and Marslen-Wilson (1997) in which T3 is specified and surface variation is resolved by a re-writing rule, along with restrictions posed by phonological constraints.

Notes

1. RT~Prime Tone*Lexicality + (1+ Prime Tone |Item) + (1+ (Prime Tone*Lexicality) |Subject)
2. RT~Prime Tone + (1|Item) + (1|Subject)
3. RT ~ Prime Tone + (1 |Subject) + (1 |Item)
4. RT~Prime Tone + (1+ Prime Tone|Item) + (1|Subject)
5. RT~ Prime Tone + (1+Prime Tone|Item) + (1|Subject)
6. RT~ Prime Tone + (1|Item) + (1|Subject)

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the H2020 European Research Council [grant number 695481].

References

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, *5*(2), 183–207.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer [Computer program]. http://www. *praat. org/*

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, *5*(6), e10729. https://doi.org/10.1371/journal.pone.0010729

Chao, Y. R. (1968). *A grammar of spoken Chinese*. University of California Press.

Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects*. Cambridge University Press.

Cheng, L. L. (1987). Derived domains and Mandarin third tone sandhi. *Chicago Linguistic Society*, *23*(2), 16–29.

Chien, Y. F., Sereno, J. A., & Zhang, J. (2016). Priming the representation of Mandarin tone 3 sandhi words. *Language, Cognition and Neuroscience*, *31*(2), 179–189. https://doi.org/10.1080/23273798.2015.1064976

Cornell, S. A., Lahiri, A., & Eulitz, C. (2013). Inequality across consonantal contrasts in speech perception: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 757–772. https://doi.org/10.1037/a0030862

Duanmu, S. (1999). Metrical structure and tone: Evidence from Mandarin and shanghai. *Journal of East Asian Linguistics*, *8*(1), 1–38. https://doi.org/10.1023/A:1008353028173

Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience*, *16*(4), 577–583. https://doi.org/10.1162/089892904323057308

Friedrich, C. K., Kotz, S. A., Friederici, A. D., & Gunter, T. C. (2004). ERPs reflect lexical identification in word fragment priming. *Journal of Cognitive Neuroscience*, *16*(4), 541–552. https://doi.org/10.1162/089892904323057281

Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of*

Experimental Psychology: Human Perception and Performance, *24*(1), 145–158. https://doi.org/10.1037/0096-1523.24.2.380
Goldsmith, J. A. (1990). *Autosegmental and metrical phonology* (Vol. 1). Basil Blackwell.
Gussenhoven, C., & Jacobs, H. (2017). *Understanding phonology*. Routledge.
Honbolygó, F., & Csépe, V. (2013). Saliency or template? ERP evidence for long-term representation of word stress. *International Journal of Psychophysiology*, *87*(2), 165–172. https://doi.org/10.1016/j.ijpsycho.2012.12.005
Hyman, L. M., & Schuh, R. G. (1974). Universals of tone rules: Evidence from West Africa. *Linguistic Inquiry*, *5*(1), 81–115.
Ikeda, K., Hayashi, A., Hashimoto, S., Otomo, K., & Kanno, A. (2002). Asymmetrical mismatch negativity in humans as determined by phonetic but not physical difference. *Neuroscience Letters*, *321*(*3*), 133–136. http://doi.org/10.1016/S0304-3940(01)02408-9
Kazanina, N., Phillips, C., & Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences*, *103*(30), 11381–11386. https://doi.org/10.1073/pnas.0604821103
Kuo, Y. C., Xu, Y., & Yip, M. (2007). The phonetics and phonology of apparent cases of iterative tonal change in standard Chinese. In C. Gussenhoven, & T. Riad (Eds.), *Tones and tunes, Vol. 2: Experimental studies in word and sentence prosody* (pp. 211–237). Mouton de Gruyter.
Lahiri, A., & Kotzor, S. (Eds.). (2017). *The speech processing lexicon: Neurocognitive and behavioural approaches* (Vol. 22). De Gruyter Mouton.
Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, *38*(3), 245–294. https://doi.org/10.1016/0010-0277(91)90008-R
Leben, W. R. (1973). *Suprasegmental phonology.* (Unpublished doctoral dissertation), MIT, Cambridge, MA.
Lee, C. Y. (2007). Does horse activate mother? Processing lexical tone in form priming. *Language and Speech*, *50*(1), 101–123. https://doi.org/10.1177/00238309070500010501
Li, Q., & Chen, Y. (2016). An acoustic study of contextual tonal variation in tianjin mandarin. *Journal of Phonetics*, *54*, 123–150. https://doi.org/10.1016/j.wocn.2015.10.002
Li, X., & Chen, Y. (2015). Representation and processing of lexical tone and tonal variants: Evidence from the mismatch negativity. *PloS one*, *10*(12), e0143097. https://doi.org/10.1371/journal.pone.0143097
Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., … Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, *385*(6615), 432–434. https://doi.org/10.1038/385432a0
Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026
Nixon, J. S., Chen, Y., & Schiller, N. O. (2015). Multi-level processing of phonetic variants in speech production and visual word processing: Evidence from Mandarin lexical tones. *Language, Cognition and Neuroscience*, *30*(5), 491–505. https://doi.org/10.1080/23273798.2014.942326
Politzer-Ahles, S., Schluter, K., Wu, K., & Almeida, D. (2016). Asymmetries in the perception of Mandarin tones: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(10), 1547–1570. https://doi.org/10.1037/xhp0000242
Reetz, H., & Kleinmann, A. (2003, August 3–9). *Multi-subject hardware for experiment control and precise reaction time measurement*. Proceedings of the 15th International congress of phonetic sciences (pp. 1489–1492), Barcelona, Spain.
Sereno, J. A., & Lee, H. (2015). The contribution of segmental and tonal information in Mandarin spoken word processing. *Language and Speech*, *58*(2), 131–151. https://doi.org/10.1177/0023830914522956
Shih, C. (1997). Mandarin third tone sandhi and prosodic structure. In J. Wang, & N. Smith (Eds.), *Studies in Chinese phonology* (pp. 81–123). Mouton de Gruyter.
State Language Commission of China. (1988). *Xian Dai Han Yu Zi Biao* [List of frequently used modern Chinese characters]. Language Press.
Steriade, D. (1987). Redundant values. *CLS*, *23*(2), 339–362.
Wetterlin, A., & Lahiri, A. (2013). The diachronic development of stød and tonal accent in North Germanic. In *Historical Linguistics 2013*, 53–68. John Benjamins.
Xinhua Zidian (10th ed.). [Xinhua Dictionary]. (2004). Beijing, China: Shang wu yin shu guan.
Xu, Y. (1993). *Contextual tonal variations in Mandarin Chinese* (Unpublished doctoral dissertation), The University of Connecticut.
Yip, M. (1980). *The tonal phonology of Chinese* (Unpublished doctoral dissertation), Massachusetts Institute of Technology.
Ylinen, S., Strelnikov, K., Huotilainen, M., & Näätänen, R. (2009). Effects of prosodic familiarity on the automatic processing of words in the human brain. *International Journal of Psychophysiology*, *73*(3), 362–368. https://doi.org/10.1016/j.ijpsycho.2009.05.013
Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, *27*(1), 153–201. https://doi.org/10.1017/S0952675710000060
Zhou, X., & Marslen-Wilson, W. D. (1997). The abstractness of phonological representation in the Chinese mental lexicon. In H. C. Chen (Ed.), *Cognitive processing of Chinese and related Asian languages* (pp. 3–27). The Chinese University Press.
Zhou, X., Qu, Y., Shu, H., Gaskell, G., & Marslen-Wilson, W. D. (2004). Constraints of lexical tone on semantic activation in Chinese spoken word recognition. *Acta Psychologica Sinica*, *36*((04|4)), 379–392.
Zhou, X., & Zhuang, J. (2000). Lexical tone in the speech production of Chinese words. In *Sixth International Conference on Spoken Language Processing*. Retrieved from http://www.isca-speech.org/archive/icslp_2000/i00_2051.html

Appendix

Appendix A. Stimuli used in experiments 1.

Prime syllable	Target	Prime syllable	Target
zou	走廊	xiang	想念
jiang	讲座	da	打仗
si	死刑	qing	请教
jin	紧急	lian	脸色
kou	口音	fan	反射
guan	管道	yang	养殖
gai	改进	qiang	抢救
guang	广播	chan	产业
ding	顶峰	qi	起床
kao	考察	chang	厂长
zi	子弹	niu	扭曲
zheng	整洁	hui	毁灭
gui	鬼神	fu	辅助
kong	孔雀	du	赌博
dan	胆量	wan	挽回
gun	滚动	nao	脑袋
zhen	诊断	ting	挺拔
zong	总结	chao	吵架

Appendix B. Stimuli used in experiments 2.

Prime syllable	Target	Prime syllable	Target
zou	步行	xiang	思索
jiang	说话	da	攻击
si	逝世	qing	恳求
jin	只是	lian	面孔
kou	嘴巴	fan	颠倒
guan	负责	yang	培育
gai	变更	qiang	夺取
guang	宽大	chan	制造
ding	尖端	chuan	呼吸
kao	测验	sha	愚蠢
shua	玩弄	niu	摇摆
dou	振动	hui	破坏
kun	缠绕	fei	强盗
shang	奖励	du	阻塞
dan	勇气	she	放弃
gun	旋转	nao	头部
zhen	治疗	ting	相当
gao	从事	chao	争执

Appendix C. Stimuli used in experiments 3.

Prime syllable	Mediator	Target	Prime syllable	Mediator	Target
tu	土匪	强盗	kao	考古	坟墓
chao	炒股	证券	zhen	诊所	保健
ke	可口	美味	ou	偶尔	不时
ju	举止	行为	fan	反感	讨厌
ying	影响	打扰	lao	老鼠	耗子
wa	瓦解	粉碎	liao	了解	懂得
ben	本领	能力	ya	雅典	希腊
cai	采取	运用	yi	以往	从前
niu	扭转	改变	zong	总理	首相
qi	起码	至少	nao	脑海	往事
wang	往往	总是	da	打扫	清理
gao	搞鬼	惹祸	suo	所以	因此
fu	辅导	学习	fan	反响	回应
xiao	小品	相声	chu	处理	案件
ling	领导	请示	zou	走狗	帮凶