Journal of Neurolinguistics 62 (2022) 101060

Contents lists available at ScienceDirect

Journal of Neurolinguistics

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

The neural encoding of productive phonological alternation in speech production: Evidence from Mandarin Tone 3 sandhi

Jie Zhang [a,*], Caicai Zhang [b], Stephen Politzer-Ahles [b], Ziyi Pan [b], Xunan Huang [c], Chang Wang [a], Gang Peng [b], Yuyu Zeng [a]

[a] *Department of Linguistics, University of Kansas, Lawrence, KS, USA*
[b] *Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University, Hong Kong*
[c] *School of Foreign Languages, University of Electronic Science and Technology, Chengdu, Sichuan, China*

ARTICLE INFO

Keywords:
Tone sandhi
Mandarin Chinese
Speech production
Event-related potentials
Phonological alternation
Word frequency

ABSTRACT

The understanding of alternation is a key goal in phonological research. But little is known about how phonological alternations are implemented in speech production. The current study tested the hypothesis that the production of words that undergo a highly productive alternation, Mandarin Tone 3 sandhi, is supported by a computation mechanism, which predicts that this alternation is subserved by neural activity in a time-window associated with post-lexical phonological and phonetic encoding regardless of word frequency. ERPs were recorded while participants sub-vocally produced high- and low-frequency disyllabic words that do or do not require sandhi. Sandhi words elicited more positive ERPs than non-sandhi words over left anterior channels around 336–520 ms after participants saw the cue instructing them to initiate sub-vocal production, but this effect was not significantly modulated by word frequency. These findings are consistent with predictions of the computation mechanism and have implications for current psycholinguistic models of speech production. (150 words)

1. Introduction

A core issue in psycholinguistic studies of speech production is the time-course of various mental processes that take place in a speaker's brain leading up to the articulation of speech sounds. According to Levelt and colleagues' serial-order model (Hagoort & Levelt, 2009; Levelt, 1999; Schriefers et al., 1990), word production involves several consecutive mental processes with different brain localizations. Indefrey (2011) summarized these processes as follows: (1) conceptual preparation (0–200+ms), (2) lemma retrieval (~200–275 ms), (3) phonological code retrieval (~275–355 ms), (4) phonological encoding (~355–455 ms), (5) phonetic encoding (~455–600 ms), and (6) articulation (600 ms+). Using intracranial electrophysiology, Sahin et al. (2009) probed into the time-course of word production in English speakers and reported patterns of neural activity that are largely consistent with the model above. Specifically, the authors found that lexical retrieval occurred about 200 ms after the onset of a visually presented word to elicit production, and phonological and phonetic encoding occurred about 450 ms afterwards.

What complicates the picture above, however, is that the same morpheme may be pronounced differently depending on the phonological context in which it appears, a phenomenon commonly known as phonological alternation. For instance, the past tense form *-ed* in English is produced as [ɪd] in words ending with coronal stops, as [d] in words ending with other voiced sounds, and as [t]

* Corresponding author. Department of Linguistics, The University of Kansas, 1541 Lilac Lane, Blake Hall 427, Lawrence, KS, 66045, USA.
E-mail address: zhang@ku.edu (J. Zhang).

https://doi.org/10.1016/j.jneuroling.2022.101060
Received 19 July 2021; Received in revised form 13 January 2022; Accepted 19 January 2022
Available online 2 February 2022

in words ending with other voiceless sounds. In generative grammar, it is assumed that the regular allomorphs of a morpheme (i.e., different pronunciations of the same meaningful unit that are fully predictable by contexts) share one underlying representation, and an input-output mapping mechanism maps the underlying representation to the different surface representations in the appropriate contexts as part of the competence grammar (Cole & Hualde, 2011; Kenstowicz & Kisseberth, 1979). Although phonological analyses along this line are generally considered to be formal models of speakers' grammatical competence and do not make the claim that speakers perform the mapping online, there is the question of how words that involve phonological alternation are implemented in speech production, and the answer to this question remains unclear. Major psycholinguistic models of speech production, e.g., the serial-order model sketched out above, also do not specify the mechanisms for the production of alternating forms.

The key theoretical debate in the psycholinguistic studies of morphology has often revolved around whether a morphologically complex form is computed online or lexically stored and how morphophonological regularity bears on the issue. For instance, arguments have been made from studies on productivity and aphasia that irregularly inflected forms (e.g., *went*) are lexically stored, while regularly inflected forms (e.g., *walked*) are computed (e.g., Pinker & Prince, 1994; Prasada & Pinker, 1993; Ullman et al., 1997). This position has been subsequently challenged on two different fronts: on the one hand, studies have shown that lexical measures of regularly inflected forms such as their frequency influence their processing (e.g., Baayen et al., 2003; Meunier & Segui, 1999; Stemberger & MacWhinney, 1986), suggesting that regularly inflected forms are lexically stored; on the other hand, priming studies (e.g., Crepaldi et al., 2016; Marslen-Wilson & Tyler, 1998; Morris & Stockall, 2012) have shown that there are priming effects between irregularly inflected and uninflected forms, and that these effects are independent of semantic priming, indicating that irregularly inflected forms are decomposed. The current paper focuses the production mechanism of morphological regular forms, in particular, whether regular forms *with productive phonological alternation* are computed or stored. This issue is closely related to the morphological regularity discussion above in the following ways. First, as reviewed above, there is competing evidence whether morphologically regular forms are computed on the fly or lexically stored. Furthermore, morphologically regular forms may involve phonological alternation, as exemplied by the allomorphs [t], [d], and [ɪd] in English regular past tense. Crucially, there are many different types of alternation attested in the world's language depending on whether it has lexical exceptions, is variable, or applies productively to new words. The regularity and productivity of phonological alternation may have a similar effect on the production mechanism of the morphological complex words as morphological regularity, in that regular and productive phonological alternation favors computation, while irregular and unproductive alternation favors storage, yet this issue is understudied (see Arndt-Lappe & Ernestus, 2020 for a recent review). The current study aims to contribute to our understanding of the effect of the former type of alternation from the perspective of a well-known phonological alternation in Mandarin Chinese: Tone 3 sandhi.

1.1. Tone 3 sandhi in Mandarin Chinese

Tone sandhi is a phonological alternation of lexical tones according to the linguistic context in which the tone appears. Tone sandhi is a common occurrence in Chinese languages, all of which are tonal, and tone sandhi patterns differ greatly in their phonological properties and productivity in different languages (see J. Zhang, in press, for an overview). The current study focuses on a tone sandhi pattern in Mandarin Chinese, commonly known as Tone 3 sandhi, which changes a Tone 3 (T3; a low falling-rising tone) to a Tone 2 (T2; a high rising tone) before another T3 syllable (e.g., 忍者/ren**3** zhe3/[1] 'ninja' becomes homophonous to 仁者 [ren**2** zhe3] 'benevolent man').[2] This alternation is highly regular in Mandarin disyllabic T3 sequences regardless of the morphosyntactic structure of the sequence, and previous studies have consistently shown that native Mandarin speakers can apply T3 sandhi to several types of novel disyllabic words, including meaningless combinations of real morphemes (尺洒 /chi3 sa3/ 'ruler' 'to spill') (Xu, 1997; J.; Zhang & Lai, 2010; Chen et al., 2019), accidental gaps where the combination of an existing base syllable and T3 yields no meaningful morpheme (e.g.,/pan3 mai3/) (J. Zhang & Lai, 2010; Chen et al., 2019), and novel syllables that are non-existing combinations of legal initials and rimes in Mandarin Chinese (e.g., /fia3 sua3/) (C. Zhang & Peng, 2013; Chen et al., 2019). These findings provide compelling evidence that T3 sandhi is a regular and productive phonological alternation in Mandarin Chinese.

Two mechanisms have been proposed to subserve the production of tone sandhi: a phonology-based computation mechanism and a word-based (or storage-based) lexical mechanism (Hsieh, 1970, 1975, 1976; X.; Wang, 1993; J.; Zhang & Lai, 2010; J.; Zhang et al., 2011; Yuan & Chen, 2014; Li & Chen, 2015; C.; Zhang et al., 2017). The computation mechanism computes the surface forms of an underlying representation according to the phonological contexts. The lexical mechanism, on the other hand, retrieves the word with the context-appropriate allomorph stored in memory. These two mechanisms are also associated with different assumptions about the nature of the mental representation of spoken words: the computation mechanism assumes that words are stored in their underlying forms, whereas the lexical mechanism assumes that words are stored in their surface forms. The productivity of Tone 3 sandhi in novel words in Mandarin Chinese, especially those involving novel syllables in the sandhi position, suggests that Mandarin speakers must be able to use the computation mechanism for Tone 3 sandhi. But whether the computation mechanism is only restricted to novel words

[1] The four tones in Mandarin are simply transcribed as 1, 2, 3, and 4 for brevity here. Tone 1 is a high level tone, Tone 2 is a high rising tone, Tone 3 is a low dipping tone, and Tone 4 is a high falling tone. In the tradition of Chao's tone letters (Chao, 1930), which range from 1 to 5, with 1 referring to the lowest pitch and 5 referring to the highest pitch, the four tones are often transcribed as 55, 35, 213, and 51, respectively. The segments are transcribed in Pinyin, which is the official romanization system for Standard Mandarin Chinese.

[2] The neutralization between Tone 3 and Tone 2 is incomplete acoustically, with the sandhi Tone 3 having a slightly lower pitch than Tone 2, based on both experimentally elicited and corpus data (e.g., Peng, 2000; Yuan & Chen, 2014). This small acoustic difference does not influence conscious identification (Peng, 2000), but whether it can be used in online processing remains an open question.

without existing allomorphs or is applicable to both real and novel words remains an open question. For real words, while it is certainly possible that Mandarin speakers compute and convert an underlying T3 to T2 given the tonal context on the fly, it remains a possibility that they are stored with their surface allomorphs in long-term memory and produced directly from these surface allomorphs, especially for highly frequent words (e.g., 可以 /ke3 yi3/ 'can', pronounced as [ke2 yi3]). This is in line with the findings that certain familiar expressions can be stored as a chunk in rote memory, which differs from the propositional language that assembles grammatical units according to rules (Van Lancker Sidtis, 2012; Van Lancker Sidtis & Postman, 2006).

Behavioral studies on the spoken word recognition of Mandarin T3 sandhi words have shed some light on their representation and processing. Using an auditory-auditory priming lexical decision paradigm, Chien et al. (2016) found that spoken disyllabic words with T3 sandhi (e.g., 辅导 /fu3 dao3/ 'counsel', pronounced as [fu2 dao3]) were significantly primed by a preceding monosyllabic word that shared the pronunciation of the underlying form of the target's first syllable (辅 /fu3/ 'guide') instead of one that shared the pronunciation of its surface form (服 /fu2/ 'assist'), and this pattern was not significantly moderated by word frequency. This result suggests that, regardless of word frequency, disyllabic T3 sandhi words are likely represented in the underlying /T3-T3/ form, and this is the form that is first accessed in spoken word recognition by Mandarin speakers. Meng et al. (2021) used cross-modal semantic priming to investigate the representation of T3 sandhi words in Mandarin. Their results indicated that both T3 and T2 induced semantic activation of disyllabic T3-sandhi words (e.g., both 打 /da3/ 'to beat' and 答 /da2/ 'to answer' activated 打扫 /da3 sao3/ 'to clean' and induced semantic priming of 清理 /qing1 li3/ 'to tidy up'); but only T3, not T2, induced semantic activation of T3-initial words in non-sandhi contexts (e.g., 脑 /nao3/ 'brain', but not 挠 /nao2/ 'to scratch', primed 头部 /tou2 bu4/ 'head'). The authors argued that these results are consistent with the view that T3 is fully specified in its underlying tone, and T2 is derived by a rewrite rule only under the appropriate context. Both studies offered support for the computation mechanism of T3 sandhi. But the evidence is indirect, as the studies only investigated the perception, not production processing, of the alternation.

The production processing of T3 sandhi words has been investigated behaviorally using the Picture-Word Interference paradigm, in which a written word interferes with the naming of a picture at various stimulus-onset asynchronies (SOAs). Nixon et al. (2015) showed that when a disyllabic T3 sandhi word (e.g., 辅导 /fu3 dao3/ [fu2 dao3] 'to tutor') was the naming target, both a T3 distractor (e.g., 斧 /fu3/) and a T2 distractor (e.g., 服 /fu2/) facilitate the naming compared to a control distractor (e.g., 付 /fu4/) when the distractor appeared at the same time as the picture (SOA = 0); but when the distractor appeared 83 ms after the onset of the picture (SOA = 83 ms), only the T2 distractor facilitates the naming. Nixon et al.'s interpretation of the findings is that both the canonical and sandhi variants are represented in Mandarin T3, and both levels of representation are activated in the production of Mandarin tones. But there is a timing difference in the time course of activation of the two levels in that the sandhi (T2) variant is more important closer to the overt production of the T3 sandhi words.

Several mismatch negativity (MMN) studies have also examined the representation of T3 and T3 sandhi words, with inconclusive results. Li and Chen (2015) found asymmetrical MMN patterns in the perceptual processing of Mandarin T3, in that MMNs were reduced when monosyllabic T3 words served as the standards while T2 served as the deviants (e.g., in oddball sequences like *ma3 ma3 ma3 ma3 ma3 ma3* ***ma2*** *ma3 ma3 ma3 ma3* ***ma2*** …), compared to the reversed condition (where T2 syllables served as standards and T3 syllables as deviants). This asymmetry was interpreted as the co-activation of both the underlying and surface forms of T3 when T3 served as the standards, implying that the surface/sandhi form of T3 may be stored together with its underlying/non-sandhi form in the brain. Importantly, they only observed this asymmetry in the contrast between T3 and T2 syllables, and not in the contrast between T3 and T1, suggesting that it is specifically due to the alternation relation between T3 and T2. However, a similar study by Politzer-Ahles et al. (2016) found asymmetrical MMN patterns not only in the contrast between T3 and T2, but in all contrasts involving T3 (T3 vs. T1, T3 vs. T2, and T3 vs. T4). They argued that Mandarin T3 is underspecified (Lahiri & Reetz, 2008, 2010), which weakens the contrast with other tones when it serves as the standard, thus eliciting asymmetrical MMNs. Chien et al. (2020) embedded monosyllabic T2 deviants (竹 /zhu2/) in different conditions of disyllabic standards, which included T2-T4 (竹叶 /zhu2 ye4/ 'bamboo leaf'; T2 condition), T3-T4 (主页 /zhu3 ye4/ 'cover page'; T3 condition), and T3-T3 (主演 /zhu3 yan3/ [zhu2 yan3] 'to star'; Sandhi condition). Crucially, due to T3 sandhi, the first syllable of the T2 condition had the same rising contour as the first syllable of the Sandhi condition. Chien et al. (2020) found MMN on the deviant for the T2 and T3 conditions, which can be accounted for by the acoustic difference in f0 (T2 condition) or the tonal category difference (T3 condition) between the first syllables in the standards and the deviants in these conditions. However, the Sandhi condition, which had the same f0 pattern on the first syllable to the T2 condition in the standards, did not yield an MMN. Chien et al. offered two possible accounts for the data pattern: (a) the T3-T3 sandhi words are underspecified for tone in the first syllable, which caused reduced MMN on the deviant; (b) the alternation in the Sandhi condition exerts an influence in passive listening and disrupts an expected MMN. Under either account, T3 sandhi words are not represented as T2-T3. The reduction of expected MMN by T3 sandhi was also observed in Zeng et al.'s (2021) MMN study with disyllabic standard and deviant words in Mandarin, which showed that an expected segmental MMN was only observed when neither the standards nor the deviants involved T3 sandhi, suggesting that discovering the underlying representation in an alternation disrupts MMN. It is worth noting that although both Politzer-Ahles et al. (2016) and Chien et al. (2020) considered underspecification as a viable account for their results, Meng et al. (2021) provided a rebuttal against underspecification based on their cross-modal priming results.

In an event-related potentials (ERPs) study that specifically targeted T3 sandhi production, C. Zhang et al. (2015) compared the ERPs elicited when speakers produced disyllables that did or did not require tone sandhi, including both real words and novel words. Their experiment used minimal pairs of T3-T3 vs. T2-T3 disyllabic words (e.g., 赌瘾 /du3 yin3/ 'gambling addiction' vs. 毒瘾 /du2 yin3/ 'drug addiction'), with the former requiring T3 sandhi, while the latter did not. Native Mandarin speakers were instructed to listen to the two syllables spoken in isolation, one after another, in their underlying forms, and then concatenate the two syllables into a disyllabic word and produce the word as soon as the second syllable was heard, first in covert production to minimize the influence of motor commands on the ERPs, and then in overt production in order to verify that T3 sandhi has applied. Critically, significant effects of sandhi encoding were observed 230–320 ms after the onset of the second syllable, at which time the T3-T3 sequences elicited greater amplitude than T2-T3 sequences, irrespective of whether they were real or novel words. The time window in which this effect was observed is commonly associated with the P2 component of the ERP. As P2 is sensitive to phonological processing (Crowley & Colrain, 2004; Landi et al., 2012), this result was interpreted as revealing a greater effort involved in the phonological processing to compute and convert the first T3 syllable into its sandhi (T2) form in production, for both real and novel words, which is consistent with the computation mechanism. However, there are a number of issues with the design of the study that make it difficult to unambiguously attribute this effect to the phonological/phonetic encoding effort in production. First, the perceptual decoding of the second syllable temporally overlapped with the planning of T3 sandhi production in this experiment. Therefore, it is difficult to tease apart these two processes. Second, due to the prevalence of homophonous morphemes in Mandarin, the participants sometimes found it difficult to determine what the intended word was. Additionally, by employing a set of novel words in the production task (despite being presented in separate blocks from the real words), the study may have encouraged the speakers to adopt an overall computational strategy, bypassing any lexical retrieval even for the real words (C. Zhang et al., 2015).

The review of the experimental literature on Mandarin T3 sandhi above presents a complex picture on the representation of T3 sandhi words as well as the mechanism involved in their production. In terms of representation, although there is strong evidence that T3 sandhi words are represented as T3-T3, and that this is the representation first accessed by listeners upon hearing these words, whether the sandhi variant is also represented remains controversial, as indicated by the relevance of the sandhi variant in picture-word interference studies as well as the conflicting findings in MMN studies. In terms of the production mechanism, wug test results indicate that the computation mechanism must be available to native Mandarin speakers, but whether real words with T3 sandhi rely on the computation mechanism, and if so, how the mechanism is manifested neurophysiologically, require further research. The goal of the current study is to further shed light on the time course of the production mechanism of real words with T3 sandhi in Mandarin using ERPs.

1.2. *The current study*

The current study follows a similar design to C. Zhang *et* al. (2015) by using ERPs to investigate Mandarin T3 sandhi production. However, there are a number of crucial differences between the design of the current study and that of C. Zhang *et* al. (2015)'s.

As mentioned above, one of the concerns with C. Zhang *et* al.'s (2015) study was that the inclusion of novel words in the study may have encouraged the speakers to adopt an overall computational strategy. The current study focuses only on real word production to directly address the production mechanism of T3 sandhi in Mandarin real words. Crucially, we also introduced a word frequency manipulation in the current study. The effect of word frequency is well established in studies of word naming and recognition, in that the lexical access of more common words is faster than that of less common ones (Brysbaert et al., 2017; McRae et al., 1990). For ERPs, the inclusion of frequency can help shed light on the production mechanism in two ways. First, if the T3 sandhi production in real words relies on a lexical mechanism, we expect to observe a frequency effect in the time window relevant for lexical processing. According to Levelt and colleagues' model, this is an earlier time window that corresponds to lemma retrieval and the retrieval of the phonological code (~200–355 ms). Second, if the T3 sandhi production in real words relies on a computation mechanism, we should then expect an alternation effect in the time window of phonological computation *regardless* of frequency. It is assumed that this time window corresponds to the later time windows of phonological and phonetic encoding (~355–600 ms) in the serial model.

A complication in using a tone pattern to investigate the timing of production is that there are controversies regarding the inherent timing of tone encoding, i.e., whether tones are encoded late, comparable to prosodic features such as stress, which is assigned late as part of the prosodic frame of production (Chen, 1999; Roelofs, 2015), or encoded early, comparable to segments (e.g., Alderete et al., 2019; Wan & Jaeger, 1998). Evidence for the late encoding of tone comes from earlier reports of speech errors that, similar to stress errors, tone errors are rare (e.g., Chen, 1999). However, later analyses of speech error corpora have shown that tone errors are in fact not rare, and that tone errors are often characterized by contextual effects and interactive spreading effects (increased error rate when the intended word and words in the environment share phonological features), both of which are hallmarks of segmental errors and early activation (e.g., Alderete et al., 2019). Moreover, in Mandarin, T3 sandhi consistently occurs when tone errors provide the context of T3 sandhi, indicating that (a) lexical tones are accessed early, and (b) tone sandhi applies late (Wan & Jaeger, 1998). These types of evidence seem to suggest that the encoding of tone and tonal alternation may not be fundamentally different from that of segments and segmental alternation. But we recognize the issue here, and we interpret our results later in the context of this complication.

The current study also addresses two other methodological limitations of C. Zhang *et* al.'s (2015) study. The first limitation, as discussed above, was that speech processing of the second syllable temporally overlapped with the encoding of T3 sandhi in production in the previous study, which made it difficult to attribute the observed P2 effect to phonological/phonetic encoding effort. To remedy this, the onset of sandhi encoding is separated and postponed until after the offset of the second spoken syllable, to be cued by a visual prompt, in the current study. To ensure that the participants only start encoding T3 sandhi upon seeing the prompt (as opposed to starting as soon as they hear the second syllable), a distractor loudness judgment task (to compare whether the two isolated spoken syllables have identical or different loudness levels) cued by a different prompt is intermixed with the production task randomly, to render the production task unpredictable. The same set of words occurs once in the production task and once in the loudness judgment task. The second limitation was that the prevalence of homophones in Mandarin made the task difficult for the participants. This is addressed in the current study by presenting the corresponding simplified Chinese character on the computer screen simultaneously as each auditory stimulus is presented to facilitate the recognition of the target word. This crucially allowed us to achieve a more accurate control of the intended word frequency difference.[3]

To summarize, we adopted a *frequency* (high vs. low frequency) × *sandhi* (T3-T3 words vs. non-T3-T3 words) design in the current study. Two sets of predictions on the ERP signals time-locked to the covert production prompt can be formed based on the production mechanism of T3 sandhi words. According to the view of the lexical mechanism, we would expect an effect of *frequency* in a relatively early time window of the ERP signals that corresponds to lexical processing (~200–355 ms), but we would not expect an effect of *sandhi*, i.e., whether the word undergoes T3 sandhi or not, in a relatively late time window that corresponds to phonological and phonetic encoding (~355–600 ms). According to the view of the computation mechanism, we would expect an effect of *sandhi* in the late time window of phonological and phonetic encoding. Furthermore, this effect should not be regulated by *frequency*; i.e., we should not expect a *sandhi* × *frequency* interaction in this time window. Based on earlier findings on the productivity of T3 sandhi (J. Zhang & Lai, 2010; C. Zhang et al., 2015; Chen et al., 2019), the representation of T3 sandhi words accessed in spoken word recognition and passive listening (Chien et al., 2016, 2020), as well as the greater effort of producing T3 sandhi words identified in the P2 time-window of ERP signals (C. Zhang et al., 2015), we hypothesized that the second set of predictions should be borne out. Another potential set of predictions is based on the possibility that the lexical mechanism is in operation for high-frequency words and the computational mechanism for low-frequency words. This possibility will be further addressed in the Discussion.

To verify that the current design served the purpose of separating the perceptual processing and the production planning of the presented stimuli, we also make the following predictions on the ERP signals time-locked to the onset of the second syllable. First, we expect a *frequency* effect on the N400 component, in particular, a lower N400 amplitude for high-frequency than low-frequency words, in line with both the general findings on frequency effects on N400 in auditory and visual modalities (e.g., Rugg, 1990; Kutas & Federmeier, 2000, 2011) and findings specific to the reading of Chinese words (e.g., Liu & Perfetti, 2003; W.; Wang et al., 2017; Y.; Wang et al., 2021). Second, we expect a *sandhi* effect on the N400 component, with a stronger negativity for sandhi words than non-sandhi words, based on the property of the auditory stimuli. One reason that led to this hypothesis is that the concatenation of two T3 syllables, even with a 500 ms interval, may be considered by the Mandarin listeners as ill-formed due to the obligatory Tone 3 sandhi. The effects of phonotactics on N400 have been well-documented, and results have generally shown that phonotactically illegal sequences lead to a stronger N400 (e.g., Domahs et al., 2009; Ulbrich et al., 2016; White & Chiu, 2017; Wiese et al., 2017). Therefore, the sandhi words may accordingly elicit a stronger N400. Another reason is that there is an incongruity between the auditory stimuli and the participants' expectation of how the written disyllabic word should be pronounced in the sandhi condition. This may cause difficulties in the processing of the sandhi words. Given that higher amplitudes of the N400 are thought to reflect the difficulty of processing and higher cognitive processing loads (Kutas & Federmeier, 2000, 2011), we expect the processing difficulties of the sandhi words to be manifested in a strong N400 as well.

2. Materials and methods

2.1. Participants

Thirty-two participants were recruited to participate in the experiment (17 men, 15 women, mean age = 22.3, SD = 4.3, range = 17–33 years old). All participants were native Mandarin speakers from Northern China, right-handed, with normal hearing. The participants reported less than 5 years of musical training and no history of neurological disorders. They were paid a nominal fee for their participation in the experiment. Two participants were excluded from analysis as a result of excessive artifacts, leaving 30 participants in the final analysis (see EEG recording and data analysis below for details).

2.2. Stimuli

A *word frequency* (high vs. low frequency) × *sandhi encoding* (T3-T3 vs. non-T3-T3) design was adopted in the present study. Forty (40) high-frequency and 40 low-frequency disyllabic Mandarin words that undergo T3 sandhi and 40 high-frequency and 40 low-frequency disyllabic Mandarin words that do not undergo T3 sandhi were selected as the production stimuli (see the Appendix for the word list). The frequency counts (words per million) of the compound words and each character/morpheme within the compound words were obtained from the Mandarin SUBTLEX corpus; in addition, subjective familiarity ratings of the compound words on a scale of 1–9 (with 9 indicating highly familiar and 1 indicating highly unfamiliar) were obtained from 12 Mandarin speakers (7 men, 5 women) who did not participate in the EEG experiment. High-frequency sandhi (mean = 138.62, SD = 369.08) and non-sandhi words (mean = 72.49, SD = 59.04), as well as the low-frequency sandhi (mean = 0.05, SD = 0.02) and non-sandhi words (mean = 1.08, SD =

[3] We also conducted a parallel experiment in which only the spoken syllables were presented without the simultaneous display of Chinese characters. Overall, the EEG data were noisier with higher artifact rejection rates compared to the experiment with simultaneous auditory and visual presentation of the syllables/characters reported here. Some subjects also reported the tasks to be challenging, which may have contributed to the noisier data. The results of this experiment are not reported here.

0.52), were roughly matched in aggregate in terms of the compound word frequency (see the Appendix for frequency data). Subjective familiarity rating followed a similar pattern (high-frequency sandhi words: mean = 8.60, SD = 0.21; high-frequency non-sandhi words: mean = 8.64, SD = 0.17; low-frequency sandhi words: mean = 5.37, SD = 0.98; low-frequency non-sandhi words: mean = 6.28, SD = 0.80). But it is worth noting that, for the frequency data, there is a bigger frequency difference between high- and low-frequency words for sandhi words than non-sandhi words; for the rating data, the rating difference between sandhi and non-sandhi words is bigger for low-frequency words than high-frequency words. These facts will play a role in the interpretation of our data.

A female native speaker of Mandarin from Jilin, China recorded the individual syllables of these words in a randomized order three times in a soundproof booth at the Hong Kong Polytechnic University. A clear and accurately produced token was selected for each syllable. Using Praat (Boersma & Weenink, 2021), all syllables were normalized to 400 ms in duration, and two versions were created with different loudness (70 dB and 80 dB) for the distractor loudness judgment task.

All of the experiment trials, including the trials for the distractor loudness judgment task, were presented twice, resulting in 640 trials in total (40 items × 4 conditions × 2 task types × 2 repetitions). They were divided into ten blocks, with 64 trials per block. Each block contained an equal number of trials for each condition as well as for each task.

2.3. Procedure

STIM2 (Neuroscan, Charlotte, NC, U.S.) was used for the control of stimulus presentation and the collection of behavioral responses. Fig. 1 illustrates the procedure. During a trial, two isolated spoken syllables in their canonical (non-sandhi) forms were presented to the participants via headphones, with a 500 ms interval in between; when each auditory stimulus was presented, the corresponding simplified Chinese character was simultaneously shown on the screen. After the offset of the second syllable, there was a 400–600 ms jittered interval before the onset of a visual prompt associated with one of the two tasks: a 'thinking' sign prompting the participants to covertly produce the two isolated syllables together as a whole word (i.e., the critical task), or a question mark prompting the participants to judge whether the two syllables were of equal loudness (i.e., the distractor task). The interval was jittered in order to mitigate any influence of the processing of the second syllable/character persisting into the ERPs of the critical task. Covert production was followed by overt production, which was cued by a microphone sign. Overt production was used to ensure that the participants had correctly encoded the sandhi and was inspected later to exclude trials with incorrect sandhi application from ERP analysis. The loudness judgment task served to make the production task unpredictable, so that production would only be initiated upon seeing the 'thinking' sign, rather than happening while the participants were still perceiving the auditory and visual stimuli. The same set of stimuli occurred twice in the production task and twice in the loudness judgment task. The two tasks occurred randomly with equal probability (50%) in a block.

2.4. EEG data acquisition

Electroencephalographic (EEG) data were recorded using SynAmpsRT amplifier (Neuroscan, Charlotte, NC, U.S.) with a cap carrying 64 Ag/AgCl electrodes placed on the scalp at specific locations according to the extended international 10–20 system. Vertical electrooculography (EOG) was recorded using bipolar channels placed above and below the left eye, and horizontal EOG was recorded using bipolar channels placed lateral to the outer canthus of each eye. The impedance between the reference electrode (located between Cz and CPz) and any recording electrode was kept below 5 kΩ. Alternating current signals (0.05–400 Hz) were continuously recorded and digitized at the sampling rate of 1000 Hz. A Cedrus StimTracker was used to record the onset of the second syllable as well as the covert-production prompt in each trial as events in the acquisition software Curry 7.

Two sets of ERP analysis were conducted, one time-locked to the covert production prompt, one time-locked to the auditory onset of the second syllable. The EEG signals were pre-processed with BESA Version 6.1 and were filtered with a 0.03–30 Hz band-pass zero-

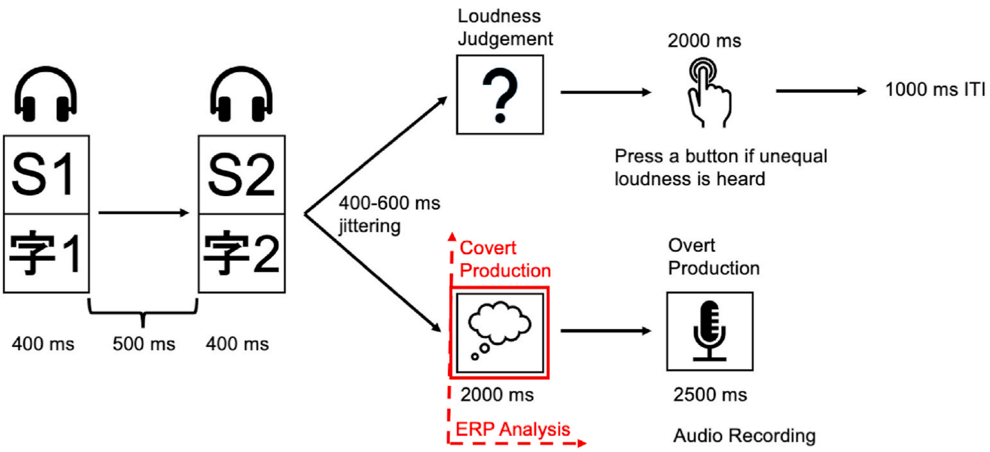

Fig. 1. Schematic representation of the experimental paradigm.

phase shift digital filter (slope 12 dB/Oct). Artifact correction was conducted using ICA by first defining eye artifacts according to the topographic distribution manually; artifacts were then corrected automatically. Baseline correction was performed relative to the onset of the covert production prompt or the onset of the second syllable within the -100 to 0 ms time-window, depending on which set of analysis was being conducted. Epochs ranging from −100 to 1000 ms after the onset were extracted for analysis of the ERP data time-locked to the covert production prompt; for the ERP data time-locked to the onset of the second syllable, however, we only analyzed the epoch up to 800 ms after the syllable onset, as this is the earliest time that the task prompt could appear. ERPs were re-referenced to the average of both mastoids for offline analysis. Epochs with amplitudes exceeding ±75 μV at any channel were excluded from averaging. Participants with fewer than 50% of critical trials remaining in their data due to either artifact rejection or incorrect production were excluded from subsequent analyses. For this reason, two participants were excluded, leaving 30 participants for the analysis of both the covert production and the second syllable. The mean and range for the number of accepted trials per participant in each condition for the covert production are as follows: sandhi high-frequency: 87.4% [100%, 57.5%]; sandhi low-frequency: 87.3% [100%, 57.5%]; non-sandhi high-frequency: 88.2% [100%, 55%]; non-sandhi low-frequency: 87.2% [100%, 57.5%]. The mean and range for the number of accepted trials per participant in each condition for the second syllable are: sandhi high-frequency: 86% [100%, 61.9%]; sandhi low-frequency: 85.6% [99.4%, 59.4%]; non-sandhi high-frequency: 85.7% [99.4%, 63.1%]; non-sandhi low-frequency: 85.5% [99.4%, 58.8%].

2.5. Data analysis

We used spatiotemporal cluster-based analysis (Maris & Oostenveld, 2007) implemented in the FieldTrip toolbox to test the effects of *sandhi* (T3-T3 vs. non-T3-T3), *frequency* (high vs. low), and their interaction. The time-window of the analysis was the entire post-stimulus epoch (0–1000 ms) for the ERPs time-locked to the covert-production prompt and 0–800 ms for the ERPs time-locked to the onset of the second syllable, as explained above. For each of the three effects above, a pairwise dependent *t*-test across the 30 participants was conducted at every sample and every channel. Clusters were formed out of samples that were spatially and/or temporally adjacent, had p-values of below 0.05 in the initial uncorrected test, and were neighbors of at least 2 other channels that also met this p-value criterion at the same time point. Cluster-level statistics were calculated by taking the sum of the t-values of all the data points within each cluster. The highest of those sums was used as the test statistic. This test statistic calculated from the actual observed data was compared to test statistics obtained by permuting the condition labels (within participants) 500 times and re-calculating the test statistic for each permutation. These parameters were chosen based on recommendations in the FieldTrip tutorial documentation (https://www.fieldtriptoolbox.org/tutorial/cluster_permutation_timelock/?s) and because they tend to yield a good balance between the ability to detect focal effects and the ability to detect broad effects. The proportion of permutation test statistics that was larger than the actual observed test statistic was the *p*-value for the test.

3. Results

3.1. ERPs time-locked to covert production prompt

For the ERPs time-locked to the covert production prompt, T3 sandhi words elicited significantly more positive ERPs than their non-sandhi counterparts ($p = .024$), and this difference was most prominent over front-central channels from about 350 to 500 ms after the appearance of the covert production prompt. This effect is illustrated in Fig. 2, which provides the waveforms for each condition at electrode Fz, the topographic distribution of the difference between sandhi and non-sandhi conditions, and a raster plot of the cluster for the sandhi vs. non-sandhi difference in the cluster-based permutation test. The raster plot indicates that the difference between the sandhi and non-sandhi conditions was driven by a cluster that lasted from 336 to 520 ms at medial fronto-central channels and briefly extended into more posterior channels at around 350 ms. (Note that, for this and all the following results, reference to “clusters” is not meant to imply that the effect is significant at these specific channels or times; see Maris and Oostenveld (2007), Sassenhagen and Draschkow, 2019, as well as https://www.fieldtriptoolbox.org/faq/how_not_to_interpret_results_from_a_cluster-based_permutation_test/ for further details.)

Furthermore, low-frequency words elicited significantly more negative ERPs than high-frequency words ($p = .002$), and this difference was largest from about 600 ms to the end of the time window and was distributed broadly over the whole scalp. The effect is shown in Fig. 3, which provides the waveforms for each condition at electrode Cz, the topographic distribution of the difference between high- and low-frequency conditions, and a raster plot of the cluster for the high-vs. low-frequency difference.

Crucially, the cluster-based permutation analysis did not identify any significant interaction between *sandhi encoding* and *frequency* in any time-window. Although the waveforms for electrodes Fz and Cz in Figs. 2 and 3 seem to suggest a numerical sandhi × frequency

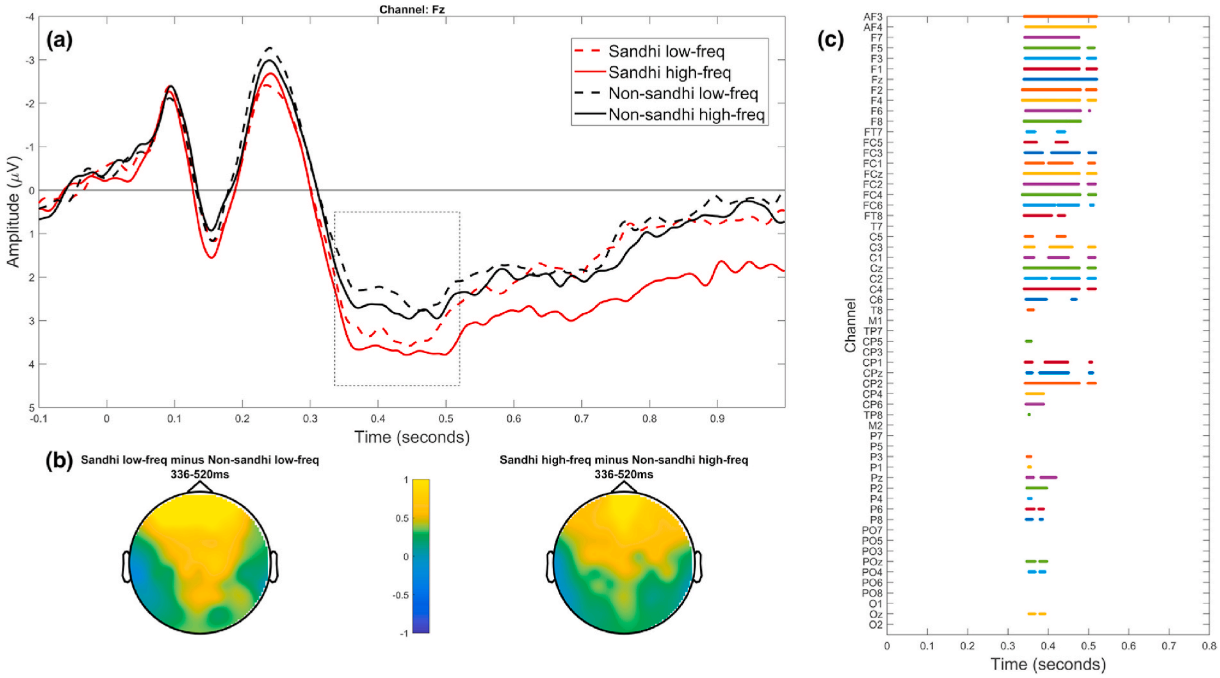

Fig. 2. Main effect of *sandhi encoding* (in ERPs time-locked to the onset of the covert production prompt [0 ms]) as revealed by the cluster-based permutation analysis. (a) Waveforms for each condition at electrode Fz. (b) Topographic distribution of the difference between sandhi and non-sandhi conditions. (c) Raster plot of the spatiotemporal extent of the cluster observed for the sandhi vs. non-sandhi difference in the cluster-based permutation test; each dot or line indicates a sample or series of samples that are part of the cluster.

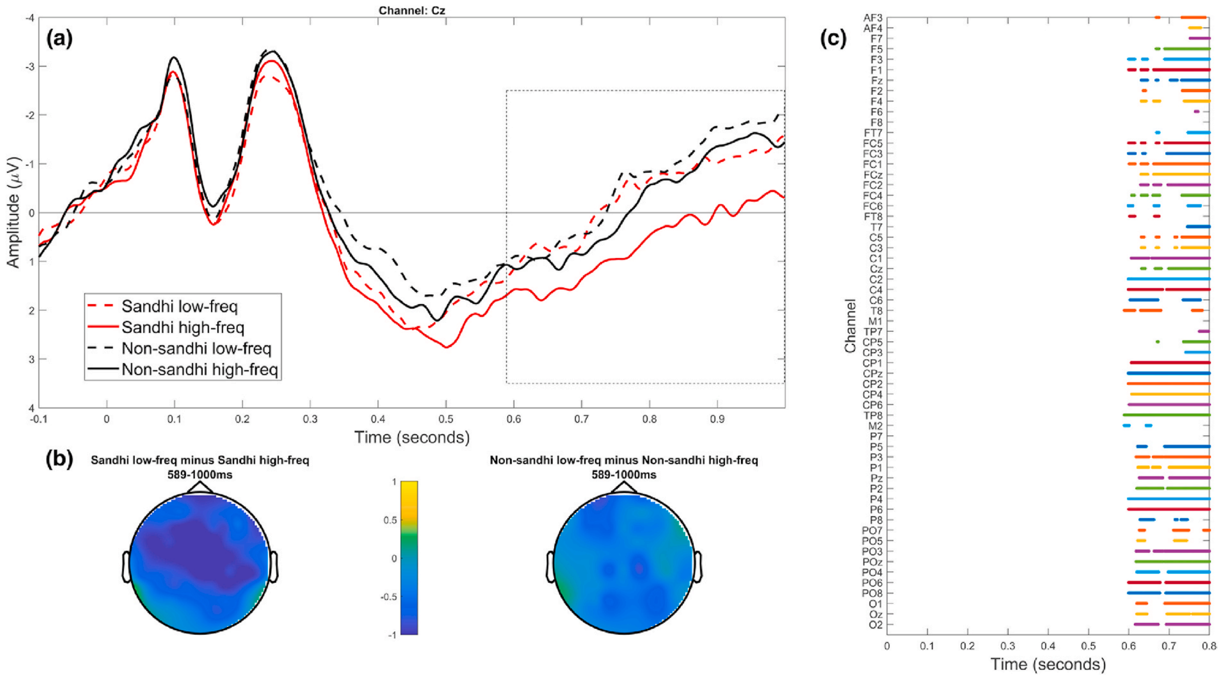

Fig. 3. Main effect of *word frequency* (in ERPs time-locked to the onset of the covert production prompt [0 ms]) as revealed by the cluster-based permutation analysis. (a) Waveforms for each condition at electrode Cz. (b) Topographic distribution of the difference between high-frequency and low-frequency conditions. (c) Raster plot of the spatiotemporal extent of the cluster observed for the high-frequency vs. low-frequency difference in the cluster-based permutation test; each dot or line indicates a sample or series of samples that are part of the cluster.

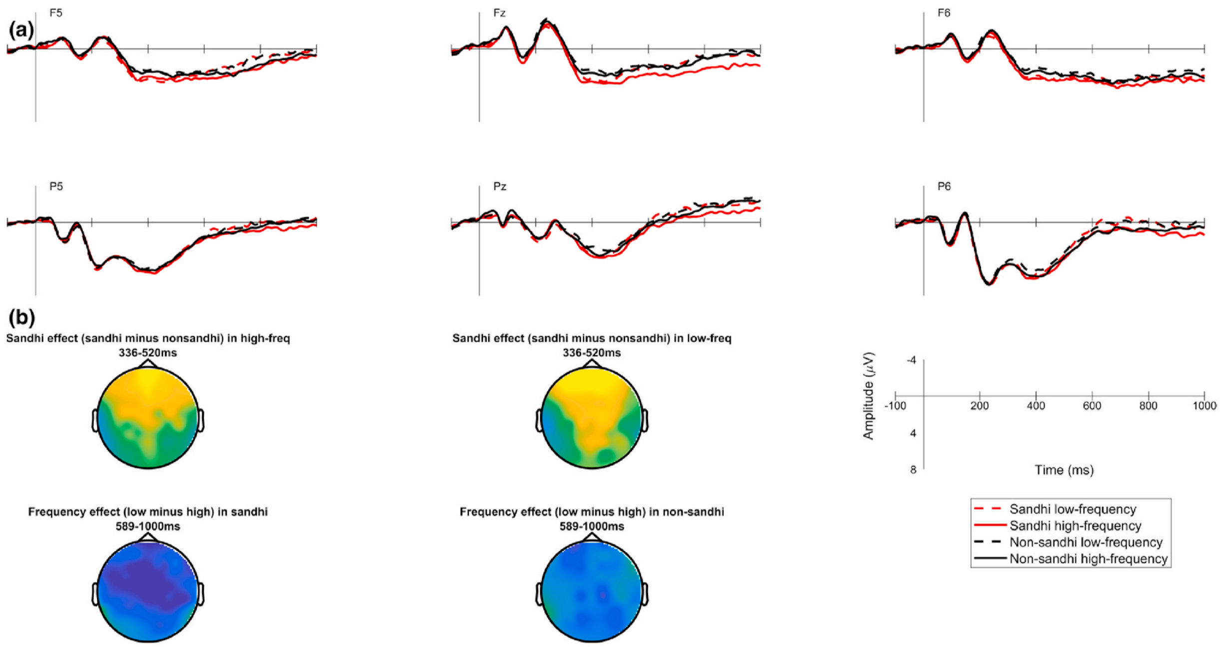

Fig. 4. ERPs waveforms and topographic plots for the sandhi effect in each level of frequency manipulation and for the frequency effect in each level of sandhi manipulation (in ERPs time-locked to the onset of the covert production prompt [0 ms]). (a) Waveforms for each condition at electrodes F5, Fz, F6, P5, Pz, and P6. (b) Topographic distributions of the sandhi effect for high- and low-frequency words respectively in the time-window of 336–520 ms and the frequency effect for sandhi and non-sandhi words respectively in the time-window of 589–1000 ms.

interaction pattern (from around 500 ms onwards after prompt onset, there appears to be a larger sandhi effect for high-frequency words than low-frequency words), no significant interaction was detected statistically. The lack of significant interaction can be more clearly seen in Fig. 4, which provides the waveforms for each condition at three frontal electrodes (F5, Fz, F6) and three posterior electrodes (P5, Pz, and P6) for broad scalp coverage, as well as the topography for the sandhi and frequency effects during the relevant windows. As can be seen in the waveforms, the numerical trend of a *sandhi* × *frequency* interaction was not consistently present across electrodes.[4]

3.2. ERPs timed-locked to the onset of the second syllable

We analyzed the ERP data time-locked to the onset of the second syllable in the same way as we analyzed the data time-locked to the covert production prompt, except that we only analyzed the epoch up to 800 ms after the syllable onset, as stated in "EEG data acquisition" above. Also, the analysis included both the production trials and the loudness judgment trials, as the participants were not aware of which task the trial belonged to when the second syllable was presented.[5] Analysis time-locked to the onset of the second syllable revealed that second syllables of sandhi-undergoing words elicited more negative ERPs than those of non-sandhi-undergoing words ($p < .001$). As shown in the raster plot in Fig. 5, the cluster where/when this difference was largest had a changing spatio-temporal pattern: on relatively frontal channels this difference was long-lasting (from about 200 to 800 ms), whereas at more posterior channels this difference was short-lived (from about 250 to 400 ms). N400s elicited by auditory stimuli tend to have earlier onset and longer duration, with a slightly more frontal distribution than visually elicited N400 (Kutas & Federmeier, 2011). The timing and scalp distribution of the effect here indicate that this is likely a combination of N400 and other components as a result of the complexity of the stimulus input as well as the challenges in accessing the lexical item presented by the different input modality.

As for frequency, second syllables of lower-frequency words elicited both significantly more negative ($p < .001$) and marginally

[4] One concern with the experimental design is that the repetition of the experimental items (twice in covert production, twice in loudness judgment) may have caused a reduction of the test effects. To address this concern, an analysis of the first repetition of the items in covert production was also conducted. The results are comparable to what has been found in the whole dataset. Words requiring T3 sandhi elicited significantly more positive ERPs than words not requiring T3 sandhi ($p = .0479$), and this difference was most prominent over fronto-central channels from about 390 to 600 ms after the appearance of the covert production prompt. Furthermore, low-frequency words elicited significantly more negative ERPs than high-frequency words ($p = .004$) from about 600 ms to the end of the time window. This difference was distributed broadly over the whole scalp, but was more prominent in centro-parietal electrodes. There was no significant interaction between sandhi and frequency.

[5] We conducted the same analysis for all trials and for only the trials with the production task. There were no robust differences in data pattern between the two analyses, indicating that the participants could not predict which task they would be asked to perform when the second syllable was presented for each trial. We opted to report the data analysis based on all trials here.

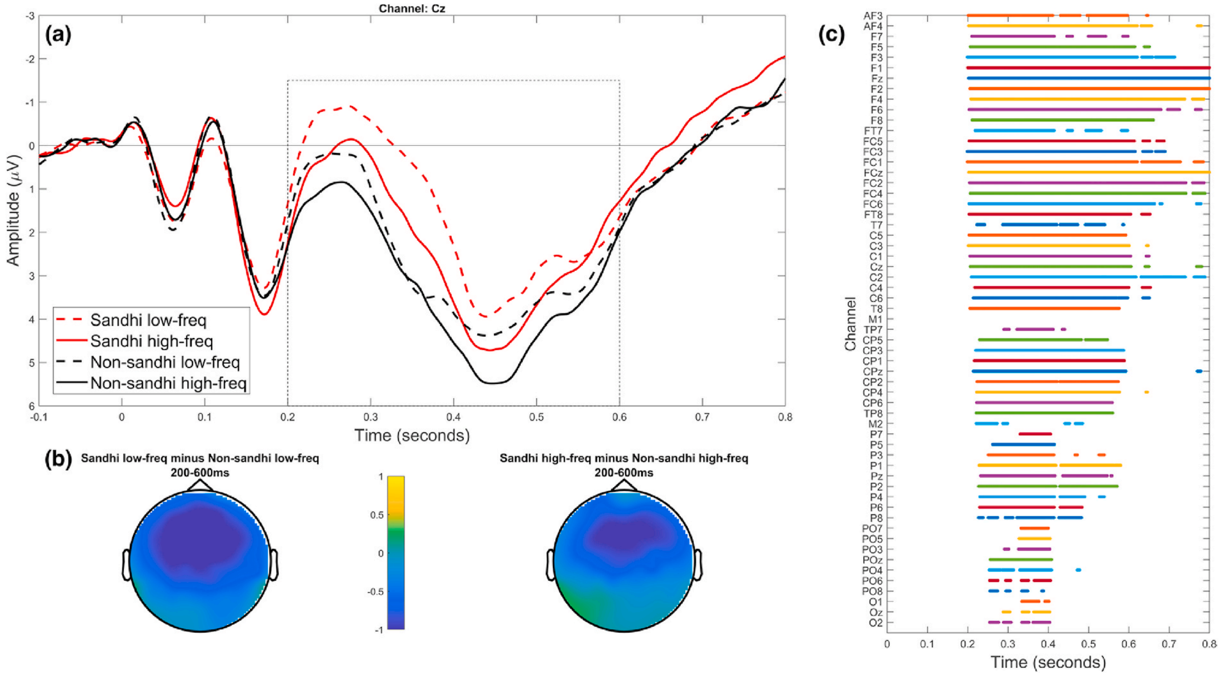

Fig. 5. Main effect of *sandhi encoding* (in ERPs time-locked to the onset of the presentation of the second syllable [0 ms]) as revealed by the cluster-based permutation analysis. (a) Waveforms for each condition at electrode Cz. (b) Topographic distribution of the difference between sandhi and non-sandhi conditions. (c) Raster plot of the spatiotemporal extent of the cluster observed for the sandhi vs. non-sandhi difference in the cluster-based permutation test; each dot or line indicates a sample or series of samples that are part of the cluster.

more positive ($p = .072$) ERPs than second syllables of higher-frequency words, although these effects had different spatiotemporal distributions. As for lower-frequency words eliciting a more negative ERP than higher-frequency words, this pattern was most prominent from about 200 to 500 ms and broadly distributed over the scalp. On the other hand, as for lower-frequency words eliciting more *positive* ERPs than higher-frequency words, this effect was most prominent after 600 ms and was limited to right posterior channels. In other words, the frequency manipulation elicited a biphasic effect somewhat reminiscent of a classic N400–P600 pattern. These effects can be seen in Fig. 6.[6]

Finally, we observed a significant interaction between *sandhi encoding* and *frequency* ($p = .012$). Specifically, the N400-like simple effect of sandhi, with sandhi-undergoing words eliciting more negative ERPs than non-sandhi-undergoing words, was bigger in low-frequency words than it was in high-frequency words. (Or, to put it another way: the simple effect of frequency, with low-frequency words eliciting more negative ERPs than high-frequency words, was bigger in sandhi-words than in non-sandhi words.) This pattern was most prominent for only a short period of time from about 300 to 400 ms and was broadly distributed, as shown in Fig. 7.

4. Discussion

4.1. Mandarin T3 sandhi: computation mechanism vs. lexical mechanism

In the current study, we tested the hypothesis that the production of real words with Mandarin Tone 3 sandhi is subserved by a computation mechanism, which predicts divergent ERP patterns between sandhi and non-sandhi words in a relatively late time-window of post-lexical phonological and phonetic encoding regardless of word frequency. Our ERP results time-locked to the covert production prompt support this hypothesis. We have observed a *sandhi* main effect, with the sandhi words eliciting more positive ERPs than non-sandhi words, and the sandhi effect appeared in a time window (about 350–500 ms) that roughly corresponds to the time window of phonological and phonetic encoding of Levelt and colleagues' model (~355–600 ms) (Hagoort & Levelt, 2009; Levelt, 1999; Schriefers et al., 1990). Moreover, this sandhi effect is not regulated by word frequency. These findings are exclusively predicted by the computation mechanism of T3 sandhi word production.

The lexical mechanism predicts a *frequency* effect in an early time window of the ERP signals (~200–355 ms) time-locked to the

[6] Baseline correction was implemented here, but its effect is difficult to see in Fig. 6 because of a strange pattern in the ERP: most of the points between −100 and 0 ms are above the zero line, but the ERP goes sharply below the zero line slightly before zero, and these points taken together average out to zero. We suspect that a reason for this slightly-before-zero effect is that the visual stimulus may have appeared on the screen slightly before the onset of the auditory stimulus (the latter was the time-locking point for the ERP analysis).

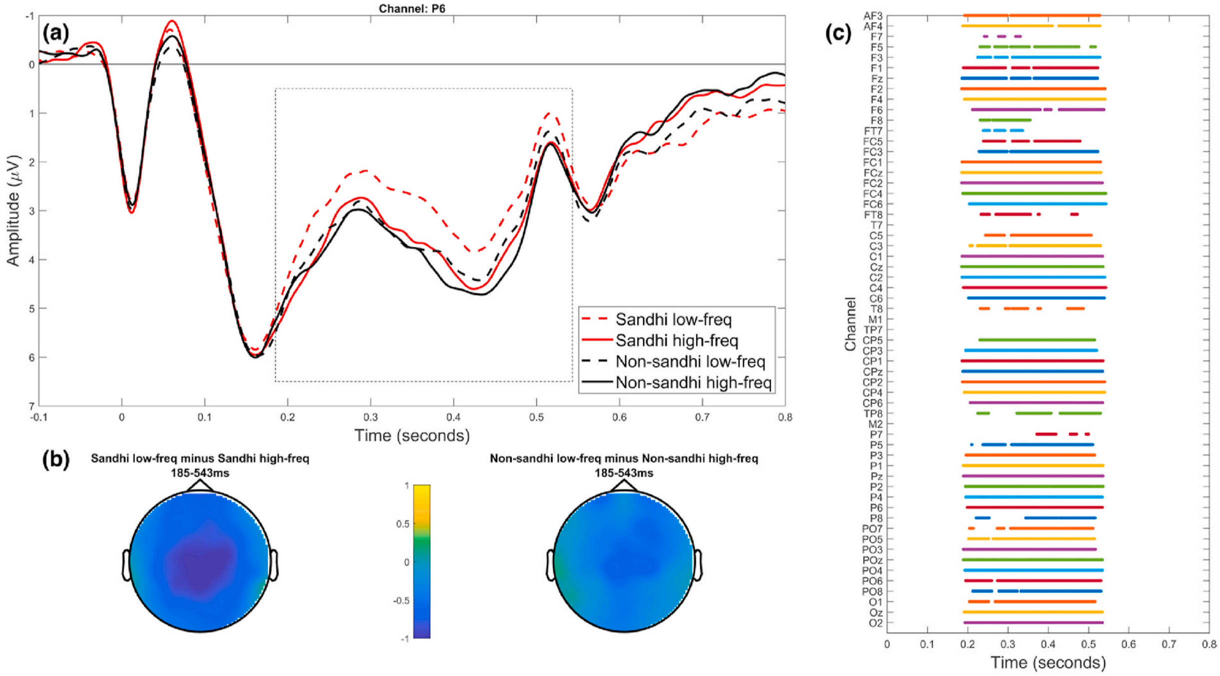

Fig. 6. Main effect of *word frequency* (in ERPs time-locked to the onset of the presentation of the second syllable [0 ms]) as revealed by the cluster-based permutation analysis – note that the topographic plot and raster plot are only showing the distribution of the early negativity, not the later positivity; the late positivity is visible in the waveform, though. (a) Waveforms for each condition at electrode P6. (b) Topographic distribution of the difference between high-frequency and low-frequency conditions in the earlier time window. (c) Raster plot of the spatiotemporal extent of the negative cluster observed for the high-frequency vs. low-frequency difference in the cluster-based permutation test; each dot or line indicates a sample or series of samples that are part of the cluster.

covert production prompt. A *frequency* effect was observed in our data, but in a very late time window (589–1000 ms), which is not only considerably later than the window expected for lexical processing, but also later than the window for phonological and phonetic encoding. Therefore, this frequency effect is very likely not the one predicted by a lexical mechanism. We surmise that this has likely stemmed from post-production processes such as self-monitoring or the rehearsal of the phonetic form in working memory. Self-monitoring is an integral part of speech production and forms the feedback loop that occurs rather late in time (e.g., 500–600 ms) (Hagoort & Levelt, 2009; Levelt, 1999; Schriefers et al., 1990). The delayed two-step production procedure employed in the current study to minimize EEG artifacts caused by production movement (first covert production and then overt production, each prompted by a separate cue) may have aggravated the need for rehearsal and self-monitoring, and that this process was impacted by word frequency.

If the production processing of T3 sandhi words involves not just the computation mechanism, but also the lexical mechanism depending on the frequency of the word, we would expect the *sandhi* effect to be regulated by *frequency* in the time window of phonological and phonetic encoding. In particular, the expectation is that higher frequency words are more likely to be stored with their surface allomorphs (Pan et al., 2020); therefore, the specific *sandhi* × *frequency* interaction predicted is that the sandhi effect would be greater for low-frequency words than high-frequency words. This interaction was not observed in our ERP data time-locked to the covert production prompt, despite the fact that the familiarity rating data for low-frequency words showed a bigger sandhi effect than those for high-frequency words, thus encouraging this effect to appear. We noted earlier that there seemed to be a numerical trend of *sandhi* × *frequency* interaction on some of the reported electrodes (e.g., Fz and Cz in Figs. 2 and 3). But even if this interaction were significant, it would not be in the direction predicted by the dual mechanism above, as the exhibited pattern showed a greater sandhi effect for high-frequency words than low-frequency words.

Another possible interpretation for the *sandhi* effect is that it may have reflected a greater effort in retrieving the appropriate phonological/phonetic code because of competing codes in the case of T3-T3 words compared to non-T3-T3 words. We did not see an earlier *frequency* effect or a *sandhi* × *frequency* interaction because the conflict was not present in the lemma (only one lemma), but in the phonological/phonetic codes (underlying and surface variants). But comparing our results with C. Zhang *et* al. (2015) results, in which they showed a main *sandhi* effect for *both real and novel words*, we believe that a more favorable interpretation of the *sandhi* effect in the current study is still that it reflects the greater effort in computing the sandhi, as the novel words in C. Zhang et al.'s study presumably did not have competing codes, yet exhibited a similar sandhi effect as in the current study.

Given that the production in this experiment was elicited by a paradigm that is quite different from natural speech or a naming experiment, the specific timing of the effects and its correspondence to the timing of the production events suggested by Levelt and colleagues' model warrant further discussion. It is possible that, when the word to be produced is triggered by hearing and seeing the

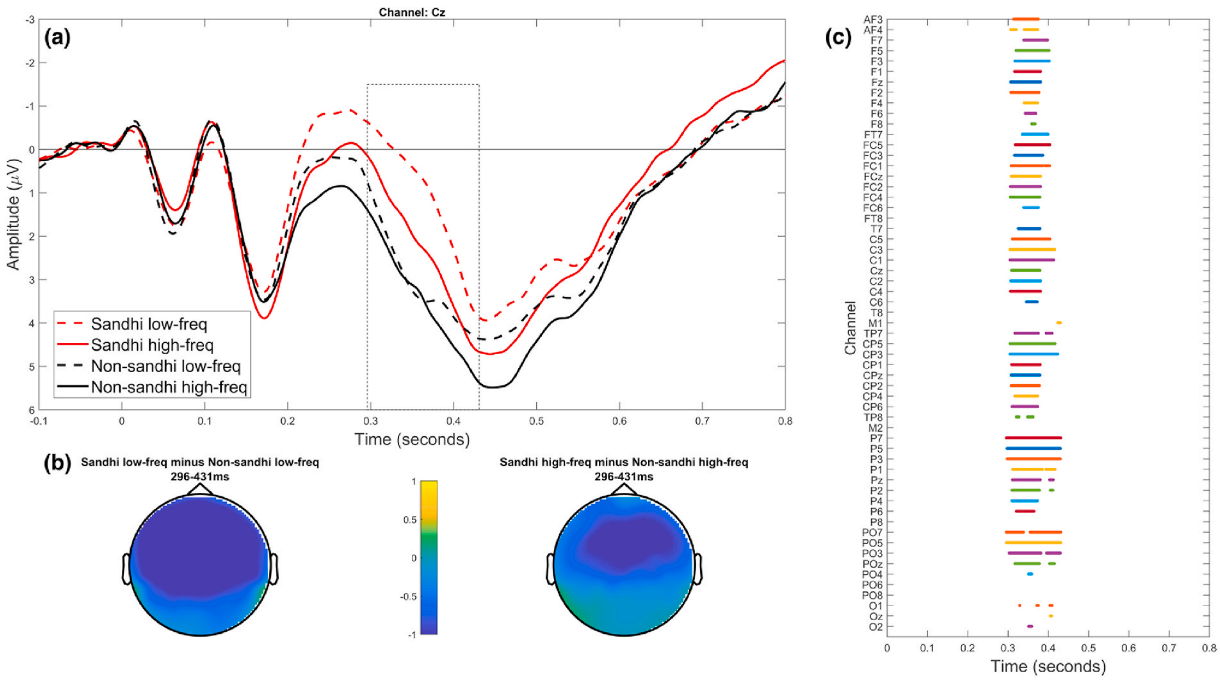

Fig. 7. Interaction effect of *sandhi encoding* and *word frequency* (in ERPs time-locked to the onset of the presentation of the second syllable [0 ms]) as revealed by the cluster-based permutation analysis. (a) Waveforms for each condition at electrode Cz. (b) Topographic distribution of the difference between sandhi and non-sandhi words, among low-frequency words (left topoplot) and among high-frequency words (right topoplot). (c) Raster plot of the spatiotemporal extent of the cluster observed for the interaction in the cluster-based permutation test; each dot or line indicates a sample or series of samples that are part of the cluster.

characters that it comprises rather than a natural speech context or a picture, the processes involved in its production may be different from those suggested in the model. But the difference is most likely in the early stages of concept formation and lemma retrieval, while the retrieval of the phonological code and subsequent phonological and phonetic encoding would still need to be involved in a similar way. The delay in the covert production to the visual prompt also may cause the timing of the production events to shift earlier due to the time that the subjects were afforded during which some of the processes may have occurred. Finally, as we recognized in the *Introduction*, even though there is evidence (e.g., from speech errors) that the encoding of lexical tone and tonal alternation may not be different from phonological segments and segmental alternation, the timing of tone encoding in production is inherently controversial. Taken these together, the fact that the timing of the sandhi effect observed in the current study overlapped with the timing of phonological and phonetic processing in the serial model may be less meaningful, but the fact that it occurred relatively late (and without a modulation of frequency differences) can still only be interpreted as a phonological processing effect.

It is difficult to compare the timing of the sandhi effect observed here (336–520 ms) with the P2 effect (230–320 ms) observed in C. Zhang *et* al. (2015) study due to the myriad differences in the methods between the two studies: the current study involves Chinese characters and a distractor task, and the ERP analysis is time-locked to the visual cue eliciting sub-vocal production rather than the beginning of the auditory presentation of the second syllable. Amplitude-wise, the current study found that the T3 sandhi words elicited more positive ERP amplitudes than non-sandhi words. Like C. Zhang et al., we interpreted this as T3 sandhi words requiring additional effort in their production processing. Topography-wise, the sandhi effect here is observed primarily in fronto-central electrodes, which overlap with the scalp distribution of the P2 where C. Zhang et al. observed the sandhi effect. Due to the poor spatial resolution of EEG and additional task differences, it is unclear where the neural sources are for the sandhi effect in the current study and the P2 in C. Zhang *et* al. (2015), or how they relate to the results of fMRI studies by H.-C. Chang et al. (2014) and C. H. C. Chang and Kuo (2016) that suggest that T3 sandhi production was mediated by activities in the right posterior IFG.

In order to separate perception from production processing, a delay in covert production was implemented. In the ERP data time-locked to the onset of the second syllable, we observed both a sandhi effect and a frequency effect in the general N400 component, with sandhi words and low-frequency words eliciting more negative ERPs. These observations are consistent with the interpretation that perceptual processing of the word, rather than production planning, occurred as the second syllable unfolded. This is also consistent with the anecdotal evidence that, in post-experiment debriefing conversations with the experimenters, participants generally stated that they did wait till the visual prompt to do covert production. We also observed a short-lived interaction between *sandhi encoding* and *frequency*, with the effect of sandhi being bigger in low-frequency words than it was in high-frequency words (or, with the effect of frequency being bigger in sandhi words than non-sandhi words). From the perception processing perspective, the direction of this effect is surprising, as one would expect the incongruity between an auditory T3-T3 sequence and an expected T2-T3 pronunciation of a sandhi word to be greater for high-frequency words. But from the sandhi and frequency main effects stated above, we also do not believe that this is a reflection of production processing. We tentatively surmise that this interaction may have stemmed from the greater familiarity rating difference between sandhi and non-sandhi words in low-frequency words than high-frequency words. Moreover, it is also consistent with the frequency data that the sandhi words had a bigger frequency difference between high- and low-frequency than non-sandhi words.[7]

Therefore, we believe that our attempt to separate the perception and production processes by introducing a secondary task and a delayed production was successful. However, this delay may have affected our ability to detect frequency effects in the earlier time window of covert production, as lexical access may have occurred earlier, which would diminish the frequency effect that could be observed during covert production. Our analysis of the ERP signals time-locked to the onset of the second syllable showed that the frequency effect indeed occurred earlier. Therefore, although our attempt to separate the perception and production processes was successful, the design may have had the negative effect of making frequency effects during covert production difficult to observe. We believe that lexical access is still relevant during covert production due to (a) the simultaneous presentation of the Chinese characters, which facilitated word access, and (b) the frequency effect observed in the late time window. But to fully address this potential confound, the current study should be complemented by future picture-naming or word-naming studies that further tap into potential frequency effects and any interaction between frequency and sandhi words in production processing. The safest conclusion we can draw from our results, then, is that the computation mechanism must be relevant for the production of Tone 3 sandhi words, but whether the lexical mechanism is relevant still remains an open question.[8]

Our interpretation that the production of T3 sandhi words in Mandarin involves a computation mechanism is in line with behavioral findings that Mandarin T3 sandhi is highly productive in novel words (J. Zhang & Lai, 2010; C. Zhang & Peng, 2013; Chen et al., 2019) and that the underlying T3 is accessed in the spoken word recognition of T3 sandhi words regardless of the frequency of the word (Chien et al., 2016). It is also consistent with Chien et al.'s (2020) and Zeng et al.'s (2021) finding that the alternation in T3 sandhi words may potentially disrupt an expected MMN.

It is worth pointing out that the computation and lexical mechanisms are not necessarily entirely incompatible with each other. We have entertained the possibility that the production of T3 sandhi words involves both mechanisms depending on the frequency of the word above and argued that our results were not consistent with the view, but we also acknowledged that our method was in a stronger position to support the computation mechanism than to reject the lexical mechanism. It is also possible that the nature of the mechanism relates to the semantic transparency and syntactic structure of the item, such that semantically more opaque and more word-like (rather than phrase-like) items are more likely to engage the lexical mechanisms, although there is little evidence that these factors impact disyllabic T3 sandhi application and Mandarin, and our study was not designed to investigate these factors. Other positions that invoke elements of both the computation and lexical mechanisms include the multi-variant view of Nixon et al. (2015) discussed in the *Introduction*, which assumes that both the non-sandhi and sandhi variants are represented as Mandarin T3, and the allomorph selection view of Tsay and Myers (1996), which, for Mandarin, would mean that the non-sandhi and sandhi allomorphs of T3 morphemes are both listed in the lexicon, and the computation involves the selection of the appropriate allomorph in context. The current study is not in a position to differentiate among these alternatives, as it is not designed to test the nature of the computation. Future experimental studies using MMN, production ERP, or other methods to provide replication or new data insights will be necessary to address the methodological issues mentioned above and further test the relevance of the computation and lexical mechanism in Mandarin T3 sandhi production.

4.2. *The role of phonological alternation in production processing*

The larger theoretical question behind the current study is the general role of phonological alternation, whereby a single morpheme has multiple surface realizations predictable by phonological context (e.g., *walk[t], jog[d], trot[ɪd]*), in production processing. This is related to, but goes beyond the computation vs. lexical storage debate of morphologically complex forms and how morphophonological regularity (e.g., *walked* vs. *went*) bears on the issue. Whether the generative approach of treating the surface allomorphs as deriving from a unique, abstract underlying representation has psychological reality in speech production is an understudied issue (Arndt-Lappe & Ernestus, 2020). The Mandarin T3 sandhi studied here furnishes an example of an alternation pattern that is regular and highly productive, and our results provided evidence for the computation mechanism of this type of alternation in line with the generative analysis.

However, this is only one type of alternation patterns attested in the world's languages. Alternations that have lexical exceptions or apply variably are common. They have been shown to have various levels of productivity when tested with novel words, and they have figured prominently in the development of phonological theory (e.g., see Bybee, 2001; Boersma & Hayes, 2001; Hayes & Londe, 2006;

[7] As in production processing, a similar controversy in perception processing is when tone is accessed. The sandhi and frequency effects we observed here seem to suggest a similar timeline for tonal access to segmental access. But we recognize that our study is not designed to address this question, and we refer the readers to Shuai and Gong (2014), Gao et al. (2019) and references within for further discussions of this controversy.

[8] An additional concern about the effect of word frequency is that it may be confounded by differences in monosyllabic morpheme frequency during both perception and production processing. One indication that the frequency effects that we observed in both sets of ERP data (timed-locked to covert production prompt and onset of the second syllable) are related to word frequency, not morpheme frequency of the second syllable, is that, due to the difficulties in matching monosyllabic morpheme frequencies in the different conditions, the second-syllable morpheme frequency of the non-sandhi ~ low frequency condition is considerably higher than that of the other conditions (see Appendix). However, this large frequency difference did not correspond to any of the observed frequency effects in ERPs in meaningful ways.

Zuraw, 2010; Becker et al., 2011; Coetzee & Pater, 2011; Coetzee & Kawahara, 2013; among many others). The phonological analyses of these patterns run the gamut from relying primarily on input-output mapping to depending mostly on lexical listing of surface allomorphs. The production and perception processing of these patterns, however, remains largely unknown: the limited literature on the recognition of spoken words with variable surface forms returned conflicting results on whether frequency regulates the representation accessed in word recognition (e.g., Connine & Pinnow, 2006; Ranbom & Connine, 2007 vs.; Sumner & Samuel, 2005), and the production processing of words that involve variable or exceptional phonological processes is severely understudied.

Another type of alternation pattern is one in which the alteration itself is regular in the language, but lacks full productivity in novel items. A tone sandhi pattern in the Taiwanese Southern Min dialect of Chinese has been consistently shown to have this property (Hsieh, 1970, 1975, 1976; X.; Wang, 1993; J.; Zhang & Lai, 2010; J.; Zhang et al., 2011). These authors have argued that this is caused by the circular chain shift nature of the tone sandhi pattern, which renders the pattern phonologically "opaque" (Kiparsky, 1973), and the surface tones must be listed in the lexical representations despite the regularity of the tone sandhi pattern in the language. This analysis is supported by Chien et al, (2017) finding that the recognition of spoken Taiwanese disyllable words with tone sandhi was facilitated by monosyllabic primes that shared the surface tone, not the underlying tone, of the syllable undergoing tone sandhi, and that the priming effect was regulated by word frequency. The neural encoding of this type of alternation in perception and production, however, also remains unknown.

The Mandarin T3 sandhi study here, therefore, may also serve as a stepping stone for future studies on other alternation patterns that will further connect theoretical phonological research and phonological typology with neurolinguistic investigations. In so doing, it will provide us with a comprehensive picture on the role of phonological alternation in production processing in general.

5. Conclusion

The current study contributes to our understanding of the neural basis of word production involving phonological alternation, a phenomenon whereby a morpheme is pronounced differently depending on the phonological context in which it appears. In particular, the study focused on a regular and productive tonal alternation named Tone 3 sandhi in Mandarin and provided ERP data on the production of Mandarin real words involving this alternation. Our ERP results time-locked to the covert production prompt showed that T3 sandhi words elicited more positive ERPs than non-sandhi words in a time window that roughly corresponds to the time window of phonological and phonetic encoding of Levelt et al.'s serial model of production, and this sandhi effect was not regulated by word frequency. Moreover, there was no frequency effect in the early time window that corresponds to lexical processing. These findings are consistent with a computation mechanism, which incurs a greater effort for alternation, but not a purely lexical mechanism, for T3 sandhi word production; indirectly, it suggests that the production mechanism accesses an underlying representation of /T3-T3/ for T3 sandhi words. What awaits further research includes collecting additional neurolinguistic data elicited from different production methods on the same type of alternation as well as investigations into alternation patterns with different regularity and productivity. In this way, phonological theory can benefit from neurolinguistic data that provide more direct evidence for representation and processing, and neurolinguistic studies can benefit from the linguistic diversity attested in the world's languages/dialects.

Funding

This work was supported by the US National Science Foundation [BCS-1826547]; and the University of Kansas General Research Fund [2140080].

Data availability statement

The data that support the findings of this study are openly available in OSF at https://osf.io/f5ng3/?view_only=3210a7faca3b4e4989e7cf0076469e0f.

Author statement

Jie Zhang: Conceptualization, methodology, writing – original draft, writing – review & editing, supervision, project administration, funding acquisition.

Caicai Zhang: Conceptualization, methodology, formal analysis, data curation, writing – original draft, writing – review & editing, supervision, project administration.

Stephen Politzer-Ahles: Conceptualization, methodology, formal analysis, data curation, writing – review & editing, visualization, supervision.

Ziyi Pan: Formal analysis, investigation, data curation, visualization.

Xunan Huang: Formal analysis, investigation, data curation, visualization.

Chang Wang: Formal analysis, investigation, data curation, writing – review & editing, visualization.

Gang Peng: Conceptualization, methodology.

Yuyu Zeng: Writing – review & editing.

Declaration of competing interest

The authors declared no conflict of interest.

Acknowledgments

We are grateful to audiences at the Society for the Neurobiology of Language in Baltimore, MD (2017), the International Workshop on the Neurolinguistic Studies of Tone at Hong Kong Polytechnic University (2019), and the PhonolEEGy Workshop on Electrophysiology and Phonological Theory at Université Côte d'Azur, Nice, France (2020) for their helpful feedback. We also thank the native Mandarin speakers who participated in our experiments.

Appendix. stimulus list

- Frequency counts are word (character)-per-million
- Familiarity ratings are from 1 (lowest) to 9 (high).

Sandhi: high-frequency words

No.	Pinyin	Sinogram	Gloss	Frequency 1st character	Frequency 2nd character	Frequency disyllabic word	Familiarity
1	bao3 xian3	保险	insurance	848.2	281.4	51.7	8.3
2	bi3 ci3	彼此	both sides	80.6	598.8	57.2	8.6
3	biao3 yan3	表演	to perform	670.2	499.0	198.5	8.8
4	cai3 qu3	采取	to adopt	67.3	264.9	26.4	8.2
5	chan3 pin3	产品	product	176.4	305.7	26.2	8.7
6	chu3 li3	处理	to deal with	613.6	877.4	193.7	8.8
7	da3 du3	打赌	to bet	1692.2	115.2	47.6	8.3
8	da3 rao3	打扰	to disturb	1692.2	121.1	98.1	8.7
9	da3 sao3	打扫	to clean	1692.2	59.7	22.2	8.8
10	gan3 ran3	感染	to infect	1303.4	69.9	36.3	8.3
11	guan3 li3	管理	to manage	621.4	877.4	32.3	8.5
12	ke3 yi3	可以	can	4540.2	3939.0	2360.8	8.8
13	lao3 ban3	老板	boss	1124.8	178.4	99.8	8.8
14	lao3 shu3	老鼠	mouse	1124.8	55.9	36.8	8.5
15	liao3 jie3	了解	to understand	21220.5	781.8	301.9	8.8
16	li3 jie3	理解	to understand	877.4	781.8	152.8	8.7
17	ling3 dao3	领导	to lead	174.4	155.3	28.3	8.6
18	lv3 guan3	旅馆	hotel	144.3	150.1	51.3	8.3
19	mei3 hao3	美好	wonderful	972.7	8811.9	66.7	8.8
20	mei3 nv3	美女	beauty	972.7	1637.3	41.4	8.7
21	na3 li3	哪里	where	1269.8	4071.1	401.6	8.7
22	ni3 hao3	你好	hello	30845.4	8811.9	37.4	8.8
23	nv3 you3	女友	girlfriend	1637.3	857.1	35.8	8.6
24	shou3 zhi3	手指	finger	1614.1	470.1	29.9	8.7
25	wu3 ru3	侮辱	to insult	16.1	37.9	21.7	8.2
26	wu3 dao3	舞蹈	dance	487.4	67.1	83.9	8.5
27	xi3 zao3	洗澡	to take a shower	175.5	36.2	28.3	8.8
28	xiang3 fa3	想法	opinion	5282.8	1296.9	136.4	8.6
29	xiang3 qi3	想起	to remember	5282.8	2723.6	55.7	8.7
30	xiao3 zu3	小组	group	2123.5	245.2	49.3	8.4
31	xuan3 shou3	选手	contestant	542.7	1614.1	111.3	8.3
32	yin3 qi3	引起	to cause	198.5	2723.6	49.3	8.7
33	ying3 xiang3	影响	to affect	361.3	127.1	92.0	8.8
34	yong3 gan3	勇敢	brave	82.9	226.0	42.6	8.8
35	yun3 xu3	允许	to allow	61.9	943.4	81.5	8.7
36	zhen3 suo3	诊所	clinic	49.9	2144.4	21.8	8.3
37	zheng3 li3	整理	to tidy up	384.7	877.4	25.0	8.8
38	zhi3 dao3	指导	to instruct	470.1	155.3	21.9	8.5
39	zong3 tong3	总统	president	636.2	285.1	165.4	8.1
40	zu3 zhi3	阻止	stop	114.9	314.5	124.6	8.8
Mean				2306.2	1214.8	138.6	8.6

(*continued on next page*)

(*continued*)

Sandhi: high-frequency words							
No.	Pinyin	Sinogram	Gloss	Frequency 1st character	Frequency 2nd character	Frequency disyllabic word	Familiarity
Sandhi: low-frequency words							
No.	Pinyin	Sinogram	Gloss	Frequency 1st character	Frequency 2nd character	Frequency disyllabic word	Familiarity
1	bang3 tui3	绑腿	leggings	81.3	106.4	0.1	4.8
2	bao3 dao3	宝岛	treasure island	402.6	88.7	0.1	6.5
3	bi3 fa3	笔法	artistic technique	121.4	1296.9	0.1	5.2
4	bi3 zhe3	笔者	the author	121.4	901.9	0.1	6.0
5	cai3 mai3	采买	to buy	67.3	385.2	0.0	4.7
6	cao3 hai3	草海	Caohai Lake	81.4	350.2	0.0	2.4
7	fa3 chang3	法场	execution ground	1296.9	741.8	0.0	4.9
8	fan3 bu3	反哺	back-feeding	339.9	2.0	0.0	5.6
9	fan3 yu3	反语	irony	339.9	230.8	0.1	5.6
10	fang3 gu3	仿古	to archaize	32.0	130.5	0.0	5.0
11	fei3 shou3	匪首	gang leader	17.5	276.0	0.1	4.4
12	gao3 ben3	稿本	manuscript	18.4	876.6	0.1	5.2
13	gu3 fa3	古法	ancient style	130.5	1296.9	0.0	4.3
14	gu3 jing3	古井	ancient well	130.5	23.8	0.1	5.7
15	gun3 shui3	滚水	boiling water	154.0	465.3	0.1	3.1
16	guo3 fu3	果脯	persevered fruit	1952.8	2.0	0.1	6.0
17	hai3 cao3	海草	seaweed	350.2	81.4	0.0	4.8
18	jian3 shi3	简史	brief history	255.2	170.8	0.1	6.7
19	jiao3 zhua3	脚爪	claw	159.1	12.0	0.1	5.8
20	ku3 zhu3	苦主	the family of the victim in a murder case	161.5	611.2	0.0	3.8
21	li3 fa3	礼法	rules of etiquette	315.3	1296.9	0.0	5.7
22	pu3 xuan3	普选	universal suffrage	187.4	542.7	0.1	5.5
23	sa3 zhong3	撒种	to sow	151.5	905.1	0.1	6.5
24	shou3 li3	守礼	to follow rules of courtesy	133.1	315.3	0.0	6.0
25	shui3 zao3	水蚤	daphnia	465.3	2.7	0.0	4.3
26	si3 huan3	死缓	death sentence with reprieve	1521.2	24.6	0.1	6.2
27	tie3 bing3	铁饼	discus	79.9	88.3	0.0	5.9
28	tie3 jia3	铁甲	armor	79.9	53.1	0.0	4.9
29	tie3 ma3	铁马	iron-clad horses	79.9	744.9	0.1	4.7
30	tie3 suo3	铁索	iron chain	79.9	184.3	0.1	5.4
31	tong3 guan3	统管	to govern	285.1	621.4	0.1	4.8
32	tu3 gai3	土改	land reform	104.0	275.7	0.0	5.3
33	tui3 fa3	腿法	foot position	106.4	1296.9	0.1	4.5
34	xue3 shui3	雪水	snow water	79.7	465.3	0.1	6.6
35	ye3 guo3	野果	wild fruit	101.9	1952.8	0.1	6.7
36	ye3 shi3	野史	unofficial history	101.9	170.8	0.0	6.6
37	yong3 wu3	勇武	valiant	82.9	139.1	0.0	5.9
38	zhi3 san3	纸伞	paper umbrella	111.5	16.4	0.1	6.5
39	zhu3 bi3	主笔	chief writer	611.2	121.4	0.0	6.3
40	zong3 lan3	总揽	to assume overall responsibility	636.2	3.1	0.0	5.9
Mean				288.2	431.8	0.06	5.4
Non-sandhi: high-frequency words							
No.	Pinyin	Sinogram	Gloss	Frequency 1st character	Frequency 2nd character	Frequency disyllabic word	Familiarity
1	ben4 dan4	笨蛋	fool	85.3	408.5	69.9	8.7
2	chao3 jia4	吵架	spat	81.2	171.3	27.2	8.6
3	cheng2 ji1	成绩	grade	1309.7	24.1	24.5	8.9
4	ci4 ji1	刺激	to stimulate	102.9	218.6	39.2	8.7
5	cun2 zai4	存在	to exist	195.5	11664.3	108.1	8.7
6	dao3 zhi4	导致	to cause	155.3	132.3	45.2	8.7
7	di4 qiu2	地球	earth	1668.0	519.9	68.4	8.4
8	fa4 xing2	发型	hairstyle	1870.8	160.2	27.2	8.5
9	fan3 dui4	反对	to oppose	339.9	4094.6	73.6	8.5
10	fei1 ji1	飞机	airplane	335.5	1044.0	149.6	8.8
11	gu1 du2	孤独	lonely	75.3	196.5	39.7	8.5
12	huai2 yi2	怀疑	to suspect	186.3	197.7	86.1	8.5
13	jie2 mu4	节目	program	371.0	480.0	186.4	8.7
14	ju4 jue2	拒绝	to refuse	74.4	341.7	85.9	8.3
15	kai1 xin1	开心	happy	2707.1	1780.9	184.9	8.9
16	ke3 lian2	可怜	pitiable	4540.2	86.0	102.2	8.6
17	la1 ji1	垃圾	trash	69.9	69.8	74.1	8.8

(*continued on next page*)

(*continued*)

Sandhi: high-frequency words							
No.	Pinyin	Sinogram	Gloss	Frequency 1st character	Frequency 2nd character	Frequency disyllabic word	Familiarity
18	lü4 se4	绿色	green	49.4	461.7	26.3	8.8
19	mian4 bao 1	面包	bread	1430.7	293.4	35.8	8.8
20	mian4 shi4	面试	interview	1430.7	573.1	30.0	8.4
21	nei4 rong2	内容	content	407.1	240.5	39.7	8.7
22	qian2 bao1	钱包	wallet	752.0	293.4	29.7	8.5
23	qiang3 po4	强迫	to force	336.4	67.4	24.9	8.5
24	qing2 kuang4	情况	happening	1479.0	283.3	318.2	8.8
25	shang1 dian4	商店	shop	147.9	235.3	36.5	8.7
26	shen2 qi2	神奇	magical	471.3	383.3	39.3	8.4
27	sheng1 ri4	生日	birthday	3474.0	494.3	113.3	8.5
28	shi2 dai4	时代	era	3471.6	323.3	46.0	8.5
29	shi2 yan4	实验	experiment	1217.1	184.5	32.3	8.7
30	shi4 he2	适合	suitable for	160.7	584.5	81.6	8.8
31	tao3 lun4	讨论	to discuss	216.7	335.9	88.4	8.8
32	tian1 qi4	天气	weather	2712.4	604.9	37.6	8.8
33	wan3 can1	晚餐	dinner	1078.4	225.2	71.8	8.8
34	xian3 ran2	显然	obviously	234.4	1953.7	106.3	8.5
35	xiang3 nian4	想念	to miss	5282.8	160.8	43.8	8.3
36	yao1 qiu2	要求	to require	7594.4	486.2	155.5	8.8
37	ye4 wan3	夜晚	night	201.5	1078.4	36.3	8.7
38	you2 jian4	邮件	mail	56.9	674.9	35.6	8.5
39	zhong4 dian3	重点	focus	840.5	2733.2	35.1	8.8
40	zui4 fan4	罪犯	criminal	304.1	378.3	43.5	8.5
Mean				1188.0	866.0	72.5	8.6

Non-sandhi: low-frequency words							
No.	Pinyin	Sinogram	Gloss	Frequency 1st character	Frequency 2nd character	Frequency disyllabic word	Familiarity
1	an4 yu4	暗喻	metaphor	110.5	591.0	1.0	6.0
2	bai4 lu4	败露	to be revealed	138.0	5762.0	1.2	5.9
3	bing4 tong4	病痛	sickness	385.0	8963.0	1.7	6.9
4	cheng2 jian4	成见	bias	1309.7	85723.0	1.8	6.9
5	duan4 dai4	缎带	satin ribbon	1.4	45283.0	0.8	4.9
6	fang1 ming2	芳名	name	6.2	43229.0	0.5	5.4
7	feng1 bing4	疯病	madness	314.9	18032.0	0.6	4.6
8	guo4 sheng4	过剩	in surplus	4233.0	5308.0	1.2	7.0
9	hua1 mi4	花蜜	nectar	371.0	2405.0	0.6	6.5
10	huang1 yuan2	荒原	wilderness	48.8	21274.0	0.8	6.8
11	huang2 quan2	黄泉	Hades	78.9	917.0	1.9	6.5
12	huo3 xian4	火线	frontline	394.5	15170.0	0.8	5.9
13	huo4 xi1	获悉	to hear about	133.1	1478.0	1.5	6.7
14	ji1 wei4	机位	machine position	1044.0	31427.0	1.3	4.3
15	ji4 feng1	季风	monsoon	140.6	13301.0	0.7	5.5
16	jia4 she4	架设	to build	171.3	9056.0	0.6	3.8
17	jian1 xian3	艰险	gruelling	34.9	13181.0	0.6	7.2
18	kang4 bian4	抗辩	to argue	85.6	2765.0	1.5	5.7
19	lie4 ji4	劣迹	bad record	13.8	5466.0	0.5	6.6
20	nian3 sui4	碾碎	to crush	3.8	3520.0	1.6	6.7
21	piao2 chong2	瓢虫	ladybug	1.2	3228.0	0.9	6.4
22	qi1 xi1	栖息	to rest	3.5	19106.0	0.8	6.2
23	qie1 pian4	切片	to slice	602.1	21431.0	2.2	6.3
24	qing1 qi4	氢气	hydrogen	4.7	28334.0	0.8	6.0
25	qiu2 ai4	求爱	to woo	486.2	65256.0	1.8	7.1
26	shui4 yi3	睡椅	sleeping chair	311.7	2116.0	1.0	5.9
27	tang2 lang2	螳螂	praying mantis	1.1	281.0	1.5	7.0
28	tie3 chui2	铁锤	iron hammer	79.9	509.0	1.9	6.0
29	tong2 yin1	同音	homophonous	917.9	18229.0	2.3	6.4
30	tun1 bing4	吞并	to annex	25.1	26283.0	1.0	7.1
31	wai4 dai4	外带	takeout	677.8	45283.0	1.3	6.9
32	xi2 xing4	习性	habit	171.8	28102.0	0.6	7.1
33	xiao3 diao4	小调	lilt	2123.5	14539.0	1.3	6.5
34	xie2 chi2	胁持	to coerce	69.8	17301.0	0.5	6.3
35	xie4 jue2	谢绝	to refuse	2576.1	16004.0	0.6	6.9
36	yin3 xian4	引线	to thread	198.5	15170.0	0.6	6.3
37	you2 bao1	邮包	mail package	56.9	13742.0	0.6	7.1
38	yue4 quan2	越权	to overstep	256.9	10283.0	1.2	6.3
39	za2 hui4	杂烩	chop suey	153.7	89.0	0.9	7.1

(*continued on next page*)

(*continued*)

Sandhi: high-frequency words							
No.	Pinyin	Sinogram	Gloss	Frequency 1st character	Frequency 2nd character	Frequency disyllabic word	Familiarity
40	zhuan1 qiang2	砖墙	brick wall	9.8	3300.0	0.5	6.5
Mean				443.7	17035.9	1.1	6.3

References

Alderete, J., Chan, Q., & Yeung, H. H. (2019). Tone slips in Cantonese: Evidence for early phonological encoding. *Cognition, 191*, 103952. https://doi.org/10.1016/j.cognition.2019.04.021

Arndt-Lappe, S., & Ernestus, M. (2020). Morpho-phonological alternations: The role of lexical storage. In V. Pirrelli, I. Plag, & W. U. Dressler (Eds.), *Word knowledge and word usage* (pp. 191–227). De Gruyter Mouton. https://doi.org/10.1515/9783110440577-006.

Baayen, R. H., McQueen, J. M., Dijkstra, T., & Schreuder, R. (2003). Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In R. H. Baayen, & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 355–390). Mouton de Gruyter. https://doi.org/10.1515/9783110910186.355.

Becker, M., Ketrez, N., & Nevins, A. (2011). The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language, 87*(1), 84–125.

Boersma, P., & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry, 32*(1), 45–86. https://doi.org/10.1162/002438901554586

Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer* (6.1.50) [Computer software] http://www.praat.org/.

Brysbaert, M., Mandera, P., & Keuleers, E. (2017). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science, 27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Bybee, J. L. (2001). *Phonology and language use*. Cambridge University Press.

Chang, C. H. C., & Kuo, W.-J. (2016). The neural substrates underlying the implementation of phonological rule in lexical tone production: An fMRI study of the tone 3 sandhi phenomenon in Mandarin Chinese. *PLOS ONE, 11*(7), Article e0159835. https://doi.org/10.1371/journal.pone.0159835

Chang, H.-C., Lee, H.-J., Tzeng, O. J. L., & Kuo, W.-J. (2014). Implicit target substitution and sequencing for lexical tone production in Chinese: An fMRI study. *PLOS ONE, 9*(1), Article e83126. https://doi.org/10.1371/journal.pone.0083126

Chao, Y.-R. (1930). A system of tone letters. *Le Maître Phonétique, 45*, 24–27.

Chen, J.-Y. (1999). The representation and processing of tone in Mandarin Chinese: Evidence from slips of the tongue. *Applied PsychoLinguistics, 20*, 289–301.

Chen, S., He, Y., Wayland, R., Yang, Y., Li, B., & Yuen, C. W. (2019). Mechanisms of tone sandhi rule application by tonal and non-tonal non-native speakers. *Speech Communication, 115*, 67–77. https://doi.org/10.1016/j.specom.2019.10.008

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

Chien, Y.-F., Sereno, J. A., & Zhang, J. (2017). What's in a word: Observing the contribution of underlying and surface representations. *Language and Speech, 60*(4), 643–657. 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*. https://doi.org/10.3389/fpsyg.2020.00646

Coetzee, A. W., & Kawahara, S. (2013). Frequency biases in phonological variation. *Natural Language & Linguistic Theory, 31*(1), 47–89. https://doi.org/10.1007/s11049-012-9179-z

Coetzee, A. W., & Pater, J. (2011). The place of variation in phonological theory. In J. A. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *The handbook of phonological theory* (2nd ed., pp. 401–434). Wiley-Blackwell. https://doi.org/10.1002/9781444343069.ch13.

Cole, J., & Hualde, J. I. (2011). Underlying representations. In *The blackwell companion to phonology* (pp. 1–26). https://doi.org/10.1002/9781444335262.wbctp0001

Connine, C. M., & Pinnow, E. (2006). Phonological variation in spoken word recognition: Episodes and abstractions. *The Linguistic Review, 23*(3), 235–245. https://doi.org/10.1515/TLR.2006.009

Crepaldi, D., Hemsworth, L., Davis, C. J., & Rastle, K. (2016). Masked suffix priming and morpheme positional constraints. *Quarterly Journal of Experimental Psychology, 69*(1), 113–128. https://doi.org/10.1080/17470218.2015.1027713

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology, 115*(4), 732–744. https://doi.org/10.1016/j.clinph.2003.11.021

Domahs, U., Kehrein, W., Knaus, J., Wiese, R., & Schlesewsky, M. (2009). Event-related potentials reflecting the processing of phonological constraint violations. *Language and Speech, 52*(4), 415–435. https://doi.org/10.1177/0023830909336581

Gao, X., Yan, T.-T., Tang, D.-L., Hang, T., Shua, H., Nan, Y., & Zhang, Y.-X. (2019). What makes lexical tone special: A reverse accessing model for tonal speech perception. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fnbeh.2014.00097

Hagoort, P., & Levelt, W. J. M. (2009). The speaking brain. *Science, 326*(5951), 372–373. https://doi.org/10.1126/science.1181675

Hayes, B., & Londe, Z. C. (2006). Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology, 23*(1), 59–104. https://doi.org/10.1017/S0952675706000765

Hsieh, H.-I. (1970). The psychological reality of tone sandhi rules in Taiwanese. In *Papers from the 6th meeting of the chicago linguistic society* (pp. 489–503). Chicago Linguistic Society.

Hsieh, H.-I. (1975). How generative is phonology? In E. F. K. Koerner (Ed.), *The transformational-generative paradigm and modern linguistic theory* (pp. 109–144). John Benjamins.

Hsieh, H.-I. (1976). On the unreality of some phonological rules. *Lingua, 38*(1), 1–19. https://doi.org/10.1016/0024-3841(76)90038-3

Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology, 2*. https://doi.org/10.3389/fpsyg.2011.00255

Kenstowicz, M., & Kisseberth, C. (1979). *Generative phonology: Description and theory*. Academic Press.

Kiparsky, P. (1973). *Abstractness, opacity and global rules*. Indiana University Linguistics Club.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463–470. https://doi.org/10.1016/S1364-6613(00)01560-6

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Lahiri, A., & Reetz, H. (2008). Underspecified recognition. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology, 7* pp. 637–676). https://doi.org/10.1515/9783110197105

Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics, 38*(1), 44–59. https://doi.org/10.1016/j.wocn.2010.01.002

Landi, N., Crowley, M. J., Wu, J., Bailey, C. A., & Mayes, L. C. (2012). Deviant ERP response to spoken non-words among adolescents exposed to cocaine in utero. *Brain and Language, 120*(3), 209–216. https://doi.org/10.1016/j.bandl.2011.09.002

Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences, 3*(6), 223–232. https://doi.org/10.1016/S1364-6613(99)01319-4
Li, X., & Chen, Y. (2015). Representation and processing of lexical tone and tonal variants: Evidence from the mismatch negativity. *PLOS ONE, 10*(12), Article e0143097. https://doi.org/10.1371/journal.pone.0143097
Liu, Y., & Perfetti, C. A. (2003). The time course of brain activity in reading English and Chinese: An ERP study of Chinese bilinguals. *Human Brain Mapping, 18*(3), 167–175. https://doi.org/10.1002/hbm.10090
Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024
Marslen-Wilson, W., & Tyler, L. K. (1998). Rules, representations, and the English past tense. *Trends in Cognitive Sciences, 2*(11), 428–435. https://doi.org/10.1016/S1364-6613(98)01239-X
McRae, K., Jared, D., & Seidenberg, M. S. (1990). On the roles of frequency and lexical access in word naming. *Journal of Memory and Language, 29*(1), 43–65. https://doi.org/10.1016/0749-596X(90)90009-O
Meng, Y., Wynne, H., & Lahiri, A. (2021). Representation of "T3 sandhi" in Mandarin: Significance of context. *Language, Cognition and Neuroscience, 36*(6), 791–808. https://doi.org/10.1080/23273798.2021.1893769
Meunier, F., & Segui, J. (1999). Morphological priming effect: The role of surface frequency. *Brain and Language, 68*(1–2), 54–60. https://doi.org/10.1006/brln.1999.2098
Morris, J., & Stockall, L. (2012). Early, equivalent ERP masked priming effects for regular and irregular morphology. *Brain and Language, 123*(2), 81–93. https://doi.org/10.1016/j.bandl.2012.07.001
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
Pan, J., Zhang, C., Huang, X., & Yan, M. (2020). Sandhi-tone words prolong fixation duration during silent sentence reading in Chinese. *Reading and Writing*. https://doi.org/10.1007/s11145-020-10093-7
Peng, S.-H. (2000). Lexical versus phonological representations of Mandarin sandhi tones. In M. Bro, & J. Pierrehumbert (Eds.), *Language acquisition and the lexicon: Papers in laboratory phonology V* (pp. 152–167). Cambridge University Press.
Pinker, S., & Prince, A. (1994). Regular and irregular morphology and the psychological status of rules of grammar. In S. D. Lima, R. Corrigan, & G. Iverson (Eds.), *The reality of linguistic rules* (pp. 321–351). John Benjamins.
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
Prasada, S., & Pinker, S. (1993). Generalisation of regular and irregular morphological patterns. *Language & Cognitive Processes, 8*(1), 1–56. https://doi.org/10.1080/01690969308406948
Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language, 57*(2), 273–298. https://doi.org/10.1016/j.jml.2007.04.001
Roelofs, A. (2015). Modeling of phonological encoding in spoken word production: From Germanic languages to Mandarin Chinese and Japanese. *Japanese Psychological Research, 57*, 22–37. https://doi.org/10.1111/jpr.12050
Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Memory & Cognition, 18*(4), 367–379. https://doi.org/10.3758/BF03197126
Sahin, N., Pinker, S., Cash, S., Schomer, D., & Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within broca's area. *Science, 326*(5951), 445–449. https://doi.org/10.1126/science.1174481
Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology, 56*(6), Article e13335. https://doi.org/10.1111/psyp.13335
Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29*(1), 86–102. https://doi.org/10.1016/0749-596X(90)90011-N
Shuai, L., & Gong, T. (2014). Temporal relation between top-down and bottom-up processing in lexical tone perception. *Frontiers in Behavioral Neuroscience, 8*(97). https://doi.org/10.3389/fnbeh.2014.00097
Stemberger, J., & MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory & Cognition, 14*(1), 17–26. https://doi.org/10.3758/BF03209225
Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final/t. *Journal of Memory and Language, 52*(3), 322–338. https://doi.org/10.1016/j.jml.2004.11.004
Tsay, J., & Myers, J. (1996). Taiwanese tone sandhi as allomorph selection. In *Proceedings of the 22nd annual meeting of the berkeley linguistics society: General session and parasession on the role of learnability in grammatical theory* (pp. 395–405). Berkeley, CA.
Ulbrich, C., Alday, P. M., Knaus, J., Orzechowska, P., & Wiese, R. (2016). The role of phonotactic principles in language processing. *Language, Cognition and Neuroscience, 31*(5), 662–682. https://doi.org/10.1080/23273798.2015.1136427
Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J., & Pinker, S. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience, 9*(2), 266–276. https://doi.org/10.1162/jocn.1997.9.2.266
Van Lancker Sidtis, D. (2012). Formulaic language and language disorders. *Annual Review of Applied Linguistics, 32*, 62–80. https://doi.org/10.1017/S0267190512000104
Van Lancker Sidtis, D., & Postman, W. A. (2006). Formulaic expressions in spontaneous speech of left- and right-hemisphere-damaged subjects. *Aphasiology, 20*(5), 411–426. https://doi.org/10.1080/02687030500538148
Wang, X. (1993). The psychological reality of tone sandhi in Taiwanese. *Tsing Hua Journal of Chinese Studies, 23*, 175–192.
Wang, Y., Jiang, M., Huang, Y., & Qiu, Y. (2021). An ERP study on the role of phonological processing in reading two-character compound Chinese words of high and low frequency. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.637238
Wang, W., Lu, A., He, D., Zhang, B., & Zhang, J. X. (2017). ERP evidence for Chinese compound word recognition: Does morpheme work all the time? *NeuroQuantology, 15*(3). https://doi.org/10.14704/nq.2017.15.3.1105
Wan, I.-P., & Jaeger, J. J. (1998). Speech errors and the representation of tone in Mandarin Chinese. *Phonology, 15*, 417–461.
White, J., & Chiu, F. (2017). Disentangling phonological well-formedness and attestedness: An ERP study of onset clusters in English. *Acta Linguistica Academica, 64*(4), 513–537. https://doi.org/10.1556/2062.2017.64.4.2
Wiese, R., Orzechowska, P., Alday, P. M., & Ulbrich, C. (2017). Structural principles or frequency of use? An ERP experiment on the learnability of consonant clusters. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.02005
Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics, 25*(1), 61–83. https://doi.org/10.1006/jpho.1996.0034
Yuan, J., & Chen, Y. (2014). 3rd tone sandhi in standard Chinese: A corpus approach. *Journal of Chinese Linguistics, 42*(1), 218–236.
Zeng, Y., Fiorentino, R., & Zhang, J. (2021). Electrophysiological signatures of perceiving alternated tone in Mandarin Chinese: Mismatch negativity to underlying tone conflict. *Frontiers in Psychology, 12*, 735593. https://doi.org/10.3389/fpsyg.2021.735593
Zhang, J. (in press). Tonal processes defined as tone sandhi. In C.-R. Huang, Y.-H. Lin, & I.-H. Chen (Eds.), The Cambridge handbook of Chinese linguistics. Cambridge University Press.
Zhang, C., Huang, X., Politzer-Ahles, S., Zhang, J., & Peng, G. (2017). Neural encoding of T3 sandhi in Mandarin Chinese speakers in speech production. In *The 9th annual meeting of the society for the Neurobiology of Language*. SNL 2017.
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

Zhang, J., Lai, Y., & Sailor, C. (2011). Modeling Taiwanese speakers' knowledge of tone sandhi in reduplication. *Lingua, 121*(2), 181–206. https://doi.org/10.1016/j.lingua.2010.06.010

Zhang, C., & Peng, G. (2013). Productivity of 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* (pp. 256–282). City University of Hong Kong.

Zhang, C., Xia, Q., & Peng, G. (2015). Mandarin third tone sandhi requires more effortful phonological encoding in speech production: Evidence from an ERP study. *Journal of Neurolinguistics, 33*, 149–162. https://doi.org/10.1016/j.jneuroling.2014.07.002

Zuraw, K. (2010). A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language & Linguistic Theory, 28*(2), 417–472. https://doi.org/10.1007/s11049-010-9095-z