

JOURNAL ARTICLE

The representation of variable tone sandhi patterns in Shanghai Wu

Hanbo Yan¹, Yu-Fu Chien^{2,3} and Jie Zhang⁴

¹ School of Chinese Studies and Exchange, Shanghai International Studies University, CN

² Department of Chinese Language and Literature, Fudan University, CN

³ Department of Modern Languages, DePaul University, US

⁴ Department of Linguistics, The University of Kansas, US

Corresponding author: Yu-Fu Chien (chien_yc@fudan.edu.cn)

Disyllabic verb-noun (V-N) items in Shanghai Wu have variable surface tone patterns: They can undergo either a rightward extension tone sandhi, which extends the lexical tone of the first syllable over the entire word, or tonal reduction on the first syllable. The current study investigates how the phonological properties of these alternation processes as well as variation influence how Shanghai speakers represent and access such words. We conducted an auditory-auditory priming lexical decision experiment on Shanghai V-N items that can undergo either tonal extension or tonal reduction with native Shanghai speakers. Each disyllabic target was preceded by monosyllabic primes with the canonical tone, the tonal-extension tone, the surface tone, or a tone unrelated to the tone of the first syllable of the targets. Results showed both canonical and tonal-extension priming effects, but no surface priming effect. Moreover, although more familiar V-Ns were recognized with shorter reaction time, the priming effect did not interact with speakers' familiarity ratings or sandhi preference ratings of the targets. These data are consistent with the interpretation that both the canonical and tonal-extension forms are represented in Shanghai speakers' mental lexicon due to tone sandhi variation, but the representation does not seem to be modulated by the frequencies of the variants. Also, together with findings from auditory priming studies of other tone sandhi patterns, the current study suggests that certain phonological properties of an alternation, such as its locality and transparency, influence the representation of words undergoing the alternation; but whether the alternation is structure-preserving does not seem to impact the representation.

Keywords: Priming; spoken word recognition; tone sandhi; phonological variation; Shanghai Wu

1. Introduction

Variation in phonological patterns has received increasingly close attention from both modern phonological theories and models of spoken word recognition. Variation patterns in American English *t/d*-deletion (Coetsee, 2004; Guy, 1994; Labov, 1989), English schwa deletion (Patterson, LoCasto, & Connine, 2003), forms of reduplication in Ilokano (Coetsee, 2006; Boersma & Hayes, 2001; Hayes & Abad, 1989), and Japanese geminate devoicing (Coetsee & Kawahara, 2012), for instance, have been investigated both empirically and formally in the phonology literature. Of particular interest here is how listeners process phonological variability during word recognition (e.g., Connine, 1994, 2004; Connine, Ranbom, & Patterson, 2008; Deelman & Connine, 2001; Johnson, 2004; Luce, MacLennan, & Charles-Luce, 2001; MacLennan, Luce, & Charles-Luce, 2003; Sumner & Samuel, 2005). Most of the findings so far, however, have come from segmental processes that can also be characterized as reduction. We complement this line of research by investigating a

morpho-syntactically conditioned variation pattern in tonal alternation (tone sandhi) in the Shanghai Wu dialect of Chinese.

Shanghai tone sandhi has been reported to have two different patterns—tonal extension or tonal reduction—depending on the properties of syntactic structure, semantic transparency, and usage frequency: Some lexical items can only undergo tonal extension, some can only undergo tonal reduction, and some can have both forms (Xu, Tang, & Qian, 1981, 1982; Yan, 2018). Not only will the phonological alternation, in this case, tone sandhi, lead to a mismatch between surface and stored representations (Chien, Sereno, & Zhang, 2016; Nixon, Chen, & Schiller, 2015), but the variation pattern may also make the recognition process more challenging. The question we address in this study is how listeners decode the variable acoustic sandhi forms and map them onto the stored representation during spoken word recognition. We address this question using an auditory-auditory priming lexical decision paradigm.

1.1 Tone sandhi in Shanghai Wu

Shanghai Wu belongs to the Taihu sub-dialect area of Northern Wu (Fu, Cai, Bao, Fang, Fu, & Zhengzhang, 1986; Xu & You, 1984). The Shanghai Wu variety under investigation here is the main dialect spoken in the urban districts of Shanghai. Syllables in Shanghai can be non-checked (open or sonorant-closed) or checked (closed by a glottal stop [ʔ]), with tones 53, 24, and 13 occurring on the former and tones 55 and 13 occurring on the latter.¹ Shanghai has maintained the historical voicing distinction on the initial consonants, and this voicing distinction has cooccurrence restrictions with the tones in the language: High register tones, tones 53, 24, and 55, only occur in syllables beginning with voiceless consonants (the historical *yin* tones), and low register tones, tones 13 and 13, only occur in those starting with voiced ones (the historical *yang* tones). For the phonetic realization of this voicing contrast, see Chen (2011), Gao and Hallé (2017), Tian and Kuang (2019), and Zhang and Yan (2018).

Most of the lexicalized items in Shanghai undergo the rightward tonal extension tone sandhi (Xu et al., 1981), which extends the tone on the first syllable across an entire word, as shown on the left side of **Table 1**. According to Xu et al. (1981), modifier-noun compounds and a portion of the verb-noun, verb-modifier, subject-predicate, and coordinate compounds can *only* undergo rightward tonal extension. We refer to these items as *Compound-type* items.

Table 1: Shanghai disyllabic tone sandhi ('X' refers to any tone in the tonal inventory).

Rightward tonal extension		Tonal reduction ²
$\sigma_2 = \text{CV or CVN}$	$\sigma_2 = \text{CV?}$	
53-X → 55-31	53-X → 55- <u>22</u>	53-X → 44-X
24-X → 33-44	24-X → 33- <u>44</u>	24-X → 44-X
13-X → 22-44	13-X → 22- <u>44</u>	13-X → 33-X
<u>55</u> -X → <u>33</u> -44	<u>55</u> -X → <u>33</u> - <u>44</u>	<u>55</u> -X → <u>44</u> -X
<u>13</u> -X → <u>11</u> -13	<u>13</u> -X → <u>11</u> - <u>33</u>	<u>13</u> -X → <u>22</u> -X

¹ This article uses Chao's (1930) system of tone numbers to indicate pitch levels, with '1' as the lowest pitch and '5' the highest pitch. Tones on checked syllables are underlined.

² Xu et al. (1981) described this pattern as a second type of tone sandhi. However, Takahashi (2013) and Zhang and Meng (2016) both showed that for this type of sandhi, the tones on the first syllable maintained the pitch properties of the canonical tones and claimed that instead of undergoing sandhi, the forms undergo phonetic contour flattening. Whether it is a phonological process or a phonetic flattening is not the focus of the current study. We refer to this pattern as 'tonal reduction' in this paper.

Non-lexicalized items undergo tonal reduction, in which the final syllable in the sandhi domain keeps the canonical tone, while the preceding syllables are reduced to a level tone as shown on the right side of **Table 1**. These can be referred to as *Phrase-type* items, and some verb-noun and subject-predicate structures fall under this type (Xu et al., 1981).

The majority of verb-noun, verb-modifier, subject-predicate, coordinate, and adverb-adjective structures can undergo *either* rightward tonal extension or tonal reduction. We refer to these items as the *either-Compound-or-Phrase-type*. For example, when /ts^hã24 ku53/ ‘to sing + song’ means ‘to sing’ and is used as a compound, it undergoes rightward tonal spreading, [ts^hã33 ku44]; but when it means ‘to sing a particular song’ and is used as a phrase, it undergoes tonal reduction, [ts^hã44 ku53].

According to Xu et al. (1981), whether a verb-noun or subject-predicate structure belongs to the *Phrase-type* or the *either-Compound-or-Phrase-type* depends on its lexical frequency and semantic transparency (the degree of connection between the components). For example, in **Table 2**, when the verb /baʔ13/ ‘to pull’ is combined with three different nouns: ‘river,’ ‘grass,’ and ‘tree,’ the sandhi applications are different. The compound /baʔ13 u13/ ‘to pull + river,’ a common, semantically-opaque idiomatic expression, only undergoes rightward tonal extension, while /baʔ13 zi13/ ‘to pull + tree,’ a semantically transparent phrase, undergoes tonal reduction. The phrase /baʔ13 ts^hɔ24/ ‘to pull + grass,’ a phrase with an intermediate degree of semantic connection between the two morphemes, can either undergo rightward spreading sandhi or tonal reduction.

Some items can belong to both the *Compound-type* and the *either-Compound-or-Phrase-type* due to their multiple meanings. For example, when /ts^hɔ24 vɛ13/ ‘to fry + rice’ means ‘fried rice’ as a compound noun, it can only undergo extension sandhi [ts^hɔ33 vɛ44]; when it means ‘to fry rice’ as a verb phrase, it can undergo extension sandhi or tonal reduction [ts^hɔ44 vɛ13].

Yan (2018) further investigated the factors that influence the variation pattern of disyllabic tone sandhi in Shanghai. The study reported a goodness rating experiment for the variant forms with native Shanghai speakers in tandem with semantic transparency and subjective frequency ratings from the same speakers. The results showed that the preference for disyllabic tone sandhi application in Shanghai was primarily determined by syntactic structure. Speakers preferred rightward tonal extension for modifier-noun (M-N) compounds and tonal reduction for verb-noun (V-N) items. Semantic transparency had no effect on sandhi preference for M-Ns, but did have an influence for V-Ns, in that speakers preferred the tonal extension form for more semantically-opaque V-Ns. More interestingly, the nature of the frequency effect on sandhi preference interacted with syntactic structure. Participants preferred the tonal extension form for more frequent M-Ns, but the reduction form for more frequent V-Ns. Yan’s interpretation was that M-Ns are words in Shanghai, and the extension sandhi, which reduces the number of underlying tones, applies more readily to higher frequency items, comparable to the relation between

Table 2: The variable application of rightward tonal extension in Shanghai.

Examples	Rightward tonal extension	Tonal reduction
/baʔ13 u13/ ‘tug of war’ to pull + river	[baʔ11 u13]	
/baʔ13 ts ^h ɔ24/ ‘to weed’ to pull + grass	[baʔ11 ts ^h ɔ13]	[baʔ22 ts ^h ɔ24]
/baʔ13 zi13/ ‘to pull out a tree’ to pull + tree		[baʔ22 zi13]

word frequency and phonological reduction processes attested elsewhere (e.g., vowel deletion in English (Hooper, 1976), t/d deletion in English (Coetzee & Kawahara, 2012), and geminate devoicing in Japanese (Coetzee & Kawahara, 2012). V-N items, on the other hand, are phrases and there is a paradigmatic correspondence between the tone on the verb of V-N and the tone of the verb syllable used elsewhere. Given that paradigmatic leveling towards regularization is more likely to occur in less frequent items (Bybee, 2001), and that the tonal extension form (rather than the tonal reduction form) should be considered the regular form in Shanghai tone sandhi due to its wider application contexts, less frequent V-N items are then more likely to undergo the tonal extension form.

1.2 The effects of phonological variation and tone sandhi alternation on spoken word recognition

Theories of spoken word recognition differ in the degree of abstraction posited in the mental lexicon. One view is that the mental lexicon contains abstract linguistic features so that surface acoustic inputs have to be computed before being mapped onto the stored representations (Gaskell & Marslen-Wilson, 2002; Marslen-Wilson & Warren, 1994; Norris, McQueen, & Cutler, 2000). Another view is that the mental lexicon consists of rich episodic information to which surface acoustic inputs are mapped relatively directly (Clarke, 2003; Goldinger, 1998; Norris, McQueen, & Cutler, 2003; Pierrehumbert, 2001, 2003). Variant forms resulted from phonological processes provide further challenges for theories of spoken word recognition.

1.2.1 The effects of phonological variation

Most of the earlier studies mainly focused on variation of phonetic reduction processes whose application is closely tied to speech register. For instance, regarding word recognition of variant forms, Deelman and Connine (2001) investigated American English word-final t/d deletion, due to which a word-final alveolar stop can be articulated with or without a release burst. In a phoneme monitoring experiment, reaction times to words ending in voiced or voiceless alveolar stops were measured. The presented words either had a final release or not. Results showed that, consistent with corpus frequency counts that 59% of the alveolar stops had a release (Crystal & House, 1988), release-bearing tokens were responded to more quickly than no-release ones. In a study that investigated the representation of words with variable schwa deletion, Connine et al. (2008) asked listeners to indicate whether a schwa was present or absent for tokens that were manipulated in schwa duration. The results showed that words undergoing schwa deletion with a low rate (below 50%) were judged as schwa-present more often than those undergoing a high rate (above 50%) of schwa deletion. That is to say, speakers detected a schwa in a word with a rate that correlated with the frequency of the schwa-present variant in real speech. The effect of variant frequency on the recognition of phonological variant forms was found in studies of American English word medial flap (Connine, 2004; Patterson & Connine, 2001) and nasal flap (Ranbom & Connine, 2004). For example, Ranbom and Connine (2004) investigated the effect of variable nasal flapping in American English (e.g., *twenty* can be realized as either [t^hwentɪ] or [t^hwenɾɪ], with the nasal flap occurring in nearly 82% of the productions of like words based on an analysis of the Switchboard database) on lexical access and found that words with greater than 50% occurrence of nasal flaps were responded to faster and more accurately than those with infrequent nasal flaps in a lexical decision task. Based on these results, Connine and Pinnow (2006) and Ranbom and Connine (2007) suggested that both forms of a phonological alternation contribute to lexical access, and frequency of occurrence of the phonological variants may influence how strongly a particular variant is activated. They argued for a hybrid

model in which multiple phonological variant forms of a given word may be stored in the lexical representation. However, Sumner and Samuel (2005) investigated the three final /t/ variants ([t] versus [ʔt] versus [ʔ]) in American English, and found that all regular variant forms were equally effective in activating lexical representations in the short term semantic priming paradigm, but the canonical form (basic [t] form) had a stronger advantage in the long term. The frequency of a certain phonetic variant seemed to be irrelevant to immediate activation at the semantic level.

The processing of morpho-syntactically conditioned phonological variation has received considerably less attention. Under the generative theory of phonology, speech register-based processes are typically considered to apply in a later component of phonology than morpho-syntactically conditioned processes (see Coetzee & Pater, 2011). In the current study, we examine how the variant forms due to the latter type of phonological alternation are represented and recognized using tone sandhi in Shanghai Wu as a test case.

1.2.2 The effects of tone sandhi alternation

Regarding how lexical items with tonal alternation are stored in the mental lexicon, Zhou and Marslen-Wilson (1997) proposed three representational views. The surface representation view states that tone sandhi words are represented based on the surface form, while the canonical representation view suggests that the tone sandhi words are represented as the combination of citation forms of their constituent morphemes. The latter view is more consistent with the assumption of traditional generative phonology that the surface form of a tone sandhi word is derived from an underlying representation using a tone sandhi rule (Chen, 1987, 2000; Shih, 1997). The third view, the abstract representation view, assumes that lexical representations of tone sandhi words are underspecified—a view that Zhou and Marslen-Wilson (1997) eventually rejected.

Recent psycholinguistic and neurolinguistic studies have used different experimental tasks to further investigate the tonal representation of words undergoing tone sandhi in Mandarin Chinese (Chien et al., 2016; Chien, Yang, Fiorentino, & Sereno, 2020), Tianjin Chinese (Li, 2016), Taiwanese Southern Min (Chien, Sereno, & Zhang, 2017), and Shanghai Wu (Yan, Chien, & Zhang, 2020). These studies showed that the representation of the sandhi words accessed in word recognition is affected by a number of phonological properties of the tone sandhi pattern. These properties include whether the tone sandhi pattern is local (whether the sandhi tone occurs within a syllable), structure-preserving (whether the sandhi tone is another isolated tone in the tonal inventory of the language), and phonologically opaque.³

The most widely studied tone sandhi pattern in this literature is the Mandarin tone 3 sandhi, which turns a tone 3 into tone 2 before another tone 3 (T3 → T2 / _ T3). This is a local, structure-preserving, and phonologically transparent sandhi in that the tone sandhi occurs on one syllable, the sandhi tone is a tone in the inventory of Mandarin, and there is an exceptionless phonotactic generalization that motivates the sandhi (*T3-T3). This sandhi pattern has been shown to be highly productive in novel words (Zhang & Lai, 2010; Zhang & Peng, 2013). Chien et al. (2016) used an auditory-auditory priming lexical decision task for T3 sandhi words in Mandarin and found that the underlying tone 3 had a facilitation effect on the recognition of words undergoing tone 3 sandhi regardless of word frequency, but the surface tone 2 did not show priming effects. These results suggest that tone 3 sandhi words in Mandarin are represented in their underlying T3-T3. This

³ A phonological rule $P, A \rightarrow B / C_D$, is opaque if the surface structures are any of the following: (a) instance of A in the C_D environment, or (b) instance of B derived by P in environments other than C_D (Kiparsky, 1973).

finds further support in an MMN experiment by Chien et al. (2020), which showed that the presence of T3 sandhi in standards (disyllabic T3-T3 words) affected the occurrence of MMNs in monosyllabic T2 deviants, indicating that the underlying T3 exerts an influence in the passive listening of T3 sandhi words.

The cognate of the Mandarin T3 sandhi in Tianjin Chinese, a dialect closely related to Mandarin, was investigated by Li (2016). Tianjin Chinese also has a four-tone inventory, with the four tones corresponding to the four tones in Mandarin, but with different acoustic realizations. The T3 sandhi $T3 \rightarrow T2 / _ T3$ applies in Tianjin as well. Using a visual world paradigm with an auditory word recognition task, Li (2016) found that the underlying tone (tone 3) of the target words was activated at an earlier stage (200-400 ms), while the surface tonal contour (tone 2) had a facilitation effect at a later stage (500-700 ms). Similar to the findings in Mandarin, this result suggests that T3-T3 is at least part of the representation of words undergoing the local, structure-preserving, and phonologically transparent tone 3 sandhi, and that this representation is accessed first in spoken word recognition. But the later activation of tone 2 indicates that the surface representation may be relevant as well.

The effect of opacity was observed in the comparison between Mandarin tone 3 sandhi and tone sandhi patterns in Taiwanese Southern Min. Four of five tones in Taiwanese Southern Min are involved in a circular chain shift in non-phrase-final positions: $51 \rightarrow 55 \rightarrow 33 \rightarrow 21 \rightarrow 51 / _ X$ (Cheng, 1968). Although these sandhi patterns are still local and structure-preserving, they are opaque in that they cannot be motivated by phonotactic generalizations (e.g., 51 is a legal tone in non-phrase-final positions). Zhang, Lai, and Sailor (2011) showed that the sandhi patterns in the circular chain shift are largely unproductive when tested with novel reduplications. Chien et al. (2017) used a similar auditory-auditory priming lexical decision paradigm to Chien et al. (2016) for Taiwanese words undergoing the $51 \rightarrow 55$ sandhi and found that the surface tone 55 led to a significantly stronger facilitation effect on word recognition than the underlying tone 51. These results suggest that, for a phonologically opaque tone sandhi pattern that lacks phonotactic motivation, the sandhi words are represented with their surface tones.

The tonal extension tone sandhi pattern in Shanghai Wu differs from both Mandarin and Taiwanese Southern Min tone sandhi in that it is non-local and non-structure-preserving. This is because the sandhi spreads the tone on the first syllable over multiple syllables in the word, and the resulting tones on each syllable do not correspond to existing citation tones in the tonal inventory of Shanghai. The sandhi pattern is transparent, however, as it can be motivated by the ban on pronounced contour tones on a single syllable (Zhang, 2007; Zhang & Meng, 2016). Yan et al. (2020) examined the *Compound-type* M-N words in Shanghai Wu, which obligatorily⁴ undergo the rightward tonal extension sandhi, using a similar auditory-auditory priming lexical decision paradigm. They found that, despite the phonological transparency of the sandhi pattern, canonical tone primes (underlying tone from the first syllable of disyllabic M-N words) failed to elicit a facilitation effect for the sandhi words for either younger speakers (infrequent Shanghai users) or older speakers (frequent Shanghai users), while surface extension tone primes facilitated word recognition of the tone sandhi words for infrequent users, but not frequent users. They proposed that due to the non-local nature of the rightward tonal extension pattern, whereby a legal disyllabic sandhi melody can only be formed by combining two non-existing surface tones together, the sandhi words should be facilitated (phonetically)

⁴ Yan (2018) used a sandhi-goodness-rating paradigm, so it was possible for the participants to provide the rating of 'I sometimes say it in this way' for the tonal reduction form for M-N words. However, they consistently rated the tonal extension form as more likely.

by the surface tone primes, but not by the canonical (underlying) tone primes. For the lack of surface tone facilitation for the frequent users, their interpretation was that the phonetic priming effect was canceled out by a lexical competition effect: Since more frequent users of the Shanghai dialect would have a larger semantic knowledge base, more words having the same surface forms as the surface primes and the first syllable of sandhi targets were activated and competed for lexical access, exhibiting a stronger lexical competition effect; for infrequent users, due to their smaller semantic knowledge base, there is less lexical competition, and this allowed the emergence of the facilitation effect at the phonetic level. But what the authors were not able to tease apart was whether the priming pattern (lack of canonical tone priming, limited surface tone priming) was due to the non-local or non-structure-preserving aspect of the sandhi, as these properties are conflated in the rightward tonal extension pattern in Shanghai. It is possible that non-structure-preservation, by which a non-lexical tone occurs as the sandhi tone, leads to the absence of canonical tone facilitation.

Taken together, these previous studies focused on how words undergoing different sandhi rules are represented and processed in spoken word recognition. For local and structure-preserving tone sandhi, opacity plays a key role in the representation of sandhi words, with phonologically transparent Mandarin and Tianjin tone 3 sandhi words ([+local, +structure-preserving, +transparent]) being represented in the canonical form, and the opaque Taiwanese 51 → 55 sandhi words ([+local, +structure-preserving, -transparent]) being represented mainly in the surface form. For a [-local, -structure-preserving, +transparent] sandhi pattern like rightward tonal extension in Shanghai, sandhi words are not represented in their canonical underlying forms, but whether this is due to the non-local or non-structure-preserving property of the sandhi remains to be seen. In order to tease apart the contribution of locality and structure-preservation to the representation of tone sandhi words, the current study investigates the tonal reduction pattern in *Phrase-type* items in Shanghai ([+local, -structure-preserving, +transparent]). A comparison between Mandarin tone 3 sandhi and the tonal reduction pattern in *Phrase-type* items in Shanghai will allow us to isolate the effect of structure-preservation on the representation of tone sandhi words. In addition, comparing the rightward tonal extension sandhi in *Compound-type* items with the tonal reduction sandhi in *Phrase-type* items in Shanghai will allow us to better understand the effect of locality on the representation of tone sandhi words.

1.3 Current study

Given that for V-N items in Shanghai, rightward tonal extension and tonal reduction can apply variably in a lexically specific manner, the pattern provides us with an opportunity to investigate the representation of words undergoing a variable tonal process, as well as the effects of locality and structure-preservation on the processing of spoken words undergoing tone sandhi. In this paper, we investigated the representation and processing of these V-N items using an auditory-auditory priming lexical decision paradigm, with V-N disyllables in tonal reduction form as the targets,⁵ preceded by a canonical tone prime, an extension tone prime, a reduction surface tone prime, and an unrelated tone prime. The canonical tone prime had the same tone as the first syllable in citation form. The tonal-extension tone prime shared the same tone with the first syllable of the tonal extension form. The reduction surface tone prime shared the same tone with the first syllable of the tonal reduction form. The unrelated tone prime was not related in any way to the first

⁵ The speaker of the current study produced all V-N items in the experiment in tonal reduction form during the recording. See more details in Section 2.2.

syllable of targets in tone. All primes shared the same segments with the first syllable of the target. Based on the [+local, -structure-preserving, +transparent] properties of tonal reduction in Shanghai as well as the variation between tonal reduction and rightward tonal extension for V-N items, various predictions can be made, and the corroborations of these predictions have different theoretical implications.

First, given that Shanghai tonal reduction only differs from Mandarin tone 3 sandhi in structure preservation, if it elicits a similar priming effect to Mandarin, i.e., if the canonical tone prime, but not the reduction surface tone prime, elicits shorter reaction times during lexical decision, then it will suggest that structure preservation does not play a significant role in the representation of tone sandhi words. As long as the tone sandhi pattern is local and transparent, tone sandhi words are represented in the canonical form regardless of whether the sandhi is structure-preserving or not. This will in turn allow us to tease apart the effects of locality and structure preservation in Shanghai M-N tonal extension (Yan et al., 2020) and provide evidence that the surface priming effect found in Shanghai M-N tonal extension words with infrequent users was due to locality, not to structure preservation. But if the reduction surface prime facilitates V-N recognition, then it will indicate that non-structure-preservation leads to the surface priming, and that the surface priming effect observed in Shanghai tonal extension sandhi in M-Ns for infrequent users may have been caused by the feature of non-structure-preservation.

Second, given that variation between tonal reduction and rightward tonal extension for V-N items has been reported (Xu et al., 1981), and that both tonal reduction and tonal extension forms in V-Ns were acceptable by native speakers (Yan, 2018), according to the hybrid model of lexical representation proposed by Connine and Pinnow (2006) and Ranbom and Connine (2007), both variant forms may be stored. Based on Yan et al.'s (2020) finding, we expect the extension form to be represented in its surface form; the reduction form is hypothesized to be represented in its canonical form (if non-structure-preservation does not influence the canonical representation). These will lead to the hypotheses that both the extension prime and the canonical prime will elicit facilitative priming effects.

Third, based on Yan's (2018) finding that the frequency of V-N items influences Shanghai speakers' preference for sandhi application, in that less frequent ones are more likely to undergo tonal extension, as well as the relevance of frequency in the representation of variable segmental processes (e.g., Connine, 2004; Connine & Pinnow, 2006; Deelman & Connie, 2001; Ranbom & Connine, 2007), we also predict that the priming effect may be modulated by the familiarity rating and sandhi preference rating of the V-N items, in that the canonical primes may facilitate the recognition of more frequent V-N items that are more likely to have tonal reduction, while the extension tone prime may have a facilitation effect for less frequent V-N items that tend to undergo tonal extension.

2 Methodology

An auditory lexical decision experiment with auditory priming was conducted with monosyllabic primes and disyllabic Shanghai V-N targets that may undergo either rightward tonal extension or tonal reduction.

2.1 Participants

Forty native speakers of Shanghai (33 females, 7 males) participated in the experiment. Thirty-nine of them were born and raised in urban districts of Shanghai. The remaining one was born in a suburban area of Shanghai, but lived in the Shanghai urban area since the age of 12. She also had a native Shanghai-speaking parent, and spoke Shanghai at home as a child. Participants ranged from 41 to 65 years old, with an average age of 54,

and they all lived in Shanghai at the time of the experiment. According to self-reports, on average, the participants in the current experiment use Shanghai 71% of the time in their daily lives.

Three of them (2 females and 1 male) were excluded from the analysis because they claimed that they did not fully understand the task of the priming experiment afterwards. All participants were paid 50 RMB (about 8 USD) for their participation.

2.2 Stimuli

Twenty disyllabic V-N items were selected from the *Shanghai Dialect Dictionary* (Li, Xu, & Tao, 1997) as critical targets (see Appendix A). Tone 24 was chosen as the canonical tone of the first syllable to limit the number of level tone primes in the experiment. The second syllable was always tone 13 underlyingly to avoid any influence from repeating the unrelated tone prime. For the 24 + 13 targets, the canonical tone prime was tone 24, the extension tone prime was tone 33, the surface tone prime was tone 44, and the unrelated one was 53. Examples of four prime types for critical targets in different tonal combinations are listed in **Table 3**.

All stimuli were recorded by a 36-year-old female native speaker in a quiet room in downtown Shanghai, using a TASCAM DR-100MKIII recorder and an Electro-Voice 767 microphone with a 22,050 Hz sampling rate. Given the variable sandhi pattern of V-N, the speaker was asked to produce the disyllables as naturally and comfortably as possible. She used the tonal reduction form for all V-N critical targets, which was consistent with Yan's (2018) finding that Shanghai speakers preferred to apply tonal reduction for V-N items. Additionally, in an earlier pilot study, we found that when listeners heard the rightward tonal extension form of V-N targets, they tended to judge them as nonwords (53%). Hence, the tonal reduction form was used for all the V-N targets in the current study. Due to the nature of the extension sandhi and reduction patterns, the extension and surface tone prime do not exist in Shanghai real monosyllables, but the other two types of monosyllabic primes are both real Shanghai words. The canonical primes, control primes, and target stimuli were recorded first. To get the non-existing monosyllabic tone primes (extension and surface primes), the speaker was asked to produce all targets with slow and medium speaking rates in both the extension and reduction sandhi forms, and the extension and surface primes were selected from the first syllables cut out from these disyllabic sandhi forms so that their durations matched the durations of the canonical and control primes as much as possible. With the relatively slow speaking rates of the disyllabic sandhi forms, we were able to achieve a close match between the durations of extension and surface primes and those of the canonical and control primes recorded in isolation. None of the stimuli, including primes and targets, was further manipulated.

The average duration of the primes was 454 ms, and the average duration of the targets was 667 ms. A one-way ANOVA showed that the duration of primes was not regulated by prime types ($F(3) = 0.95, p = 0.42$). An acoustic analysis was also conducted in Praat

Table 3: Examples of the four prime types for a sandhi target.

24 + 13 → 44 + 13 (tonal reduction) or 33 + 44 (tonal extension)		
Prime Type		Target
Canonical Prime	[kɔ24]	'to teach'
Extension Prime	[kɔ33]	[kɔ44 zã13] / kɔ24 zã13/
Surface Prime	[kɔ44]	'to complain about someone'
Control Prime	[kɔ53]	'tall'

(Boersma & Weenink, 2020) to compare the tones of the canonical prime, the extension prime, the surface prime, and the first syllable of the targets. An F0 measurement was taken at every 10% of the duration for the monosyllabic primes and the first syllable of the targets using ProsodyPro (Xu, 2013). Based on visual inspection of the comparison, the F0 curves of the canonical and extension primes were acoustically quite different from those of the targets' first syllables, while the F0 curves of the surface primes, which were taken from slower readings of the tonal reduction forms, had substantial overlap with those of the targets' first syllables. The average F0 curves across all critical prime and target syllables are plotted in **Figure 1** using the R package ggplot2 (Wickham, 2016).

In addition to the twenty critical V-N targets, twenty M-Ns and twenty V-Ns were included as fillers to mask the purpose of the experiment, as shown in **Table 4**. The speaker was asked to produce the disyllables as naturally and comfortably as possible. Among the twenty filler V-Ns, she used the tonal reduction form for eighteen items, and the tonal extension form for two of them. All twenty filler M-N items were produced with the rightward tonal extension sandhi. For the forty fillers, twenty of which were M-Ns and twenty were V-Ns, fifteen were primed by an extension tone with the same syllable (seven M-Ns and eight V-Ns), five were primed by a canonical tone with the same syllable

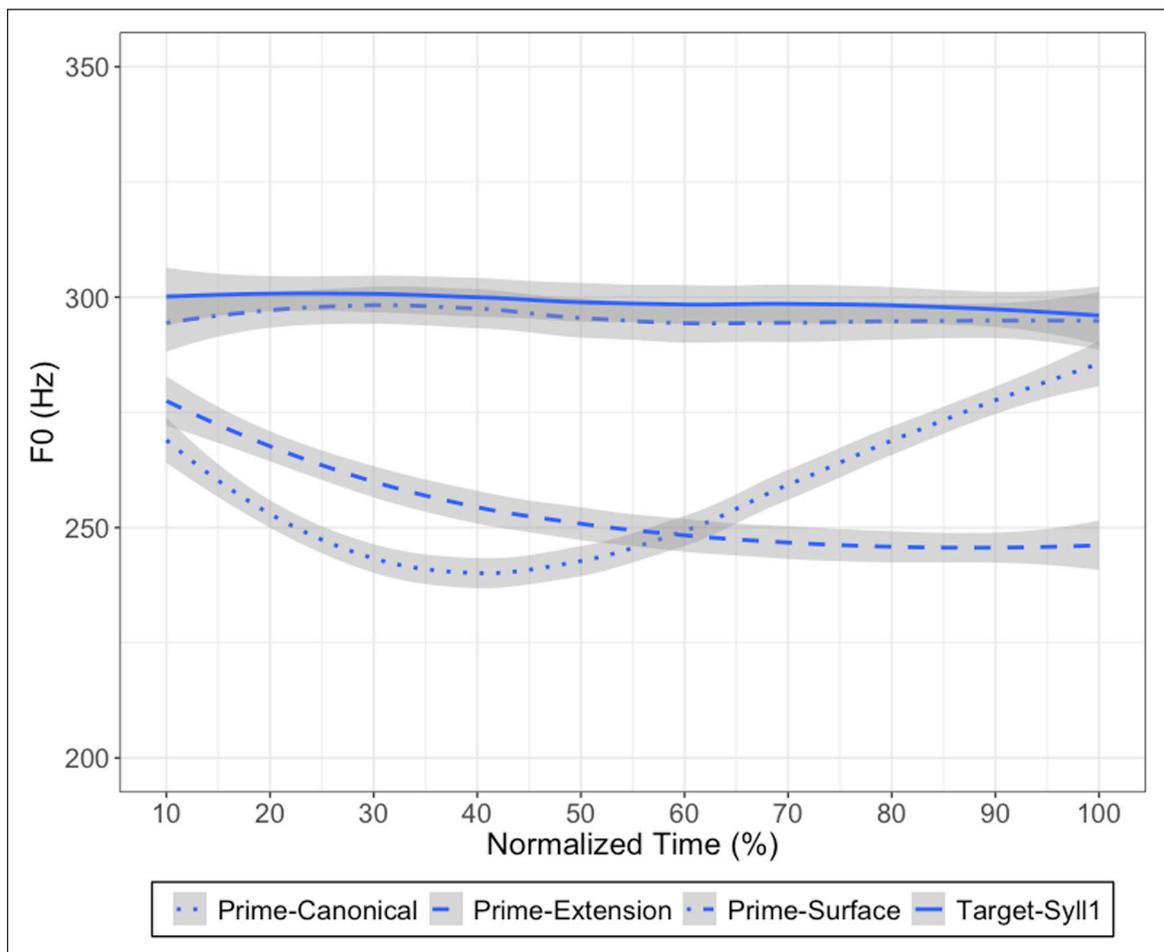


Figure 1: F0 data for the tones of the canonical prime, the extension prime, the surface prime, and the targets' first syllables. The dotted line represents the average F0 curve for the canonical primes, the dashed line represents the average F0 curve for the extension primes, the dotdashed line represents the average F0 curve for the surface primes, and the solid line represents the average F0 curve for the tones of the first syllables of the targets with the tonal reduction form. Shadows indicate the 95% confidence interval of the F0 data.

Table 4: Tonal combinations involved in critical, filler words, and nonwords.

Tonal combination	V-N	M-N	Nonwords
Tone 53 + tone 24	10 (filler)	10 (filler)	20
Tone 13 + tone 53	10 (filler)	10 (filler)	20
Tone 24 + tone 13	20 (critical)		20

(three M-Ns and two V-Ns), and twenty were primed by an unrelated tone with a different syllable (ten M-Ns and ten V-Ns). Sixty nonword targets were also included, with twenty in each of the three tonal combinations. The distributions of canonical, extension, surface, and unrelated primes for the nonwords were the same as those used in real words. The number of tones were balanced across critical targets, filler words, and nonwords. A full list of filler words and nonwords is given in Appendix B.

2.3 Procedures

The experiment was implemented in Paradigm (Tagliaferri, 2015), and all instructions were presented in Shanghai Wu. Twelve randomized practice trials were presented first before the 120 randomized main trials. The 20 critical V-N targets were presented in a Latin Square design such that each participant only heard each critical V-N once, preceded by one of the corresponding primes (canonical, extension, surface, or control). All participants heard the same filler words and nonwords.

Participants wore a pair of SONY headphones during the experiment. In each trial, participants first heard a monosyllabic prime. After 250 milliseconds (ms), they heard a disyllabic target, and then they needed to judge whether it was a real Shanghai word or not as quickly and accurately as possible by clicking the mouse (left button for ‘real word,’ right button for ‘nonword’). There was a 3000-ms interval between trials.

Without a Shanghai spoken-word frequency corpus, we conducted an auditory familiarity rating task after the priming experiment with the same participants to obtain the subjective familiarity of the sandhi targets. All participants rated the stimuli in a random order in Paradigm (Tagliaferri, 2015), with a response scale ranging from 0 ‘this word is rarely used’ to 3 ‘this word is often used.’ The rating was averaged across different participants to get the average familiarity rating for each target. The justification of using subjective frequency ratings to estimate frequency comes from Balota, Pilotti, and Cortese (2001), who collected subject frequency ratings for 2,938 monosyllabic English words from different groups of adult English speakers with varying ages and educational backgrounds and showed that there was a tight correlation between subjective estimates and objective log frequency estimates (Kučera & Francis, 1967; Baayen, Pipenbrock, & Gulikers, 1995). Therefore, instead of using a Mandarin corpus, whose word frequencies are most likely different from those of Shanghai, to estimate Shanghai word frequency, subjective ratings were collected to serve as an estimate for the relative frequency of exposure to a lexical item in the dialect. Since Shanghai shares the same writing system with Mandarin, to avoid the influence of Mandarin, the participants only heard the stimuli during the rating task.

After the priming and familiarity rating experiments, the participants were asked to do a sandhi goodness rating task (Yan, 2018), in which they rated the two variant sandhi forms of each V-N critical target (tonal extension form and tonal reduction form). Participants first saw a word on the screen and two speaker icons. They needed to click on the speaker icon on the top first to listen to the first pronunciation, and then consider if they say this word, how likely they would say it as the recording they just heard, with a response scale

ranging from 0 ‘I never say it in this way’ to 3 ‘I always say it in this way.’ They were then asked to click on the bottom speaker icon to listen to the second pronunciation, and similarly give a rating for this pronunciation according to their preference. The order of the two sandhi forms was counterbalanced. That is to say, for half of the V-Ns, participants heard the tonal extension form first, and for the other half, the tonal reduction form first. The rating for the reduction form was then subtracted from the rating for the extension form for each subject. This rating difference was averaged across different participants to get the average sandhi preference for each target. The scale of this measure ranged from -3 to 3, with a positive value indicating a preference for tonal extension, and a negative value a preference for tonal reduction.

In addition, to get a clearer picture of these speakers’ knowledge of monosyllabic tones, after the three tasks, participants were asked to produce the first syllable of the twenty critical items in isolation (twenty tone 24 monosyllables), as well as that of the twenty monosyllabic fillers (ten tone 53 and ten tone 13 monosyllables), in a random order. The monosyllables were presented one by one in PowerPoint. Recordings were made by a TASCAM DR-100MKIII recorder using an Electro-Voice 767 microphone with a sampling rate of 22,500 Hz. The monosyllabic production data were then classified by the first author, a phonetically-trained Shanghai speaker, into ‘Correct’ and ‘Incorrect.’ If either the segment or the tone was produced unexpectedly,⁶ it was classified as ‘Incorrect.’ The whole experiment took about 30 minutes.

3 Results

The reaction times and errors from the lexical decision task were examined. The overall accuracy rate for the critical V-N stimuli was 89% (658/740 trials, 20 critical items per participant * 37 participants). For the reaction time analyses on the critical stimuli, inaccurate responses (11%, 82/740 trials) as well as responses over two standard deviations from the mean reaction time (2%, 18/740 trials) were excluded. There were overlaps among these excluded items, and the final analysis was conducted on 640/740 trials.

To reduce the skewness of the raw data, reaction times were log-transformed, and then modeled with a series of Linear Mixed-Effects models with participant and item as random effects, and prime (canonical, extension, surface, control), familiarity, sandhi preference as fixed effects. The ratings of familiarity and sandhi preference for each item were averaged across all participants. For the effect of prime, the control prime was selected as the baseline to which canonical, extension, and surface primes were compared.

Nine models (A’, A, B, C, D, E, F, G, and H), including random-intercepts and by-subject random slopes for prime, were run and compared using likelihood ratio tests to determine the effects of prime, familiarity, sandhi preference, and their interactions, as shown in **Table 5**. All analyses were performed using the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2014), and *p* values were estimated using the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2016).

Table 5 shows that adding by-subject random slopes (Model A’) for prime did not significantly improve the null model with only random intercepts (Model A). Another seven random-intercepts-only models (B, C, D, E, F, G, and H) were run and compared using likelihood ratio tests to determine the effects of the fixed variables. The optimal model was chosen by removing interactions and factors that do not explain significant

⁶ ‘Expected’ productions are productions in accordance with the *Shanghai Dialect Dictionary* (Li et al., 1997). Productions with ‘expected’ tones but documented variant vowel forms were also considered as ‘expected.’

Table 5: Reaction time likelihood ratio tests on fixed effects: model comparison.

Model	Factor 1	Factor 2	Factor 3
A' (with random slopes)	Not applicable (N/A)	N/A	N/A
A	N/A	N/A	N/A
B	Prime	N/A	N/A
C	Prime	Familiarity	N/A
D	Prime	Familiarity	Prime: Familiarity
E	Prime	Familiarity	Sandhi preference
F	Prime	Familiarity	Prime: Sandhi preference
G	Prime	Familiarity	Familiarity: Sandhi preference
H	Prime	Familiarity	Prime: Familiarity: Sandhi preference
Model comparison	χ^2	df	p
A versus A'	5.237	9	0.813
B versus A	14.194	3	0.003**
C versus B	27.131	1	1.901e-07***
D versus C	3.709	3	0.295
E versus C	1.664	1	0.197
F versus C	4.498	4	0.343
G versus C	1.422	1	0.233
H versus C	5.330	4	0.255

amount of variance (Baayen, Davidson, & Bates, 2008). The best model was the one with prime and familiarity, of which the parameter estimates are listed in **Table 6**.

As shown in **Table 6**, both of the canonical prime and the extension sandhi tone prime elicited significantly faster logged reaction times for V-N recognition than the control prime did, while the surface prime tone did not. These can also be observed in the form of raw reaction time in **Figure 2**. Reaction time elicited by the extension and canonical primes was shorter than that by the surface and control primes.

The coefficient for familiarity was $-.090$, showing that with the increase of 1 unit in the familiarity rating, the logged reaction time decreased $.090$. It indicates that it took listeners less time to recognize more familiar V-N targets, as we can see in **Figure 3**. Finally, the lack of the prime and familiarity interaction suggests that both canonical and extension primes facilitated V-N recognition regardless of familiarity rating, and the degree of facilitation was not modulated by familiarity rating either.

To further compare the effects of different primes, we re-ran Model C, but used the surface prime as the baseline. We did not find a significant difference between the canonical and the surface prime ($\beta = -.013$, $SE = .010$, $t = -1.289$, $p = .20$), nor did we find a difference between the control and the surface prime ($\beta = .015$, $SE = .010$, $t = 1.489$, $p = .14$), but there was a marginal difference in logged reaction times elicited by the extension and the surface prime ($\beta = -.020$, $SE = .010$, $t = -1.958$, $p = .051$).

For the accuracy analysis, we conducted a set of logit mixed-effects models on accuracy with participant and item as random effects, and prime (canonical, extension, surface, control) and familiarity as fixed effects. The outliers were excluded, but both correct and incorrect response were included. The control prime was used as the baseline to which the canonical, extension, and surface primes were compared. Likelihood ratio tests were conducted to evaluate the effects of prime, familiarity, sandhi preference, and their

Table 6: Fixed effect estimates (top) and variance estimates (bottom) for model results of logged reaction times in Shanghai disyllabic V-N.

Fixed effect	Coefficient	SE of estimate	t	p
(Intercept)	3.311	0.029	112.466	<2e-16***
Prime-Canonical	-0.028	0.010	-2.797	0.005**
Prime-Extension	-0.035	0.011	-3.489	0.001**
Prime-Surface	-0.015	0.010	-1.489	0.137
Familiarity	-0.090	0.012	-7.495	2.56e-07**
Random Effects	s^2			
Residual	0.008			
participant (intercept)	0.003			
item (intercept)	0.001			

. $p \leq 0.1$ * $p \leq 0.05$ ** $p \leq 0.01$ *** $p \leq 0.001$.

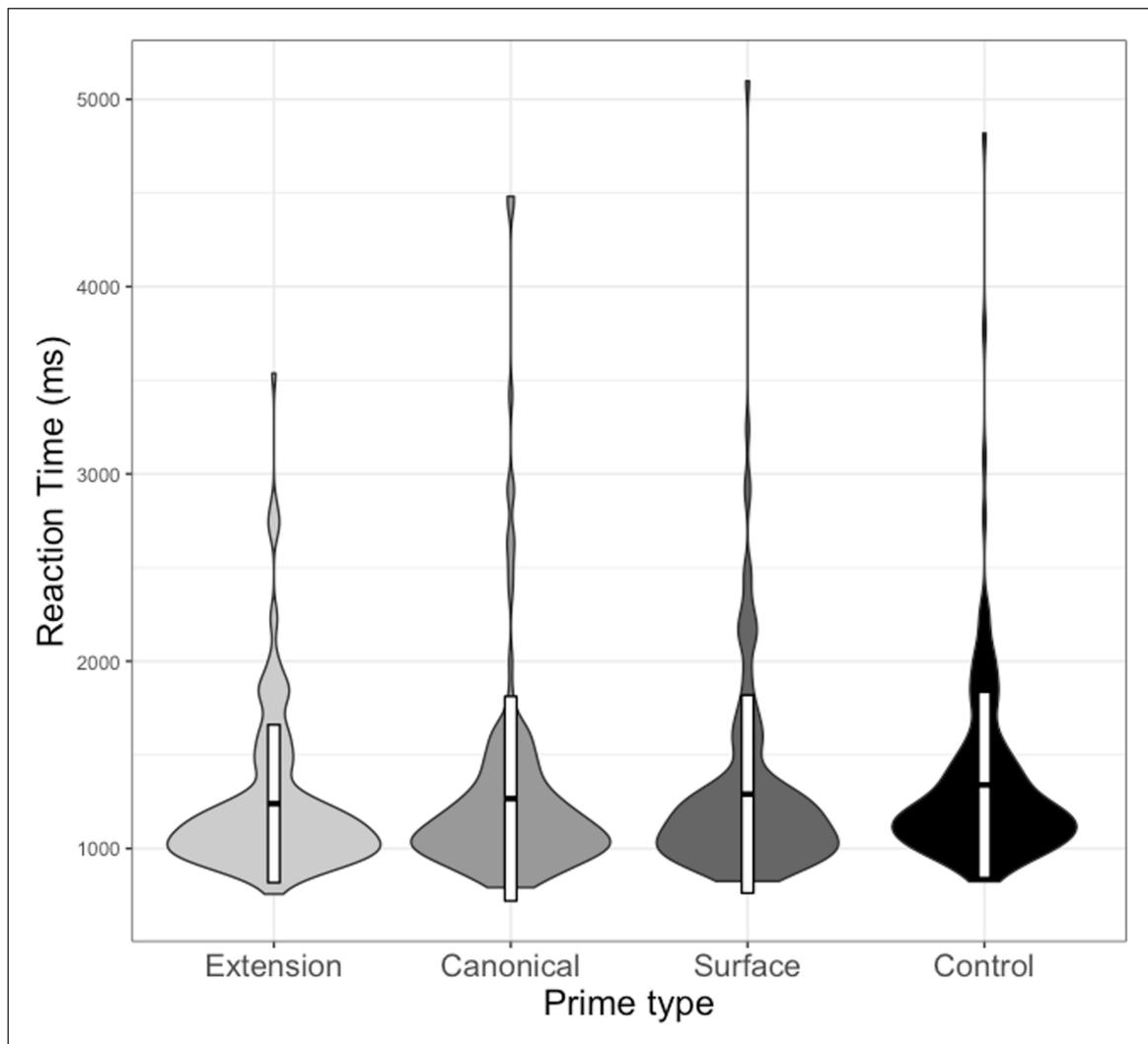


Figure 2: Reaction times for the four prime types presented as violin plots and boxplots. The boxplots indicate the mean \pm standard deviation, sorted by mean value (ascending order). The outlines of the violin plot show the kernel probability density, i.e., the width of the shaded area represents the probability of distribution of the data points.

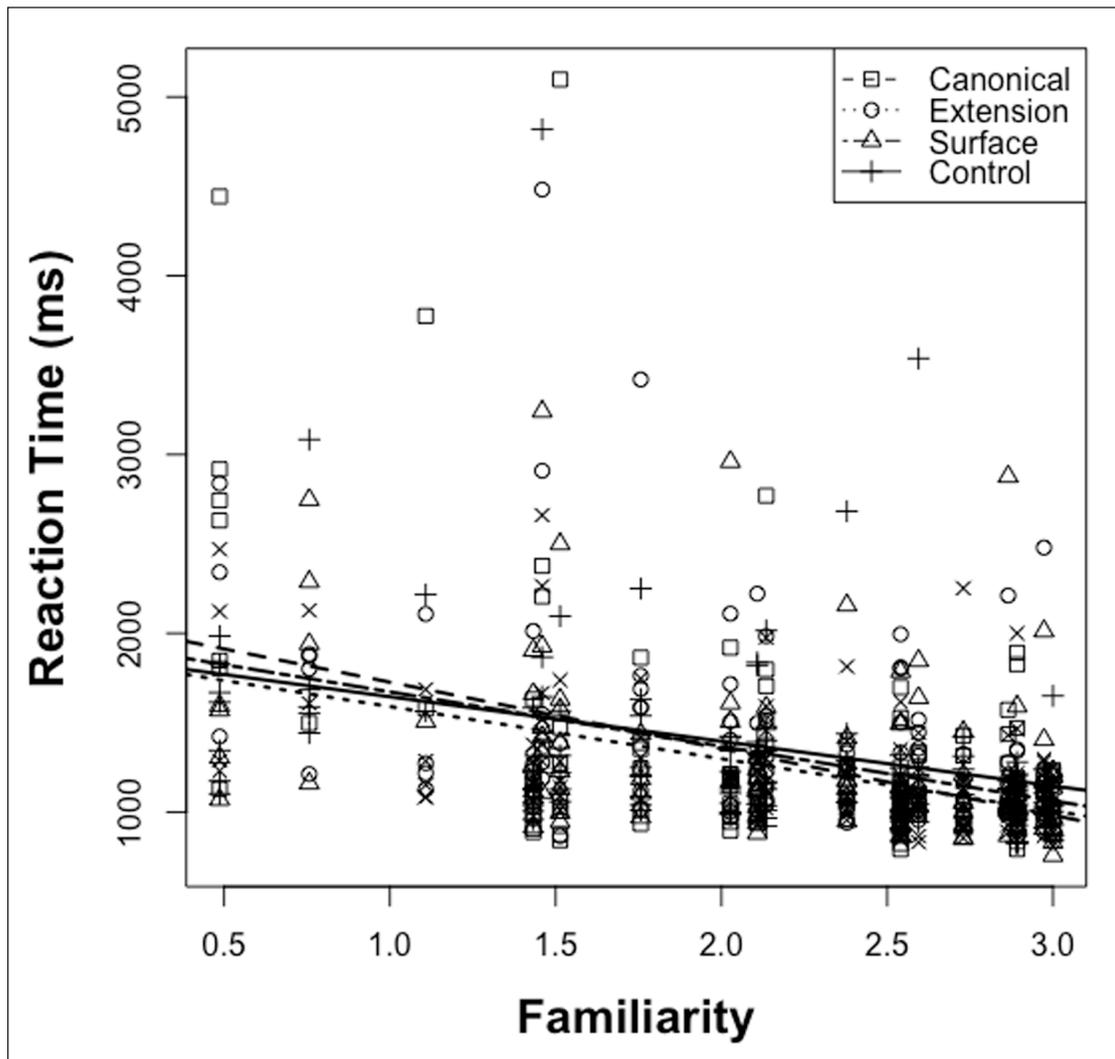


Figure 3: Reaction time elicited by the canonical, extension, surface, and control prime as a function of familiarity for Shanghai V-N items (each point represents a target word primed by a certain type of primes, which was differentiated by symbols).

interactions. As shown in **Table 7**, adding by-subject random slopes for prime (Model A') did not significantly improve the random-intercepts-only null model (Model A). Another seven random-intercepts-only models (B, C, D, E, F, G, and H) were run and compared using likelihood ratio tests to identify the optimal mixed-effects model and to determine the effects of the fixed variables. The results of the best model, Model C, showed that more familiar targets had higher accuracy than less familiar ones ($z(715) = 6.004, p < .001$), but the accuracy was not regulated by prime types or sandhi preference; see **Table 8**.

In all of the analyses above, sandhi preference did not influence the reaction time or accuracy of V-N recognition. To further examine its effect, a regression test was run between familiarity rating and sandhi preference of each participant on each item, and there was a marginal correlation between these two ($t(638) = -1.916, p = .056$). There was also a negative correlation between these two factors when they were averaged across the participants by item ($t(638) = -6.315, p < .001$), suggesting that the more familiar the item was, the lower the sandhi preference value it had, i.e., it had a stronger preference towards tone reduction. This finding replicates the finding in Yan (2018) that Shanghai speakers preferred the tonal reduction form for more familiar V-Ns.

Table 7: Accuracy likelihood ratio tests on fixed effects: model comparison.

Model	Factor 1	Factor 2	Factor 3
A' (with random slopes)	Not applicable (N/A)	N/A	N/A
A	N/A	N/A	N/A
B	Prime	N/A	N/A
C	Prime	Familiarity	N/A
D	Prime	Familiarity	Prime: Familiarity
E	Prime	Familiarity	Sandhi preference
F	Prime	Familiarity	Prime: Sandhi preference
G	Prime	Familiarity	Familiarity: Sandhi preference
H	Prime	Familiarity	Prime: Familiarity: Sandhi preference
Model comparison	χ^2	Df	p
A versus A'	4.561	9	0.871
B versus A	3.934	3	0.269
C versus B	23.168	1	1.484e-06***
D versus C	2.845	3	0.416
E versus C	0.058	1	0.810
F versus C	2.617	4	0.624
G versus C	0.076	1	0.783
H versus C	1.786	4	0.775

Table 8: Fixed effect estimates (top) and variance estimates (bottom) for model results of accuracy in Shanghai disyllabic V-N.

Fixed effect	Coefficient	SE of estimate	z	p
(Intercept)	-1.379	0.742	-1.858	0.063
Prime-Canonical	-0.135	0.414	-0.327	0.743
Prime-Extension	0.279	0.433	0.643	0.520
Prime-Surface	-0.522	0.405	-1.289	0.197
Familiarity	2.307	0.384	6.004	1.93e-09***
Random Effects	s ²			
participant (intercept)	0.208			
item (intercept)	0.686			

. $p \leq 0.1$ * $p \leq 0.05$ ** $p \leq 0.01$ *** $p \leq 0.001$.

Finally, the production study of the monosyllables showed that the participants were highly familiar with the canonical tone of the monosyllabic morphemes: The average accuracy rate of their production was 94% for all items ($SD = 7\%$), and 92% for the critical targets ($SD = 13\%$).

4 Discussion

Using an auditory priming method whereby participants heard monosyllabic primes followed by disyllabic targets for which they had to conduct lexical decision, the present study investigated how native Shanghai Wu speakers represent and process V-N disyllables that have variant sandhi forms (tonal extension and tonal reduction). The results showed

that overall, both the canonical prime and the extension prime facilitated the recognition of V-N targets, but the surface prime did not have a facilitation effect. In addition, more familiar V-N targets were recognized more quickly than less familiar ones, but there was no interaction between prime type and familiarity, or between prime type and sandhi preference. The facilitation of both the canonical and extension primes is likely due to the tone sandhi variation of V-N items, suggesting that both canonical and extension forms are represented in the lexicon. These results have the following implications.

First, the current results help us further understand the effects of locality and structure preservation on the representation and processing of tone sandhi words. Although Shanghai V-N reduction differs from Mandarin tone 3 sandhi with regard to structure preservation, the canonical tone primes elicited faster reaction times for word recognition in both of the sandhi patterns (cf. Chien et al., 2016), suggesting that non-structure-preservation does not in and of itself lead to a facilitation effect from the surface form and therefore has little influence on tone sandhi word representation. Relatedly, this also supports Yan et al.'s (2020) contention that the surface facilitative effect found in the Shanghai rightward tonal extension sandhi of M-Ns was due to non-local tonal spreading, not the non-structure-preserving property of the tonal extension sandhi, as the non-structure-preserving and local Shanghai reduction sandhi did not yield a surface facilitative effect.

Second, the current study found that both canonical and extension tone prime facilitated the recognition of V-Ns in tonal reduction form (the preferred form). This supports the prediction of the hybrid model of lexical representation (e.g., Connine & Pinnow, 2006; Mitterer, Chen, & Zhou, 2011; Ranbom & Connine, 2007), which states that multiple phonological variant forms of a given word may be stored in the lexical representation. The fact that we observed an extension tone priming effect in V-N even though the current study used frequent Shanghai users, while Yan et al. (2020) only observed extension tone priming in M-N for infrequent users, requires further comments. Yan et al.'s (2020) account for the lack of extension tone priming effect for frequent users of Shanghai was that the facilitative effect from the phonological overlap between the prime and the first syllable of the target could be canceled out by lexical competition that resulted from the multiple word candidates activated by the prime on the lexical level, and this lexical competition was particularly strong for frequent Shanghai users, who had a large semantic knowledge base. We contend that the extension tone prime played a different role in V-N than in M-N. For M-N, the extension tone prime was the surface tone of the first syllable of the extension sandhi form, and it may activate a number of disyllabic words with the same initial syllable and the same sandhi melody, causing lexical competition; but for V-N, the extension tone prime was the tone of the first syllable of a dispreferred variable sandhi pattern (Yan, 2018) that the participants did not hear in the disyllabic targets. When hearing the extension tone prime, lexical candidates with the extension sandhi form may be activated, including the one used as the target in its dispreferred extension sandhi form; however, this particular lexical candidate was presented in its preferred tonal reduction form as the target to the participants. The surface discrepancy between the target (tonal reduction form) and the previously activated lexical competitors (extension form) may have resulted in reduced lexical competition. This allowed the facilitative effect of the extension tone to emerge even for frequent users of Shanghai.

Third, although the current study replicated Yan's (2018) finding that Shanghai speakers preferred the tonal reduction variant more for more familiar V-N items, we did not find an interaction between prime type and familiarity rating. Compared to the control prime, both the canonical prime and the extension prime facilitated the recognition of V-N targets regardless of familiarity ratings. The sandhi preference ratings also had no direct bearing on the recognition of the targets. There are two possible interpretations for the lack of

prime type and familiarity interaction. One is that, although the effect of familiarity on sandhi preference is significant in both the current study and Yan (2018), there is overall a very strong preference for tonal reduction for V-N items. This is indicated by the fact that all V-N test items were pronounced with tonal reduction by the native speaker who provided the auditory stimuli. Therefore, although the listeners were aware of the tonal extension sandhi form as a possible variant, the effect of item familiarity (proxy for item frequency) on the frequencies of the variants might be very small, leading to the lack of interaction between prime type and familiarity. Another possibility is that the lexical representation of words with variable pronunciations is not sensitive to the frequencies of the variants. The fact that the tonal reduction variant is the strongly preferred, and therefore most likely more frequent variant than the tonal extension variant for V-N items, yet they elicited similar facilitative priming for V-N items, provides further support for this possibility. If so, then the finding here is consistent with that of Sumner and Samuel (2005), who showed that for English final-/t/ variation, all regular variant forms were equally effective in facilitating immediate processing regardless of their frequency. This casts doubts on exemplar-based models. The support of the current study on the hybrid model of lexical representation (e.g., Connine, 2004; Connine & Pinnow, 2006; Deelman & Connie, 2001; Ranbom & Connine, 2007) is then only partial, as the frequency modulation encoded in this type of models is not found in our study.

Finally, the current study showed the importance of underlying representations in spoken word recognition for at least certain types of alternation. Together with the findings from Chien et al. (2016, 2017) and Yan et al. (2020), we have provided evidence that words undergoing a local and transparent phonological alternation are represented in their underlying form regardless of whether the alternation is structure-preserving, and that the listeners access this representation in spoken word recognition. This is consistent with earlier works from priming that suggested that listeners can access abstract phonological representations in processing when there is predictable alternation such as place or voicing assimilation (e.g., Gaskell & Marslen-Wilson, 1996; Gow, 2001, 2002; Lahiri, Jongman, & Sereno, 1990; Luce et al., 2001). For instance, Gow (2001) showed that an item that has undergone regressive place assimilation (e.g., *gree[m] beans*) primes the unassimilated form in isolation (e.g., *green*). However, the lack of surface form priming in the current study as well as in Chien et al. (2016) is inconsistent with findings from Connine and colleagues and Sumner and Samuel (2005) that suggested that non-structure-preserving variants of phonemes (e.g., as the result of flapping or /t/-glottalization in English) are part of the lexical representation. We propose that this difference may have resulted from the different nature between phonological alternations that are morpho-syntactically conditioned or impact entire morphemes, such as Shanghai tonal reduction and Mandarin tone 3 sandhi, and speech-register conditioned reduction processes on the word level. Take English flapping as an example. The alternation of an underlying /t/ or /d/ to a flap occurs on one transient segment and has relatively little morphosyntactic impact. Storing the surface form of the whole word could accelerate the speed of spoken word recognition. For Shanghai tonal reduction in V-N items, on the other hand, the alternation occurs on the phrasal domain with paradigmatic ramifications; moreover, even though the reduction tone sandhi is non-structure-preserving, the process impacts the whole syllable and hence the morpheme. The Mandarin tone 3 sandhi also impacts the whole morpheme and creates multiple additional homophones to boot. For these types of processes, if the surface form is also represented, it will likely have a more negative effect on morphological parsing. This could be the reason that the surface form did not facilitate the recognition of Shanghai V-N items, which undergo tonal reduction, nor did it yield a facilitative effect for Mandarin tone 3 sandhi words (Chien et al., 2016). However, it

is important to note that the surface forms are indeed activated when the task involves production (Chen, Shen, & Schiller, 2011; Nixon et al., 2015; Politzer-Ahles & Zhang, in press).

5 Conclusion

Disyllabic verb-noun (V-N) items in Shanghai Wu can variably undergo either a rightward extension tone sandhi or tonal reduction on the first syllable. The current study investigates how the phonological properties of these alternation processes as well as the variation influence how Shanghai speakers represent and access such words. An auditory-auditory priming lexical decision experiment on Shanghai V-N items, with the targets preceded by monosyllabic primes with different tones, showed that primes with the canonical tone and the tonal-extension tone of the first syllable of the targets elicited facilitation, but surface tone primes did not. Moreover, although more familiar V-Ns were recognized with shorter reaction time, the priming effect did not interact with speakers' familiarity ratings or sandhi preference ratings of the targets. These data are consistent with the interpretation that both the canonical and tonal-extension forms are represented in Shanghai speakers' mental lexicon due to tone sandhi variation, supporting a hybrid model of lexical representation, but the representation does not seem to be modulated by the frequencies of the variants. Also, together with findings from auditory priming studies of other tone sandhi patterns, the current study suggests that the phonological properties of an alternation, such as its locality and transparency, influence the representation of words undergoing the alternation; but whether the alternation is structure-preserving does not seem to impact the representation.

Additional Files

The additional files for this article can be found as follows:

- **Appendix A.** Critical target. DOI: <https://doi.org/10.5334/labphon.264.s1>
- **Appendix B.** Disyllabic filler word and non-word target. DOI: <https://doi.org/10.5334/labphon.264.s2>

Acknowledgements

We are grateful to the experimental participants in Shanghai for taking part in our study. We also thank the General Editor Dr. Mirjam Ernestus, Associate Editor Dr. Alan Yu, and two anonymous reviewers, whose comments substantially improved this article.

Funding Information

This study was funded by Shanghai Planning Office of Philosophy and Social Sciences, China, awarded to the first author, under grant [2018EYY002]. Neither the individuals and institutions cited herein nor the funding agency, however, should be held responsible for the views in this article.

Competing Interests

The authors have no competing interests to declare.

References

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412. DOI: <https://doi.org/10.1016/j.jml.2007.12.005>
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database (CD-ROM). Philadelphia, PA: Linguistic Data Consortium.

- Balota, D. A., Pilotti, M., Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition*, 29(4), 639–647. DOI: <https://doi.org/10.3758/BF03200465>
- Bates, D., Maechler, M., Bolker, B. M., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *Journal of Statistical Software*. Retrieved from <http://arxiv.org/abs/1406.5823>. DOI: <https://doi.org/10.18637/jss.v067.i01>
- Boersma, P., & Hayes, B. (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry*, 32, 45–86. DOI: <https://doi.org/10.1162/002438901554586>
- Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer. Retrieved from <https://www.fon.hum.uva.nl/praat/>.
- Bybee, J. L. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511612886>
- Chao, Y. R. (1930). A system of tone letters. *Le Maître Phonétique* (3rd series) 45, 24–47.
- Chen, M. Y. (1987). The syntax of Xiamen tone sandhi. *Phonology Yearbook*, 4, 109–149. DOI: <https://doi.org/10.1017/S0952675700000798>
- Chen, M. Y. (2000). Tone sandhi: Patterns across Chinese dialects. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511486364>
- Chen, Y.-Y. (2011). How does phonology guide phonetics in segment–f0 interaction? *Journal of Phonetics*, 39(4), 612–625. DOI: <https://doi.org/10.1016/j.wocn.2011.04.001>
- Chen, Y., Shen, R., & Schiller, N. O. (2011). Representation of allophonic tone sandhi variants. In *Proceedings of Psycholinguistics Representation of Tone. Satellite Workshop to ICPHS*, Hongkong (pp. 38–41).
- Cheng, R. L. (1968). Tone sandhi in Taiwanese. *Linguistics*, 41, 19–42. DOI: <https://doi.org/10.1515/ling.1968.6.41.19>
- Chien, Y.-F., Sereno, J. A., & Zhang, J. (2016). Priming the representation of Mandarin Tone3 sandhi words. *Language, Cognition and Neuroscience*, 31(2), 179–189. DOI: <https://doi.org/10.1080/23273798.2015.1064976>
- Chien, Y.-F., Sereno, J., & Zhang, J. (2017). What’s in a word: Observing the contribution of both underlying and surface representations. *Language and Speech*, 60, 643–657. DOI: <https://doi.org/10.1177/0023830917690419>
- Chien, Y.-F., Yang, X., Fiorentino, R., & Sereno, J. A. (2020). The role of surface and underlying forms when processing tonal alternations in Mandarin Chinese: A Mismatch Negativity study. *Frontiers in Psychology*, 11, 646. DOI: <https://doi.org/10.3389/fpsyg.2020.00646>
- Clarke, C. M. (2003). *Processing time effects of short-term exposure to foreign-accented English*. Doctoral dissertation, University of Arizona, Tucson.
- Coetzee, A. W. (2004). *What It Means to Be a Loser: Non-Optimal Candidates in Optimality Theory*. Doctoral dissertation, University of Massachusetts Amherst.
- Coetzee, A. W. (2006). Variation as accessing “non-optimal” candidate. *Phonology*, 23, 337–385. DOI: <https://doi.org/10.1017/S0952675706000984>
- Coetzee, A. W., & Kawahara, S. (2012). Frequency biases in phonological variation. *Natural Language & Linguistic Theory*, 31, 47–89. DOI: <https://doi.org/10.1007/s11049-012-9179-z>
- Coetzee, A. W., & Pater, J. (2011). The place of variation in phonological theory. In J. Goldsmith, J. Riggle & A. Yu (Eds.), *The Handbook of Phonological Theory: 2nd Edition* (pp. 401–434). Cambridge: Blackwell. DOI: <https://doi.org/10.1002/9781444343069.ch13>
- Connine, C. M. (1994). Vertical and horizontal similarity in spoken word recognition. *Perspectives on sentence processing*, 107–120.

- Connine, C. M. (2004). It's not what you hear, but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review*, *11*, 1084–1089. DOI: <https://doi.org/10.3758/BF03196741>
- Connine, C. M., & Pinnow, E. (2006). Phonological variation in spoken word recognition: Episodes and abstractions. *The Linguistic Review*, *23*, 235–245. DOI: <https://doi.org/10.1515/TLR.2006.009>
- Connine, C. M., Ranbom, L. J., Patterson, D. J. (2008). On the representation of phonological variant frequency in spoken word recognition. *Perception & Psychophysics*, *70*, 403–411. DOI: <https://doi.org/10.3758/PP.70.3.403>
- Crystal, T., & House, A. (1988). The duration of American-English stop consonants: An overview. *Journal of Phonetics*, *16*, 285–294. DOI: [https://doi.org/10.1016/S0095-4470\(19\)30503-0](https://doi.org/10.1016/S0095-4470(19)30503-0)
- Deelman, T., & Connine, C. M. (2001). Missing information in spoken word recognition: Non-released stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 656–663. DOI: <https://doi.org/10.1037/0096-1523.27.3.656>
- Fu, G., Cai, Y., Bao, S., Fang, S., Fu, Z., & Zhengzhang, S. (1986). Wuyu de Fenqu (Gao) [The division of Wu dialects (manu.)]. *Fangyan [Dialects]*, *1*, 1–7.
- Gao, J.-Y., & Hallé, P. (2017). Phonetic and phonological properties of tones in Shanghai Chinese. *Cahiers de Linguistique Asie Orientale*, *46*, 1–31. DOI: <https://doi.org/10.1163/19606028-04601001>
- Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(1), 144–158. DOI: <https://doi.org/10.1037/0096-1523.22.1.144>
- Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology*, *45*, 220–266. DOI: [https://doi.org/10.1016/S0010-0285\(02\)00003-8](https://doi.org/10.1016/S0010-0285(02)00003-8)
- Goldinger, S. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279. DOI: <https://doi.org/10.1037/0033-295X.105.2.251>
- Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language*, *45*, 133–159. DOI: <https://doi.org/10.1006/jmla.2000.2764>
- Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 163–179. DOI: <https://doi.org/10.1037/0096-1523.28.1.163>
- Guy, G. R. (1994). The phonology of variation. In K. Beals (Ed.), *CLS 30: Papers from the 30th Regional Meeting of the Chicago Linguistic Society. Volume 2: The Parasession on Variation in Linguistic Theory* (pp. 133–149). Chicago: Chicago Linguistic Society.
- Hayes, B., & Abad, M. (1989). Reduplication and syllabification in Ilokano. *Lingua*, *77*, 331–374. DOI: [https://doi.org/10.1016/0024-3841\(89\)90044-2](https://doi.org/10.1016/0024-3841(89)90044-2)
- Hooper, J. B. (1976). *An Introduction to Natural Generative Phonology*. New York: Academic Press.
- Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.), *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium* (pp. 29–54). Tokyo, Japan: The National International Institute for Japanese Language.
- Kiparsky, P. (1973). Abstractness, opacity and global rules. In O. Fujimura (Ed.), *Three dimensions in phonological theory* (pp. 57–86). Tokyo, Japan: TEC Company.
- Kučera, M., & Francis, W. N. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.

- Kuznetsova, A., Brockhoff, B., & Christensen, H. (2016). Tests in linear mixed effects models. <https://cran.r-project.org/web/packages/lmerTest/index.html>. DOI: <https://doi.org/10.18637/jss.v082.i13>
- Labov, W. (1989). The child as linguistic historian. *Language Variation and Change*, 1, 85–97. DOI: <https://doi.org/10.1017/S0954394500000120>
- Lahiri, A., Jongman, A., & Sereno, J. (1990). The pronominal clitic [dər] in Dutch: A theoretical and experimental approach. In G. E. Booij & J. van Marle (Eds.), *Yearbook of Morphology*, 3, 115–127. Dordrecht: Foris Publications. DOI: <https://doi.org/10.1515/9783112420744-009>
- Li, Q. (2016). *The production and perception of tonal variation: Evidence from Tianjin Mandarin*. Doctoral dissertation, Leiden University.
- Li, R., Xu, B., & Tao, H. (1997). *Shanghai Fangyan Cidian [Shanghai Dialect Dictionary]*. Nanjing: Jiangsu Education Press.
- Luce, P. A., McLennan, C. T., & Charles-Luce, J. (2001). Abstract and specificity in spoken word recognition: Indexical and allophonic variability in long-term repetition priming. In J. Bowers & C. Marsolek (Eds.), *Rethinking implicit memory* (pp. 197–214). Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780192632326.003.0009>
- MacLennan, C., Luce, P., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29, 539–553. DOI: <https://doi.org/10.1037/0278-7393.29.4.539>
- Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, 101, 653–675. DOI: <https://doi.org/10.1037/0033-295X.101.4.653>
- Mitterer, H., Chen, Y., & Zhou, X. (2011). Phonological abstraction in processing lexical-tone variation: Evidence from a learning paradigm. *Cognitive Science*, 35(1), 184–197. DOI: <https://doi.org/10.1111/j.1551-6709.2010.01140.x>
- 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. DOI: <https://doi.org/10.1080/23273798.2014.942326>
- Norris, D., McQueen, J., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, 23, 299–370. DOI: <https://doi.org/10.1017/S0140525X00003241>
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47, 204–238. DOI: [https://doi.org/10.1016/S0010-0285\(03\)00006-9](https://doi.org/10.1016/S0010-0285(03)00006-9)
- Patterson, D., & Connine, C. M. (2001). Variant frequency in flap production: A corpus analysis of variant frequency in American English flap production. *Phonetica*, 58, 254–275. DOI: <https://doi.org/10.1159/000046178>
- Patterson, D., LoCasto, P., & Connine, C. M. (2003). Corpora analyses of frequency of schwa deletion in conversational American English. *Phonetica*, 60, 45–69. DOI: <https://doi.org/10.1159/000070453>
- Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency Effects and Emergent Grammar* (pp. 137–157). Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/tsl.45.08pie>
- Pierrehumbert, J. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 177–228). Cambridge, MA: The MIT Press.
- Politzer-Ahles, S., & Zhang, J. (in press). Evidence for the role of tone sandhi in Mandarin speech production. *Journal of Chinese Linguistics* [Special Issue: Studies on Tonal Aspect of Languages].

- Ranbom, L., & Connine, C. M. (2004). On the role of phonological variant frequency in spoken word recognition. *147th Meeting of the Acoustical Society of America*, New York City, NY, USA.
- Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*, 273–298. DOI: <https://doi.org/10.1016/j.jml.2007.04.001>
- Shih, C.-L. (1997). Mandarin Third Tone sandhi and prosodic structure. In J. Wang & N. Smith (Eds.), *Studies in Chinese phonology* (pp. 81–123). Berlin: Mouton de Gruyter.
- Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, *52*, 322–338. DOI: <https://doi.org/10.1016/j.jml.2004.11.004>
- Tagliaferri, B. (2015). Perception Research Systems Inc. (Paradigm). Retrieved from <http://www.paradigmexperiments.com/>.
- Takahashi, Y. (2013). *The phonological structure of Shanghai tone sandhi*. Doctoral dissertation, Tokyo University of Foreign Studies, Tokyo, Japan.
- Tian, J., & Kuang, J.-J. (2019). The phonetic properties of the non-modal phonation in Shanghainese. *Journal of International Phonetic Association*. DOI: <https://doi.org/10.1017/S0025100319000148>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. DOI: https://doi.org/10.1007/978-3-319-24277-4_9
- Xu, B., Tang, Z., & Qian, N. (1981). Xinpai Shanghai fangyan de liandu biandiao I [Tone sandhi in new Shanghai I]. *Fangyan [Dialects]*, *2*, 145–155.
- Xu, B., Tang, Z., & Qian, N. (1982). Xinpai Shanghai fangyan de liandu biandiao II [Tone sandhi in new Shanghai II]. *Fangyan [Dialects]*, *2*, 115–128.
- Xu, B., & You, R. (1984). Sunan he Shanghai Wuyu de Neibu Chayi [The internal difference between Sunan Wu and Shanghai Wu]. *Fangyan [Dialects]*, *1*, 3–12.
- Xu, Y. (2013). ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France, 7–10.
- Yan, H. (2018). *The nature of variation in tone sandhi patterns of Shanghai and Wuxi Wu*. Singapore: Springer Nature. DOI: <https://doi.org/10.1007/978-981-10-6181-3>
- Yan, H., Chien, Y.-F., & Zhang, J. (2020). Priming the representation of left-dominant sandhi words: A Shanghai dialect case study. *Language and Speech*, *63*(2), 362–380. DOI: <https://doi.org/10.1177/0023830919849081>
- Zhang, C., & Peng, G. (2013). Productivity in Mandarin Third Tone sandhi: A wug test. In *Eastward Flows the Great River: Festschrift in Honor of Prof. William S.-Y. Wang on his 80th Birthday*, 256–282. Hong Kong: City University of Hong Kong.
- Zhang, J. (2007). A directional asymmetry in Chinese tone sandhi systems. *Journal of East Asian Linguistics*, *16*, 259–302. DOI: <https://doi.org/10.1007/s10831-007-9016-2>
- Zhang, J., & Lai, Y.-W. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, *27*(1), 153–201. DOI: <https://doi.org/10.1017/S0952675710000060>
- Zhang, J., Lai, Y.-W., & Sailor, C. (2011). Modeling Taiwanese speakers' knowledge of tone sandhi in reduplication. *Lingua*, *121*, 181–206. DOI: <https://doi.org/10.1016/j.lingua.2010.06.010>
- Zhang, J., & Meng, Y. (2016). Structure-dependent tone sandhi in real and nonce disyllables in Shanghai Wu. *Journal of Phonetics*, *54*, 169–201. DOI: <https://doi.org/10.1016/j.wocn.2015.10.004>
- Zhang, J., & Yan, H. (2018). Contextually dependent cue realization and cue weighting for a laryngeal contrast in Shanghai Wu. *The Journal of the Acoustical Society of America*, *144*(3), 1293–1308. DOI: <https://doi.org/10.1121/1.5054014>

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 other Asian Languages* (pp. 3–26). Hong Kong: The Chinese University Press.

How to cite this article: Yan, H., Chien, Y-F., & Zhang, J. 2021 The representation of variable tone sandhi patterns in Shanghai Wu. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 12(1): 15, pp.1–24. DOI: <https://doi.org/10.5334/labphon.264>

Submitted: 06 February 2020 **Accepted:** 18 July 2021 **Published:** 11 August 2021

Copyright: © 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

][*Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 