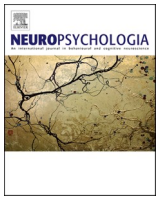
Neuropsychologia 175 (2022) 108367

Contents lists available at ScienceDirect

Neuropsychologia

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



Encoding category-level and context-specific phonological information at different stages: An EEG study of Mandarin third-tone sandhi word production

Xiaocong Chen [a,*], Caicai Zhang [a,**], Yiya Chen [b], Stephen Politzer-Ahles [a,c], Yuyu Zeng [c,d], Jie Zhang [c]

[a] Research Centre for Language, Cognition, and Neuroscience, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China
[b] Leiden University Center for Linguistics and Leiden Institute for Brain and Cognition, Netherlands
[c] Department of Linguistics, College of Liberal Arts and Sciences, University of Kansas, USA
[d] Speech Motor Neuroscience Group, Waisman Center, University of Wisconsin, Madison, USA

ARTICLE INFO

Keywords:
Mandarin T3 sandhi
Word production
Underlying form
Surface form
EEG

ABSTRACT

Pronunciation of words or morphemes may vary systematically in different phonological contexts, but it remains unclear how different levels of phonological information are encoded in speech production. In this study, we investigated the online planning process of Mandarin Tone 3 (T3) sandhi, a case of phonological alternation whereby a low-dipping tone (T3) changes to a Tone 2 (T2)-like rising tone when followed by another T3. To examine the time course of the encoding of the abstract category-level (underlying form) and context-specific phonological form (surface form) of T3, we conducted an electroencephalographic (EEG) study with a phonologically-primed picture naming task and examined the event-related potentials (ERPs) time-locked to the stimulus onset as well as speech response onset. The behavioral results showed that targets primed by T3 or T2 primes yielded shorter naming latencies than those primed by control primes. Importantly, the EEG data revealed that T3 primes elicited larger positive amplitude over broad frontocentral regions roughly in the 320–550 ms time window of stimulus-locked ERP and −500 to −400 ms time window of response-locked ERP, whereas T2 primes elicited larger negative amplitude over left frontocentral regions roughly in the −240 to −100 ms time window of response-locked ERP. These results indicate that the underlying and the surface form are encoded at different processing stages. The former presumably occurs in the earlier phonological encoding stage, while the latter probably occurs in the later phonetic encoding or motor preparation stage. The current study offers important implications for understanding the processing of phonological alternations and tonal encoding in Chinese word production.

1. Introduction

In connected speech, words or morphemes are not always pronounced in canonical forms but may change to different pronunciation variants in different phonological environments. For example, the English plural suffix "-s" is pronounced as [s] when preceded by voiceless consonants (e.g., *books*), and as [z] when preceded by voiced consonants or vowels (e.g., *girls*, *boys*). Such phonological alternation occurs not only in segments but also in suprasegments. A well-noted example is the tone sandhi in many Chinese dialects, whereby a lexical tone changes from its citation pitch form (pronunciation in isolation) to a different pitch form in the conditioned phonological context (Chen, 2000). Despite the abundance of phonological alternations in natural speech, how speakers plan the phonological alternations and select the intended variant conditioned by the phonological context in speech production has received relatively less attention in previous research. Related to this

* Corresponding author. HJ613, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, 11 Yuk Choi Rd, Hung Hom, SAR, Hong Kong, China.
** Corresponding author. Room EF741, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, 11 Yuk Choi Rd, Hung Hom, SAR, Hong Kong, China.
E-mail addresses: felcshallot@gmail.com, xiaocong.chen@polyu.edu.hk (X. Chen), caicai.zhang@polyu.edu.hk (C. Zhang).

https://doi.org/10.1016/j.neuropsychologia.2022.108367
Received 3 April 2022; Received in revised form 2 July 2022; Accepted 1 September 2022
Available online 6 September 2022

question, current speech production models do not adequately account for the encoding process of phonological alternations (Bürki, 2018; Bürki et al., 2014). To shed light on this issue, the current study seeks to investigate the online planning process of the Mandarin Tone 3 sandhi, a case of phonological alternation, during speech production.

In Mandarin, there are four lexical tones associated with a syllable to distinguish meanings: Tone 1 (T1, a high-level tone/55/,[1] e.g., *mā*,[2] 妈, 'mother'), Tone 2 (T2, a rising tone/35/, e.g., *má*, 麻, 'hemp'), Tone 3 (T3, a low-dipping tone/214/, e.g., *mǎ*, 马, 'horse') and Tone 4 (T4, a high-falling tone/51/, e.g., *mà*, 骂, 'scold'). In disyllabic words, when the T3 is followed by another T3 (e.g., *yǔ-sǎn* 雨伞, 'umbrella'), the first T3 changes from its original citation pitch form (a low-dipping tone/214/) to a T2-like high rising tone (/35/).[3] This process is known as T3 sandhi (e.g., Chao, 1948; Peng, 2000; Wang and Li, 1967; Yuan and Chen, 2014; C. Zhang and Peng, 2013; J. Zhang and Lai, 2010). It has been argued that this process cannot be attributed to a pure articulatory process like tonal co-articulation (Shih, 1986; Xu, 1997; J. Zhang and Lai, 2010). There has been an intense debate over the encoding process of Mandarin T3 sandhi in previous literature (e.g., Y. Chen et al., 2011; Nixon et al., 2015; C. Zhang et al., 2015; J. Zhang et al., 2022). Studies on phonological alternations propose that phonological alternations could be generated during speech production in at least two ways: either via the online computation (e.g., through phonological rules) operated on a presumed underlying representation for phonological variants, or via the direct retrieval of the pre-stored phonological variants from the lexicon (Arndt-lappe and Ernestus, 2020; Bürki et al., 2010, 2011; Bürki and Gaskell, 2012; Bürki et al., 2014). Regarding Mandarin T3 sandhi, different accounts have also been proposed by linguists.

Influenced by the early generative phonological framework (Chomsky and Halle, 1968), the computation account implicitly assumes that all the T3 variants are only stored as a shared underlying representation (T3 tonal category) in the lexicon and proposes that the context-specific surface form (T2-like rising sandhi variant) is computed on the fly in production (e.g., C. Zhang and Peng, 2013; C. Zhang et al., 2015; J. Zhang and Lai, 2010; J. Zhang et al., 2022; hereafter, we will label T3 as the underlying form and T2 as the surface form for convenience). Although researchers disputed over the nature of the underlying representation (see discussion in Chien et al., 2016, 2020, 2021; Meng et al., 2020; Politzer-Ahles et al., 2016; Zhou and Marslen-Wilson, 1997) and computation (e.g., Politzer-Ahles and Zhang, in press), it has been suggested that the generation of T2-like sandhi variant could be operated as an online tonal substitution rule that changes the underlying T3 to a T2 (Chang and Kuo, 2020). Under this account, the T3 sandhi production involves a serial encoding process, with the underlying T3 firstly specified and the surface T2-like sandhi tone generated at a later time point before phonetic spellout or articulation (e.g., J. Chen, 1999; Politzer--Ahles and Zhang, in press). Moreover, as the sandhi tone involves online computation and does not require direct retrieval of the surface form from the lexicon, it also assumes that the application of T3 sandhi will not be influenced by the lexical frequency and will be productively observed in pseudowords and nonwords (see below).

In contrast, the lexical storage account assumes that the disyllabic T3 sandhi words are only stored as their surface form (i.e., as a T2 + T3 sequence) in the mental lexicon (the surface representation view defined by Zhou and Marslen-Wilson, 1997, also see J. Chen et al., 2012), and the morpheme-specific sandhi variant is directly accessed from memory during production (e.g., Hsieh, 1970, 1976). Under this account, the underlying form of T3 plays no role in production. Moreover, this account also suggests that T3 sandhi will be only applied to real words and the application of T3 sandhi will be sensitive to word frequency, as the word frequency would affect the strength of the representation of the surface sandhi tone in memory and thus the subsequent retrieval process in production (see discussion in J. Zhang et al., 2022).

More recently, another alternative account, the multivariant activation account, has also been proposed (Y. Chen et al., 2011; Li and Chen, 2015). This account takes a hybrid view and assumes the storage of both the abstract T3 category and the T2-like T3 sandhi variant (which is abstracted from the T3 exemplars and stored as allotones under the general T3 category). This assumption stands in contrast to the computation account (which mainly assumes the storage of only the abstract T3 category) and the lexical storage account (which assumes the storage of only surface sandhi tone). Moreover, the multivariant activation account also assumes that the T3 tonal category and the context-specific T2-like sandhi variant are both activated during T3 sandhi production (Nixon et al., 2015; Y. Chen et al., 2011). Although this account does not make specific assumptions on the temporal order of the activation of different variants during production, results of a PWI study implied that both the T3 tonal category and the context-specific T2-like sandhi variant are activated in parallel (Nixon et al., 2015), which is different from the serial processing assumption of the computation account. Under this assumption, the intended sandhi variant is only selected later after the following tonal information is available (see a similar proposal in Miozzo and Caramazza, 1999). In addition, the multivariant account makes no specific assumptions on the role of lexical frequency in the application of T3 sandhi.

The computation account has gained support from behavioral studies testing the productivity of Mandarin T3 sandhi. An early study (Cheng, 1968) found that T3 sandhi is not only applied in Chinese real words but also in code-mixing phrases when a T3 Chinese morpheme is followed by English words beginning with an unstressed low-pitch syllable (e.g. *好 professor*, 'good professor'). Moreover, previous literature has consistently shown that T3 sandhi can be productively applied to pseudowords composed of existing morphemes (e.g., *损岭, sǔn lǐng, 'to destroy' + 'mountain'*) (Deng et al., 2003; Politzer-Ahles and Zhang, in press; Tang et al., 2019; Xu, 1997; J. Zhang and Lai, 2010), nonce words containing pseudo-syllables (where the segmental syllable exists but never occurs together with T3 in a real situation, e.g., *pǐng zěng*) (J. Zhang and Lai, 2010), and even sequences containing two non-occurring segmental syllables (e.g., *tiǎng luǎ)* (C. Zhang and Peng, 2013; C. Zhang et al., 2015).

Several recent neurolinguistic studies also provide evidence for the computation account. In an EEG study by C. Zhang et al. (2015), participants were asked to listen to two real or two non-occurring syllables (consisting of either a T2 plus T3 or two T3s) and make covert production by combining the two syllables into a disyllabic sequence. Examination of the event-related potentials (ERP) time-locked to the second syllable revealed a larger positive amplitude in the P2 time window for the T3 + T3 sequence than for the T2 + T3 sequence, no matter whether the syllables were real or non-occurring. As larger P2 is linked to processing difficulty (Kim et al., 2008) and P2 is sensitive to phonological processing (Crowley and Colrain, 2004; Landi et al., 2012), the researchers interpreted this result as a likely indication of additional effortful phonological processing caused by the T3 sandhi irrespective of the lexicality status, thus supporting the computation account. Another recent EEG study (J. Zhang et al., 2022) also found that T3 sandhi words elicited a larger positive amplitude than non-sandhi words in the left anterior regions around 336–520 ms after participants were cued to produce the target words, and this effect was not modulated by lexical frequency, which is aligned with the computation account. Other studies

[1] The tonal values are based on Chao's (1948) five-point-scale tonal notation system, with 5 standing for the highest pitch value and 1 for the lowest pitch value.

[2] The italicized forms are Hanyu Pinyin, the romanization system used in China, with tonal values marked on the vowels by diacritics.

[3] Previous studies reveal that there are some minute acoustic differences between the derived T3 sandhi rising tone and the T2 rising tone (e.g., Kuo et al., 2007 and Peng, 2000 based on laboratory speech elicited in designed experiments; Yuan and Chen, 2014 based on corpus data), but it is also shown that native speakers could not distinguish them in perception, as observed in identification tasks (e.g., Peng, 2000; Wang and Li, 1967).

(Chang and Kuo, 2016; Chang et al., 2014) using functional magnetic resonance imaging (fMRI) to examine the brain activation during T3 sandhi production also revealed a higher activation of the right posterior inferior frontal gyrus (IFG) in the condition where participants were required to produce sequences involving T3 sandhi overtly. Chang and Kuo (2020) argued that the higher activation of the right IFG in producing T3 sandhi sequences supports the existence of an extra online tone substitution operation during the T3 sandhi production. According to Chang and Kuo (2020), the frontal lobe has been suggested to engage in online phonological processing and articulation in previous research (e.g., Indefrey and Levelt, 2004), so the higher activation of the right IFG may indicate more efforts due to the extra online tonal operation in pronouncing the T3 sandhi sequences, which provided support for the computation account.

However, the tasks used in previous productivity and neurolinguistic studies typically require participants to combine isolated syllables into sequences, which may encourage participants to focus more on the underlying form and resort more to the computation mechanism (see discussion in Chien et al., 2020; J. Zhang et al., 2022). For this reason, it remains less clear whether the computation mechanism is also engaged in those more naturalistic language production tasks (e.g., word naming or picture naming).

The production of T3 sandhi words has also been investigated using other tasks. One line of the research comes from the implicit priming task, where shorter speech latencies are found when participants produced a set of disyllabic words sharing the initial tonal syllable (homogenous set) than a set of words without tonal overlap in the initial syllable or a set of words without any initial phonological overlap (heterogeneous set) (J. Chen et al., 2002). However, this line of research has yielded inconsistent findings on the T3 sandhi. For example, Y. Chen et al. (2011) and J. Chen et al. (2012) observed similar degrees of facilitatory effects in production latencies when the underlying tonal category of T3 was shared, no matter whether the surface tone was shared or not. In contrast, Politzer-Ahles and Zhang (in press) and X. Chen (2012) found that facilitation in production latencies occurred only when the surface tone was shared, but no facilitation was observed when the underlying T3 category was shared and the surface tone differed. The inconsistencies of the above studies could be attributed to different task implementation, attentional bias and ad-hoc strategies caused by the learning procedure of the implicit priming task (Politzer-Ahles and Zhang, in press) or the heterogenous language background of participants (e.g., Beijing vs. Taiwan Mandarin) across these studies. More crucially, some researchers have suggested that the implicit priming task may not be sensitive to the early phonological encoding stage (Wong and Chen, 2008, p. 1173). Thus, this task may not allow researchers to probe into the entire production process of T3 sandhi words.

To investigate the Mandarin T3 sandhi word production, Nixon et al. (2015) employed another task, the picture-word interference (PWI) paradigm, whereby participants were asked to name pictures while ignoring accompanying distractors. They further manipulated the stimulus-onset asynchrony (SOA hereafter) to tap into the time course of the activation for the category-level (underlying form) and context-specific (surface form) information in production. In their first experiment, participants saw 27 pictures of disyllabic T3 sandhi target words (e.g., 辅导, *fǔ dǎo*, 'tutor'), with a distractor Chinese character superimposed upon the pictures. The distractor character consisted of three types: a T3 morpheme (sharing the tonal category, or the underlying form, e.g., 斧, *fǔ*), a T2 morpheme (sharing the context-specific pitch contour, or the surface form, e.g. 服, *fú*), and a control morpheme (T1 or T4, e.g., 付, *fù*). Moreover, the SOA was manipulated by presenting the distractors simultaneously with the picture onset (SOA = 0) or after the picture onset (SOA = 83 ms). In general, they found that the T3 and T2 distractors both facilitated the naming latencies of T3 sandhi words compared to the control distractor, suggesting the activation of both underlying and surface forms during the production of T3 sandhi words. Moreover, further analyses revealed that both the T3 and T2 distractors led to a significant facilitation effect when the SOA was 0 ms. However, only the T2 distractor significantly facilitated the naming latencies when the SOA was 83 ms. As argued by the authors, the facilitation effect of both T3 and T2 distractors at 0 ms SOA suggested that both the category-level and context-specific phonological information of T3 were activated in an early time window. In contrast, the remaining facilitation effect only for the T2 distractor at 83 ms SOA could be because the T2-like surface form has longer activation than the T3 underlying form, or because, with both forms activated all the time, the T2 distractor could additionally benefit the later articulation preparation due to the articulatory/acoustic congruency between the T2 distractor and the final surface form.

In general, the results in Nixon et al. (2015) support the multivariant activation account assuming activation of both underlying and surface forms in production. However, the results did not conclusively confirm the parallel activation of the underlying and surface forms. Instead, the serial processing assumed by the computation account might offer an alternative explanation. It may be assumed that during T3 sandhi production, the underlying form of T3 is first activated, and the surface form is later computed on the fly. If so, the T3 distractor at 0 ms SOA could primarily facilitate the earlier phonological encoding process while the T2 distractor at 0 ms SOA only facilitates the later articulation preparation process. Thus, we could expect both T3 and T2 distractors at 0 ms SOA could yield facilitation effects on the final naming latencies. But at the late SOA (83 ms), only the later production process could be mostly influenced. As the processing of distractors takes time, the facilitation effect is thus only preserved for the T2 distractor but not for the T3 distractor. Therefore, behavioral studies may not be the most informative, because naming latencies as the end-product of production could not reveal the entire time course of production, which calls for a further investigation using EEG.

1.1. The present study

Given the unresolved controversy over the planning process of T3 sandhi word production in the previous literature, we conducted an EEG study using a phonologically-primed picture naming task to examine the time course of the encoding of the categorical-level (underlying form) and context-specific (surface form) phonological information in Mandarin T3 sandhi word production. Instead of merely relying on the final naming latencies in behavioral measurements, we utilized the high temporal resolution of EEG to directly probe into the online planning process during T3 sandhi word production to obtain a clearer picture of its time course.

Moreover, following previous EEG studies on speech production (e.g., Laganaro and Perret, 2011; Laganaro, 2014), we examined the event-related potentials (ERPs) time-locked to both the stimulus (picture) onset and response (speech production) onset. As suggested in the introduction, the encoding of the context-specific surface form could potentially occur in the later motor planning stage. As pointed out by Bürki (2017), the stimulus-locked ERP analysis could probe into the early phonological encoding processes shortly after the presence of the stimulus, but only the response-locked ERP analysis could tap into the late motor preparation processes close to articulation. However, nearly all the current EEG studies on Chinese word production only focused on the ERPs time-locked to the stimulus onset (e.g., Cai et al., 2020; Feng et al., 2019; Qu et al., 2020), which may fail to capture the later motor planning processes (Laganaro, 2014). Thus, a combination of stimulus-locked and response-locked ERP allowed us to investigate the entire production process, which could offer new insights into the time course issue.

Similar to Nixon et al. (2015), we employed a picture naming task in this study, which could cover the entire production process from conceptualization to articulation (Indefrey and Levelt, 2004) and is more naturalistic than the implicit priming task or the syllable or

morpheme combination tasks. Moreover, as Nixon et al. (2015), we used overt immediate naming instead of covert or delayed naming (e.g., C. Zhang et al., 2015; J. Zhang et al., 2022), because previous research has suggested that covert or delayed production may involve different processes from overt production (e.g., Ganushchak et al., 2011; Laganaro and Perret, 2011) and previous fMRI research found that the effect of the T3 sandhi could only be observed in the overt production rather than covert production (e.g., Chang & Kuo, 2020).

However, we modified the traditional PWI paradigm used in Nixon et al. (2015) to optimize ERP data analysis. If visual distractors are superimposed on pictures during naming like in Nixon et al. (2015), it would make it difficult to tease apart the processes of the distractor recognition and word production on the ERPs. To circumvent this problem, we introduced two changes to Nixon et al.'s (2015) PWI task. First, we adopted the priming/cueing technique in previous spoken word recognition studies (e.g., Chien et al., 2016, 2021) and employed a phonologically-primed picture naming task,[4] whereby participants were asked to name pictures of disyllabic T3 sandhi words, with different types of monosyllabic primes presented with a long negative SOA to separate the prime processing and word production. Second, unlike the typical PWI paradigm that only employed either visual or auditory distractors, we combined the use of both auditory and visual primes (which were simultaneously presented for the same short duration before the target pictures) in our current study to boost the priming effect with a long negative SOA and to minimize potential visual or auditory confusion (see more details in Methods).

To examine the encoding of the underlying and surface forms in T3 sandhi word production, we manipulated the tonal features of the preceding monosyllabic primes, with the conditions of the primes similar to Nixon et al. (2015). There were three types of primes: a T3 prime (sharing only the underlying form with the first morpheme of the target), a T2 prime (sharing only the surface form with the first morpheme of the target), and a control prime (carrying T1 or T4, with syllabic overlap but no tonal overlap with the first morpheme of the target).

Based on previous theoretical accounts, we make the following predictions about the priming effect of the T3 and T2 prime (compared to the control prime) on both behavioral and ERP data. The computation account assumes a serial encoding process from the early activation of the underlying form to the later generation of the surface form before articulation. According to previous proposals (J. Chen, 1999; Politzer--Ahles and Zhang, in press), we also speculate that the activation of the underlying form may occur in the early phonological encoding stage while the generation of the surface form may occur in the later phonetic encoding or motor preparation stage. Thus, the computation account predicts facilitatory effects on the naming latencies for both T3 and T2 prime with their effects occurring in early and late time windows respectively: the T3 prime will modulate the ERP amplitude in an earlier time window (phonological encoding stage), whereas the T2 prime will modulate the ERP amplitude in a later time window (phonetic encoding/motor preparation stage), especially in the response-locked ERPs. In contrast, the lexical storage account assumes that only the surface form is represented in the T3 sandhi words, and the surface form will be retrieved earlier in the phonological encoding stage and continue to be activated until the phonetic encoding/motor preparation stage. Thus, only the T2 prime will lead to a faciliatory effect on naming latencies and modulate the ERP amplitude from the early to late time windows (from phonological encoding to phonetic encoding/motor preparation stage). Alternatively, the multivariant activation account assumes that the underlying and surface form are both stored and activated in production and thus also predicts the priming effects for both the T3 and T2 prime. As mentioned above, the multivariant activation account did not make specific assumptions on the time course of the activation of both the underlying and surface form given that only behavioral results were collected. However, the discussion in Nixon et al. (2015) implicitly indicates that the underlying and surface form are activated in parallel from the early stage of production. If so, we would expect the T3 and T2 primes to simultaneously modulate the ERP amplitude in the early time window (the phonological encoding stage), but only the effect of the T2 prime continues to the late time window (the phonetic encoding/motor preparation stage).

Moreover, we also explored the potential modulation effect of lexical frequency on the activation of the underlying and surface form during T3 sandhi production. Although some previous studies did not find a modulatory effect of lexical frequency on the T3 sandhi production (e.g., C. Zhang et al., 2015; J. Zhang et al., 2022; Nixon et al., 2015), another study did show that when the lexical frequency is extremely high (i.e., >1000 in frequency counts of 3,431,707 words in the Xinhua newswire corpus), there is an influence on the acoustic realization of T3 sandhi words (Yuan and Chen, 2014). Thus, this issue still awaits further investigation. As mentioned earlier, only the computation account and the lexical storage account make specific assumptions on the effect of lexical frequency. The computation account assumes no role of lexical frequency in T3 sandhi production. In contrast, the lexical storage account assumes that the T3 sandhi production process may be modulated by lexical frequency. As very high-frequency words may have a strong listing of the surface form in memory, we expect that the effect of the T2 prime may be stronger for very high-frequency words than for low-frequency words (also see similar discussion in Yuan and Chen, 2014).

2. Methods

2.1. Participants

We recruited 39 native Standard Mandarin speakers, who were studying or working in Hong Kong at the time of the experiment. All the participants were born and raised in Mandarin-speaking regions in Northern China (regions to the north of the Qinling-Huaihe Line) and reported they acquired Standard Mandarin (Putonghua) before the age of seven and predominantly used Standard Mandarin in their daily life.[5] To minimize the potential influence from the usage of those Southern Chinese dialects, we excluded one participant who reported to have acquired Cantonese at an early age and currently use Cantonese as the dominant language (over 90%) in daily communication,[6] and another participant who acquired the Hakka dialect from her parents in childhood. Besides, another two participants had to be excluded due to failure to complete the whole experiment. Thus, the final sample included 35 participants (12 men and 23 women; mean age = 22.1 years, SD = 3.9, range = 18–34). All the participants were right-handed (as assessed by the Edinburgh Handedness Inventory, Oldfield, 1971), and reported normal hearing, normal or corrected-to-normal vision, and no neurological or language disorder history. The experimental procedures were

[4] The phonologically-primed task used in the current study is akin to the PWI task when the auditory distractors are used and presented in a negative SOA. However, we did not use "PWI" to refer to the current task because in the current task, the processing of distractors/primes did not overlap with the picture naming process, unlike most traditional PWI tasks. This is better described as "priming" instead of "interference".

[5] There were 12 participants reporting that they could speak a local Mandarin dialectal variety (all are Northern Mandarin varieties, including Northeastern Mandarin, Tianjin Mandarin, Central Plains Mandarin, Lanyin Mandarin, Jiaoliao Mandarin, and Jilu Mandarin). The other 23 participants reported that they did not speak any other local Mandarin variety except standard Mandarin/Beijing Mandarin.

[6] Another 7 participants reported that they learnt a little Cantonese after living in Hong Kong, but they only used it for less than 30% of their daily activities, so they were retained in the analysis.

approved by the Human Subjects Ethics Sub-committee of The Hong Kong Polytechnic University (PolyU) (Application number: HSEARS20201016001). Informed written consent was obtained from the participants in compliance with the experimental protocols and participants got paid after the experiment.

2.2. Design and stimuli

The target stimuli were 48 pictures of disyllabic Mandarin T3 sandhi words (Appendix 1 in supplementary materials). We selected these target words from a wide range of lexical frequencies taken from the Leiden Weibo Corpus (Van Esch, 2012), with a mean frequency of 14.92 occurrences per million words (SD = 38.91, range = 0.07 to 242.85). In addition to corpus frequencies, we also collected the subjective frequency ratings for the target words (on a 7-point Likert scale, 1 = 'no exposure to the word', and 7 = 'extremely frequent exposure to the word') from our participants using Qualtrics (Qualtrics, Provo, UT, USA, https://www.qualtrics.com) after they completed the EEG experiment (Mean rating score = 4.99, SD = 0.79, range = 3.11 to 6.70, highly correlated with the log-transformed Weibo corpus frequency: $r = 0.784$, $p < .001$). We selected 48 corresponding grayscale line drawings for these target words. Due to a lack of standardized pictures for T3 sandhi words, the target pictures were selected from multiple sources, with five pictures from the materials in Nixon et al. (2015), two pictures from the standardized MultiPic picture database (Duñabeitia et al., 2018), and the remaining 41 pictures adapted from the Internet clipart sources.

The target picture was preceded by a prime with a relatively long negative SOA (i.e., the time interval between the onset of the prime and that of the picture; jittered between 800 and 1000 ms, see below) following Pellet Cheneval et al. (2018). This negative SOA was adopted to avoid the overlap of the recognition processes of the primes and the planning processes of picture naming to ease the interpretation of the ERP activities following the target picture onset. In this way, we can more confidently attribute any differences found between the prime conditions to the planning processes of speech production instead of the recognition processes for different primes.

The prime was presented simultaneously in its visual (Chinese character) and auditory form (spoken word). The combined use of visual and auditory primes was motivated for several reasons. First, it was intended to boost and obtain a robust phonological facilitation effect when primes were presented with a long negative SOA, as previous research showed that the phonological facilitation effect often became weaker or absent when the SOA was negative with the only use of either visual or auditory primes (e.g., Schriefers et al., 1990; Q. Zhang et al., 2016). To confirm that a phonological priming effect could be obtained, we conducted a behavioral pilot study on another 25 native Mandarin speakers with a similar design to the current study (X. Chen et al., 2021), and we found a reliable phonological priming effect in reaction times even when the SOA was as long as −1200 ms. Second, the combination of visual and auditory primes could minimize potential visual confusion when only the visual primes were presented (e.g., misrecognition of similar-looking Chinese characters with different pronunciations such as 冶 and 治) or the wrong activation of homophones or acoustically similar sounds when only the auditory primes were presented. Moreover, only using the auditory T3 prime could potentially activate the target morpheme, leading to a confounding influence of morphological or semantic relatedness instead of phonological relatedness to the target words. Our pilot study (X. Chen et al., 2021) found that combining the auditory T3 prime with another visual homophonous character different from the target morpheme led to a much smaller priming effect than combining the auditory T3 prime with the same visual target morpheme. This suggests that using a different visual homophonous character could control for the confounding effect of activating the target morpheme introduced by the T3 auditory prime, and make the T3 prime condition better matched with the other two prime conditions in terms of morphological and semantic relatedness to the target words. Therefore, we opted to use both visual and auditory primes together in our current study.

For each target picture (e.g., 雨伞, *yǔ sǎn*, 'umbrella'), the prime is a different monosyllabic morpheme that shares the same syllable with the first morpheme of the target word. There were three types of primes (see Appendix 1 in supplementary materials): a T3 prime (underlying form, e.g., 语, *yǔ*, 'language/speech'), a T2 prime (surface form, e.g., 鱼, *yú*, 'fish'), and a control prime (tonally-unrelated, carrying T1 or T4, e.g., 玉, *yù*, 'jade'). Due to the limited number of available items, following several prior studies on speech production (e.g., Cai, et al., 2020; Feng et al., 2019; Q. Zhang and Damian, 2019), the target pictures were presented with all three types of primes for each participant. Moreover, each condition was also repeated to increase the experimental power (see Lorenz et al., 2021 for a similar practice). Thus, each target word was repeated 6 times (3 prime conditions × 2 repetitions), giving rise to a total of 96 trials for each prime type.

To further control the potential influence of other prime features, the selected visual primes paired with the auditory primes were neither orthographically nor semantically related to the target words for all three prime conditions. Moreover, we avoided using heterophonic characters (i.e., Chinese characters with multiple pronunciations in different semantic contexts) as the visual primes. In addition, we matched the three prime conditions in character frequency (log-transformed frequency per million from the Leiden Weibo corpus; T3 prime: Mean = 1.77, SD = 0.91; T2 prime: Mean = 1.77, SD = 0.80; Control prime: Mean = 1.77, SD = 0.63), visual complexity (number of strokes; T3 prime: Mean = 8.90, SD = 3.20; T2 prime: Mean = 8.90, SD = 3.02; Control prime: Mean = 8.90, SD = 3.21), and the tonal syllable frequency[7] (log-transformed raw token frequency of a syllable plus tone, based on the calculations from Da (2010); T3 prime: Mean = 5.14, SD = 0.52; T2 prime: Mean = 5.09, SD = 0.70; Control prime: Mean = 5.22, SD = 0.69).

The auditory primes were recorded by a young male native Standard Mandarin speaker from Shandong (aged 22), who acquired Standard Mandarin (Putonghua) since childhood and seldom spoke the local Mandarin dialect. The recording was conducted in a sound-proof booth with Audacity (Audacity Team, 2017) on a PC connected to an Audio-Technica AT2035 cardioid condenser microphone, with a 44,100 Hz sampling rate and 16 bits per sample. As for the T3 prime, we chose to use the half-third tone variant (/21/) instead of the full canonical form (/214/).[8] This was because the canonical form of T3 had been reported to be more likely to be perceptually confused with T2 by both native Chinese and non-native speakers (T. Huang and Johnson, 2011), partly because the final rising portion of the low dipping tone T3 shared acoustic similarity with the rising tone T2. The half-third variant of T3, on the other hand, is more acoustically distinct from both the rising tone T2 and the sandhi variant of T3. Note that a previous study showed that both the canonical (/214/) and the half-third variant of T3 (/21/) induced a similar magnitude of priming effects (Chien et al., 2021). To elicit a more natural production of the half-third variant, we asked the speaker to pronounce a disyllabic word containing the T3 target syllable followed by a T4 syllable with an obstruent onset (plosives, fricatives, or affricates), and then we extracted the first syllables from those recorded

[7] As suggested by one anonymous reviewer, we noticed that the type frequency of the tonal syllable (or the homophone density, i.e., the number of homophonous characters sharing the tonal syllable) was not perfectly matched between the three prime conditions due to the lower number of T3 homophonous characters than that of T2 and T1/T4 counterparts in Mandarin. However, including this variable as a covariate in the data analysis did not change the qualitative pattern of results, and this variable itself did not have a significant effect on RTs or ERPs.

[8] Another reason to choose the /21/ variant was that our male speaker tended to use a creaky voice when pronouncing the /214/ variant in isolation, which made the sound quality very poor after the duration normalization.

tokens. For the T2 and control primes, the speaker was asked to pronounce them in isolation as naturally as possible. For each prime, the speaker was asked to produce at least three tokens. All the audio stimuli were normalized to 300 ms in duration by Audacity to align the presentation duration of the auditory primes to that of the visual primes. The best resulting tokens, which had maximal naturalness, auditory clarity, and perceptual distinctiveness, were selected by the first author as the auditory primes. To ensure the quality of the processed audio stimuli, a trained phonetician (also a native Mandarin speaker from Northern China) was also asked to evaluate whether the recorded tokens could be recognized as the intended tones. Finally, the loudness of the selected audio stimuli was normalized by equalizing the peak amplitude (set to 0.7) by the Scale Peak algorithm in Praat (Boersma and Weenink, 2018).

To mask the experimental purpose, we included another 84 filler pictures (see Appendix 2 in supplementary materials), which were common objects or actions selected or adapted from the standardized MultiPic database (Duñabeitia et al., 2018) or other internet sources. The initial morpheme of the fillers did not contain any morpheme or syllable identical to the initial morpheme of the targets. Like the target words, the filler words were also repeated 6 times in total, resulting in 504 filler trials. All the fillers were preceded by monosyllabic visual primes (a different Chinese character from the initial morpheme of the fillers) together with the corresponding auditory forms. The primes of fillers were either phonologically related (sharing segmental syllable and tone or only sharing segmental syllable) or unrelated (no syllabic and tonal overlap) to the first morpheme of the fillers. In particular, we made the overall number of the phonologically related and unrelated primes identical across the experiment to reduce potential strategic processing caused by the phonological relatedness between primes and targets (see detailed descriptions in Appendix 2 in supplementary materials). Another eight pictures of disyllabic words, each paired with one unrelated prime, were selected for practice.

2.3. Procedure

The participants were tested in a sound-proof booth and seated approximately 70 cm in front of a 24.5-inch DELL LCD computer screen with a refresh rate of 60 Hz and a resolution of 1920 × 1080 pixels. The stimulus presentation and data collection were controlled by the software E-Prime 2.0 (Psychology Software Tools, Inc., Pittsburgh, PA) in a computer connected to a Chronos Box (a USB-based stimulus and response device). The auditory primes were presented synchronously with the visual primes via ER3 disposable form eartips connected to the Chronos Box. The Chronos Box was also connected to an Audio-Technica ATR-1100 dynamic microphone for the voice-key triggering and recording of the participants' oral responses. The recorded oral responses were saved as. wav audio files at a sampling rate of 48,000 Hz with 16 bits per sample.

To reduce the naming variability, we followed a procedure commonly adopted in previous studies on speech production (e.g., Lorenz et al., 2021) and trained participants to familiarize themselves with the names of all the pictures before the experiment. First, participants were given all the pictures together with their corresponding written names and instructed to pay attention to the mapping between the written words and the pictures. Next, they were asked to complete a picture naming test on a tablet computer, where they named all the pictures presented on the screen one by one in random order and were corrected by the experimenter if they named a picture incorrectly. Once they named all the pictures correctly, the experiment started. The familiarization procedure took about 30 min.

Each trial in the experiment underwent the same procedure as follows (Fig. 1A). It began with a fixation cross at the center of the screen lasting 500 ms. Then, a visual Chinese character prime in black 32-point

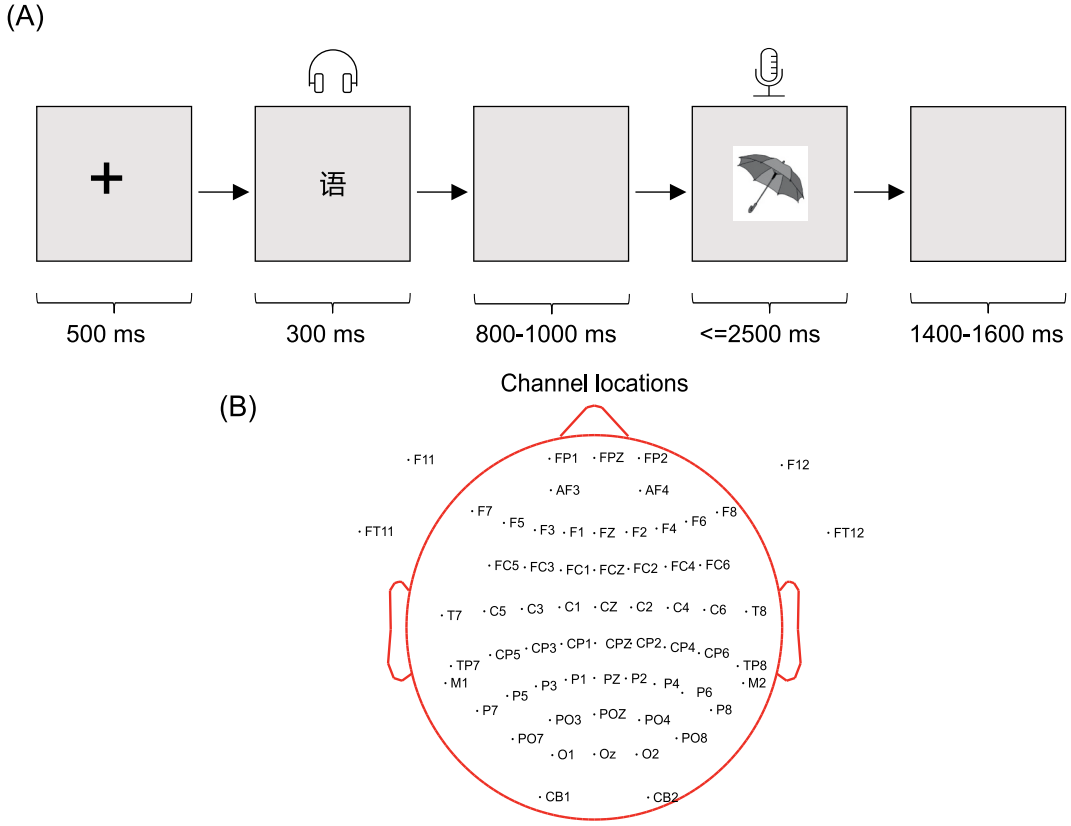


Fig. 1. (A) Illustration of the experimental procedure in one trial and (B) EEG channel locations.

Song font was presented for 300 ms at the same position on the screen. Simultaneously, the spoken form of the visual prime, with a duration of 300 ms, was played binaurally to the participants. This was followed by a blank screen randomly jittered between 800 ms and 1000 ms, with a mean duration of about 900 ms for the three conditions (with a total SOA between 1100 and 1300 ms, Mean = 1200 ms). The jittered SOA was adopted to minimize the transient effects of neural processing of the primes overlapping with the picture naming processes and reduce strategic expectancy (see a similar practice in X. Huang et al., 2014). The picture stimuli (scaled to 450 × 450 pixels, corresponding to a visual angle of 10.4°) then appeared at the center of the screen and remained there until the participants produced an oral response or the time limit of 2500 ms elapsed. Participants were instructed to overtly produce the corresponding words for the pictures as accurately and quickly as possible while ignoring the primes presented before the pictures. They were also asked to try to avoid disfluencies during their verbal responses. In addition, they were asked to remain relaxed and produce their oral responses softly to minimize the muscular artifacts. The onset of the audio recording for the verbal response for each trial was synchronized with the picture onset by E-Prime, so that the audios could be used for offline measurement of the naming latencies. After the picture disappeared, there was a blank inter-trial interval (ITI) randomly jittered between 1400 and 1600 ms (with a mean duration of about 1500 ms) before the next trial began.

The 48 target pictures were repeated in three lists, with each target picture paired with one type of prime appearing only in each session. Each list was further split into two blocks, resulting in six blocks to keep the block duration short. The three types of primes had equal probabilities to occur in each block. For each target picture, the order of the three prime conditions across the blocks was counterbalanced across the participants. Since several target words shared the same initial segmental syllable, these items were allocated into different blocks, so that the target words in each block did not have the same initial segmental syllable. The same target pictures were also not repeated in consecutive blocks. Then the six blocks were repeated an additional time, creating 12 blocks in total. The 84 filler pictures with six repetitions were also evenly divided into the 12 blocks. Thus, the experiment consisted of 792 trials (132 pictures × 6 repetitions), with each block containing 66 trials (including 24 target pictures and 42 filler pictures). Moreover, each block always started with three filler pictures. In each block, 12 out of the 42 filler pictures were paired with the visual primes from the targets (all phonologically unrelated to the first morpheme of the filler words). This was intended to prevent participants from implicitly learning the links between the primes and target words, and to minimize participants' anticipatory strategies due to target repetitions in the experiment. The order of the trials (except for the first three filler trials at the beginning) within each block was pseudo-randomized, and the primes between consecutive trials did not share the same syllable. Before the experiment, participants also received eight practice trials (another eight pictures paired with one unrelated prime) to get familiar with the procedure. Short breaks were allowed between blocks. After the whole experiment, participants were also asked to complete a subjective frequency rating task for the target words in Qualtrics. It took roughly 70 min on average to complete the picture naming task, and the entire experiment lasted about 2.5 to 3 h.

2.3.1. EEG recordings

The EEG data were recorded using SynAmps 2 amplifier (NeuroScan, Charlotte, NC, U.S.) at a sampling rate of 1000 Hz, with a cap (Quik-Cap Neo Net) carrying 64 Ag/AgCl electrodes on the scalp arranged based on the extended international 10–20 system (Fig. 1B). The vertical electrooculogram (VEOG) was recorded with two bipolar electrodes placed above and below the left eye respectively. The horizontal electrooculogram (HEOG) was recorded with another two bipolar electrodes placed lateral to each eye's outer canthi respectively. An electrode located between the CZ and CPZ served as the online reference. The impedance of all the electrodes was kept under 5 kΩ. The event markers for the picture stimuli were externally sent via the Cedrus StimTracker (Cedrus Corporation, San Pedro, CA) device to the EEG data acquisition computer.

2.4. Data pre-processing and analysis

2.4.1. Behavioral data

We did not use the automatic naming latencies registered by Chronos, as those automatic naming latencies contained inaccurate onset triggering (Roux et al., 2017). Instead, following previous speech production research (e.g., Bürki, 2017), naming latencies were measured offline from the recorded audio files in the software CheckFiles by manually annotating the acoustic onset of the oral responses (one sub-program of CheckVocal, see Protopapas, 2007) with the aid of both waveforms and spectrograms. This was done by a research assistant, a native Mandarin speaker with phonetic training who was blind to the experimental purpose. The acoustic onset was determined based on the acoustic characteristics of the beginning segment in the wideband spectrogram. To ensure the reliability of the measurement process, the first author, a trained phonetician, also measured the data of two participants at the beginning and then cross-checked the data measured by the research assistant. Latencies with a discrepancy larger than 10 ms between the two annotations (accounting for 8.16% of all trials) were examined and discussed by the first author and the research assistant to reach a consensus of the annotation criteria. Then the research assistant completed the latency measurements for the remaining participants, which were also double-checked by the first author. The first author and the research assistant also jointly examined the accuracy of participants' oral responses. Trials were coded as errors if the audio recordings of participants' oral responses fell into the following categories: (a) responding to the target picture with an incorrect word (or word fragment); (b) incomplete production of the target word; (c) containing pronunciation errors in the first morpheme of the target word, including segmental and tonal errors, or non-application of the T3 sandhi; (d) containing verbal disfluencies, including filler expressions (such as uh or oh), stuttering, repairs, long pause between the two syllables (larger than 300 ms), as well as other illegal sounds (like coughing); (e) containing no response (due to recording equipment failure or participants' attentional lapse). We excluded all the erroneous trials (2.15%), and trials with naming latencies shorter than 200 ms and longer than 2000 ms or beyond three standard deviations from the mean naming latencies of each condition for each participant (1.79%) (see a similar approach in Qu et al., 2020). The remaining naming latencies were log-transformed to reduce right skewness (Baayen, 2008) and then analyzed by the linear mixed-effect modeling (LMM) using lme4 package (Bates et al., 2015b) in R (R Core Team, 2021), with *p*-values estimated via the Satterthwaite approximation method implemented in the lmerTest package (Kuznetsova et al., 2017). Moreover, the raw naming latencies obtained from the behavioral data pre-processing were used to generate the speech onset markers, which were temporally aligned with the EEG signal for each participant.

2.4.2. EEG data

The EEG data were pre-processed using the EEGLAB toolbox (Delorme and Makeig, 2004), the plugin ERPLAB (Lopez-Calderon and Luck, 2014) and customized MATLAB (R2020b, The Mathworks, Inc.) scripts. The EEG signal was re-referenced offline against the average of left and right mastoids, bandpass filtered between 0.1 and 30 Hz using a zero-phase IIR Butterworth filter (slope: 12 dB/octave), and down-sampled to 500 Hz. Bad channels (no more than 4 channels per participant, detected with a kurtosis value larger than the threshold value of 10) were removed and later interpolated using the spherical spline method after artifact correction.

We extracted long epochs (−200 ms–2500 ms relative to the picture presentation onset) and ran the Independent Component Analysis (ICA)

to correct the ocular artifacts using the Adaptive Mixture ICA (AMICA) algorithm (Palmer et al., 2011).[9] Components associated with blinks and eye movements (2 up to 4 components for each participant) were identified by visual inspection and removed. We then removed epochs with erroneous responses (2.15%) and with extremely long naming latencies (longer than 2000 ms, 0.01%).[10] To avoid the potential influence of speech artifacts, unlike the pre-processing of the behavioral data, we followed a common practice in previous EEG research on speech production (e.g., Strijkers et al., 2010) and further excluded epochs with naming latencies shorter than 550 ms (4.38%, with the threshold selected based on the distribution of naming latencies). We then baseline-corrected the remaining epochs using the −100 ms–0 ms pre-stimulus time window (the same baseline time window was applied to both stimulus-locked and response-locked ERPs, see a similar approach in Jeong et al., 2021), and additionally rejected the epochs with amplitudes exceeding ±100 μV at any electrode in the time window between −100 and 550 ms relative to the picture presentation onset, as well as in the time window between 500 and 100 ms before the speech onset, or containing large muscular artifacts or electrode drifting 100 ms prior to the speech onset (2.15%). The average retention rate of the final epochs across the three prime conditions was 91.31% (SD = 6.87%), with 90.18% (SD = 7.87%) for the T3 prime condition, 91.64% (SD = 6.55%) for the T2 prime condition and 92.11% (SD = 7.24%) for the control prime condition. We then extracted the stimulus-locked epochs (−100 to 550 ms relative to the picture presentation onset) and the response-locked epochs (−500 ms up to the speech onset) from the pre-processed long epochs for further analysis.

Moreover, we adopted a new analytical method in recent EEG research on speech production (e.g., Bürki, 2017) and conducted the mass univariate analysis with the hierarchical general linear modeling using the LIMO plugin (Pernet et al., 2011) to investigate the effect of Prime type. Unlike conventional ERP analysis, the advantage of this new approach was to allow us to explore the effect of the predictor across the entire time course during production and across the entire scalp, and avoid the difficulty of selecting *a priori* time window(s) or electrode(s) for the analysis. Thus, this analysis was suitable for our current study, as previous literature did not provide informative knowledge of the time window for tonal encoding during the picture naming task. The LIMO plugin implemented a two-level analysis. In the first-level analysis, general linear regression models (with Prime type as the predictor) were respectively fit for each time point and electrode within each participant, and the beta coefficients for the three prime conditions were respectively estimated. Then the obtained beta coefficient matrix for the three prime conditions across the participants was entered into the second level of analysis and analyzed using a robust one-way repeated-measure ANOVA test. To make direct comparisons between different conditions of a categorical predictor, planned contrast was utilized in the ANOVA test in LIMO. Following our hypothesis, our primary contrasts of interest concerned the effect of the T3 and T2 prime against the control prime. Thus, we performed the ANOVA tests using two pre-specified coding schemes. To compare the T3 prime and the control prime, we coded the predictor as [T3: 1; T2: 0; Control: −1]; and to compare the T2 prime against the control prime, we coded the predictor as [T3: 0; T2: 1; Control: −1]. The uncorrected *p*-values of the one-way repeated-measure ANOVA tests for each time point/electrode were obtained via a bootstrap technique in LIMO (see Pernet et al., 2011 for more technical details), which was set to 5000 bootstrap times. Finally, the spatiotemporal clustering method (Maris and Oostenveld, 2007) was used to correct the multiple comparisons, with the neighboring distance set to 4 cm, the minimum number of channels for a cluster set to at least 2 (*minnbchan* = 2), and the alpha value set to 0.05 (Pernet et al., 2015). Data points with *p*-values exceeding .05 under the neighboring distance threshold were identified as a cluster. The sum of the *F* values of the data points in the clusters was computed and used as the cluster-level statistic (known as the cluster-mass value). The significance of the computed cluster-mass value was then assessed via a bootstrap technique (set to 5000 times). Any cluster(s) with a significant cluster-mass value found between the prime conditions served as an indication of the difference between conditions (for more technical details, see Pernet et al., 2015). The time window of interest was from 0 to 550 ms after the picture onset for the stimulus-locked analysis, and from 500 to 100 ms before the speech onset for the response-locked analysis. For the response-locked analysis, the last 100 ms before the speech onset was not analyzed to exclude potential contamination of the speech-related artifacts (e.g., Bürki, 2017).

Moreover, to further examine the potential interaction between the prime condition and lexical frequency, we also performed the single-trial analysis with LMM for both stimulus-locked and response-locked data. The voltage amplitudes were averaged in pre-selected time windows and electrodes for the stimulus-locked epochs or response-locked epochs in each trial. The selected time windows and electrodes were mainly based on the results from the cluster-based tests from LIMO (see similar practice from Frömer et al., 2018; Lorenz et al., 2021).

3. Results

3.1. Behavioral results

The behavioral data for the three prime conditions are summarized in Table 1 (also see Fig. 2). We did not analyze the accuracy data because the average accuracy rates were near ceiling for all the three prime conditions. We then used LMM to analyze the naming latency data and investigated the effect of Prime type and its potential interaction with Lexical frequency. We dummy coded the Prime type variable and used the control prime condition as the reference for comparison, since we focused on the effect of T3 and T2 prime against the control prime. For the Lexical frequency variable, we also ran separate analyses respectively using two z-scored frequency norms: the Weibo corpus frequency and the subjective frequency ratings collected from participants. The Weibo corpus frequency has been shown to have a superior predictive power for lexical processing data in Chinese (Sun et al., 2018) and subjective frequency has often been shown to account for a larger variance in lexical processing data than corpus frequencies and more faithfully capture the strength of lexical representations (e.g., Kuperman and van Dyke, 2013; Imai et al., 2005). The use of these two frequency norms could offer a complete picture of the role of lexical frequency in Mandarin T3 sandhi word production.

In the current study, we constructed the LMM in the following ways (including the LMM analysis on the EEG data below). Firstly, we

Table 1
Mean naming latency (in milliseconds) and mean accuracy rates with standard deviations (in parentheses) for the three prime conditions.

Prime Condition	Latency	Accuracy
T3 prime	718.1 (65.5)	97.92% (2.38%)
T2 prime	727.6 (69.0)	98.24% (2.83%)
Control prime	760.7 (74.9)	97.38% (2.87%)

[9] To achieve better ICA decomposition, the ICA was performed on another dataset high-pass filtered at 1.5 kHz (see Winkler et al., 2015) and the ICA weights were imported back to the original filtered dataset. The ICA was run using the resources provided by the NSG Portal (Sivagnanam et al., 2013, http://www.nsgportal.org).

[10] Unlike the pre-processing of the behavioral data, to get a sufficient number of trials for ERP analysis, we did not further remove those epochs with naming latencies beyond 3 standard deviations for each condition from each participant following the conventional ERP analysis (also see a similar approach by applying different trimming criteria for naming latencies and ERP data in Q. Zhang and Damian, 2019). But we did re-analyze the ERP data by additionally excluding those epochs, and obtained the same patterns as the ERP results reported in this paper.

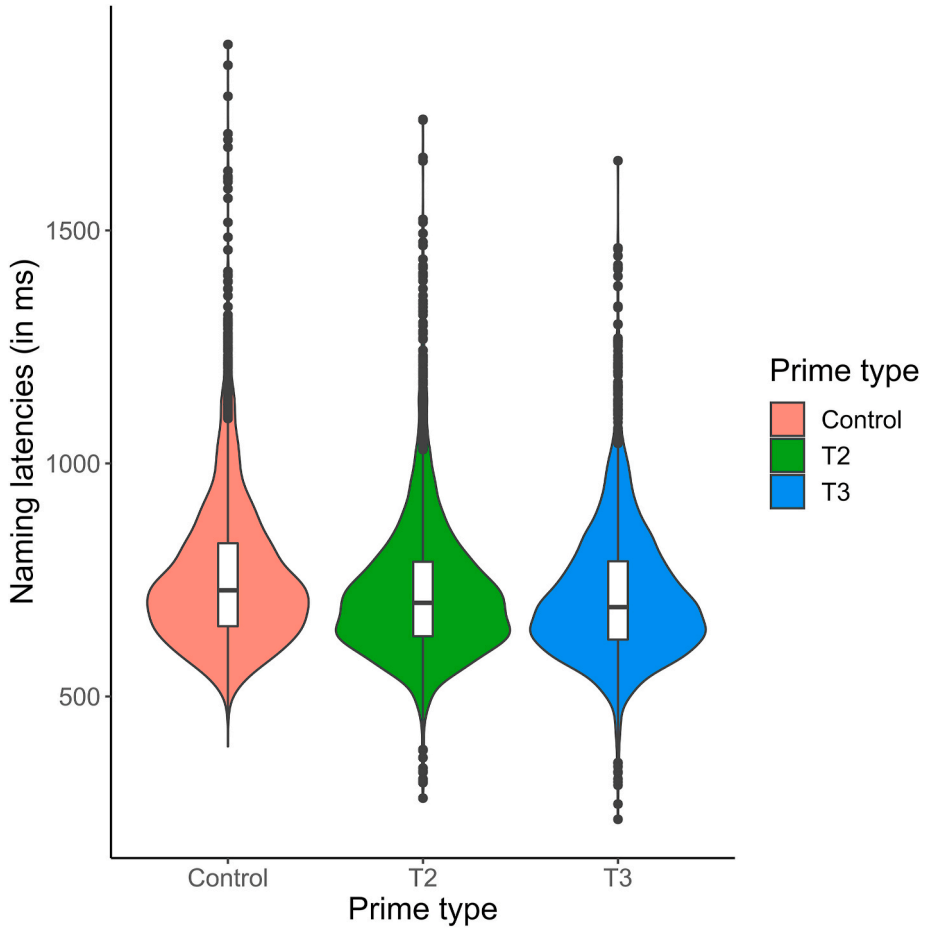


Fig. 2. Violin plots of naming latencies (in milliseconds) for the three prime conditions.

examined the effect of Prime type and built a model with only Prime type as the fixed-effect predictor. For the random effect structure, we firstly included the by-participant and by-item random intercept and slope for Prime type (Barr et al., 2013), but due to the singularity fitting and non-convergence problems caused by the complex random effect structure, we performed the principal component analysis using the rePCA function (Bates et al., 2015a) and reduced the random effect structure accordingly (i.e. by removing the correlations between random intercept and random slope, and then removing the random effects with the smallest variance) until the model was supported by the data. Then, we built additional models respectively for the two frequency norms to further explore the effect of Lexical frequency and its interaction with Prime type. For each frequency norm, we built a second model by further adding the lexical frequency norm (*i.e.*, Prime type + frequency norm) and a third model by further adding the interaction term between the lexical frequency norm and Prime type (*i.e.*, Prime type × frequency norm). Model comparisons with likelihood ratio tests were performed to assess the significance of the main effect of each lexical frequency norm and its interaction with Prime type.

Our results (see Table 2) revealed that both T3 and T2 primes elicited significantly shorter naming latencies than the control prime. For the Weibo corpus frequency, we did not observe statistical significance for its main effect (χ^2 (1) = 0.290, p = .590) or its interaction with the Prime type (χ2 (2) = 4.279, p = .118). There was only a marginally significant main effect of the subjective frequency (χ^2 (1) = 3.123, p = .077), with higher-subjective-frequency words tending to elicit shorter naming latencies than lower-subjective-frequency words. But the interaction between subjective lexical frequency and Prime type did not reach statistical significance (χ^2 (2) = 3.790, p = .150).

Table 2
Summary of the results of the final LMM for log-transformed naming latencies (Prime type was dummy coded, with Control prime as the reference baseline).

Predictors	*β*	*SE*	*df*	*t*	*p*
Intercept	6.617	0.020	72.9	337.245	**<.001**
T3 prime	−0.056	0.004	37.1	−12.917	**<.001**
T2 prime	−0.044	0.005	40.7	−9.618	**<.001**

Note. The formula of the LMM is: lmer(logLatency ~ PrimeType+(1|Subject)+(0+PrimeType|Subject)+(1|Item)+(0+PrimeType|Item); p values less than 0.05 were in bold.

3.2. EEG results

3.2.1. Mass univariate analysis

Stimulus-locked ERP: For the stimulus-locked ERPs, the mass univariate analysis revealed an overall significant main effect of Prime type (p = .002). Planned comparisons further revealed a significant difference between the T3 and the control prime (p = .001), driven mainly by a cluster between 318 and 550 ms (Fig. 3A). Moreover, the cluster was broadly distributed over the frontocentral scalp regions, which was also shown by the topographical maps (Fig. 3B). The separate average ERP waveforms for the selected electrodes in the cluster revealed more positive amplitude in the T3 prime than the control prime roughly after 320 ms of the stimulus-locked epochs (Fig. 3C). In contrast, no significant difference was found between the T2 prime and the control prime by the mass univariate analysis after cluster-test correction.

Response-locked ERP: For the response-locked ERPs, the mass univariate analysis showed that there was an overall significant main

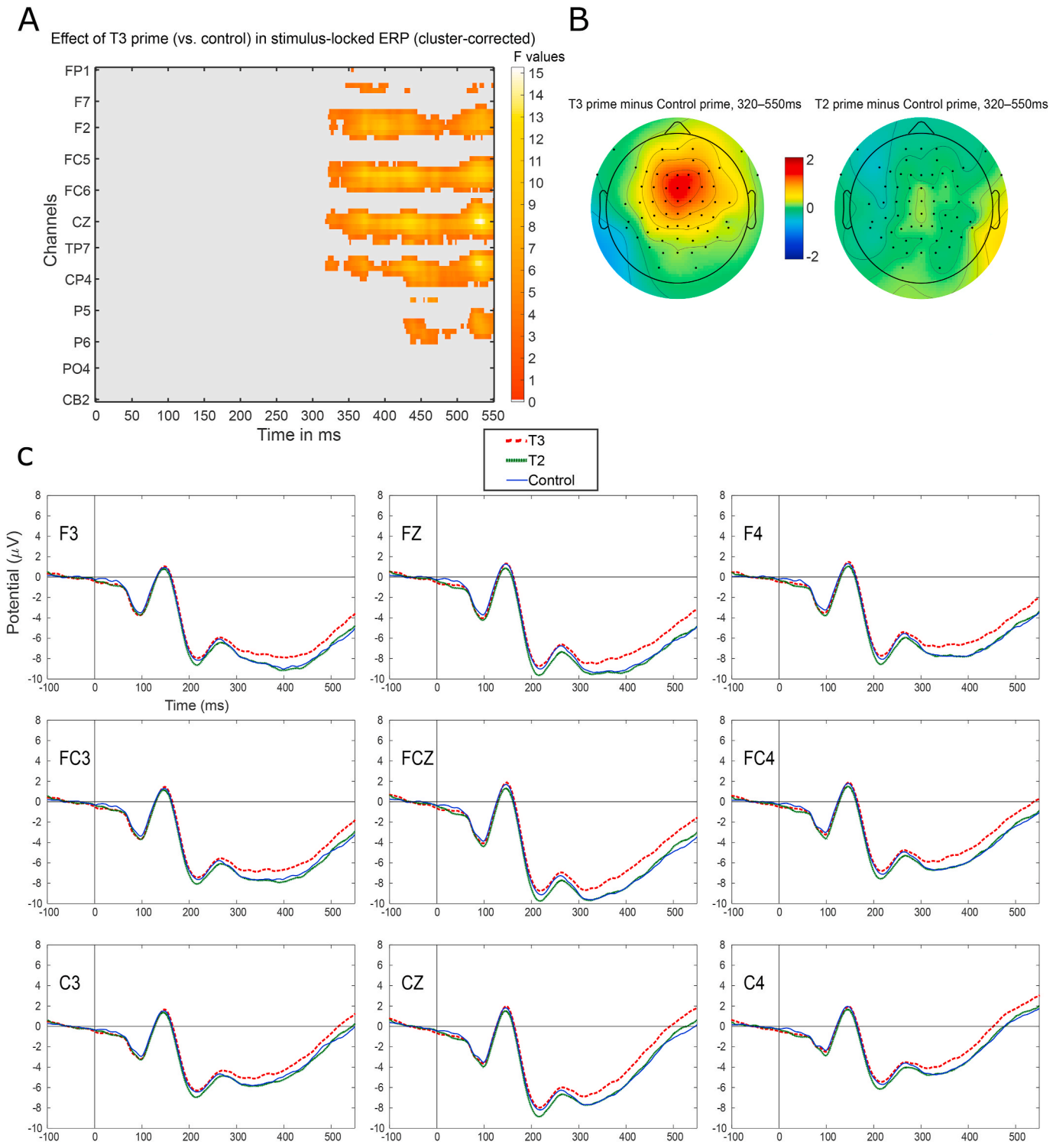


Fig. 3. (A) Results of the mass univariate analysis for the effect of T3 prime (compared to Control prime) in the stimulus-locked ERP after cluster-correction for multiple comparisons. The x-axis represents the time points of the analysis time window, and the y-axis represents individual electrode channels. The colored (orange) area represents the electrodes/time points with F values exceeding the statistical significance threshold. (B) Topographic maps of the difference between T3 and Control prime, and between T2 and Control prime in the 320–550 ms time window of stimulus-locked ERP. (c) Grand-average stimulus-locked ERPs for the three prime conditions in representative electrodes from frontocentral scalp regions. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

effect of Prime type ($p = .001$). Planned comparisons further revealed significant differences between the T3 and the control prime ($p = .020$), as well as between the T2 and the control prime ($p = .010$), but the differences seemed to occur in different time windows. The difference between the T3 and control prime was driven by a cluster between 500 and 406 ms before the speech onset (Fig. 4A). In contrast, the difference between the T2 and control prime was driven by another cluster between 236 and 100 ms before the speech onset (Fig. 4B). Moreover, the cluster for the T3 prime effect was broadly distributed across the anterior and central regions on the scalp, which had a similar scalp

distribution to the T3 prime effect found in the stimulus-locked ERPs (time-locked to the picture presentation onset). In contrast, the cluster for the T2 prime effect mainly had a left anterior and central distribution on the scalp. The separate average ERP waveforms for the selected electrodes in the two clusters (Fig. 4C) showed that the T3 prime elicited more positive amplitude than the control prime approximately in the −500 to −400 ms time window of the response-locked ERPs (time-locked to the speech onset). In contrast, the T2 prime elicited more negative amplitude mostly in the left frontal and central electrodes approximately in the −240 to −100 ms time window of the response-locked ERPs. Topographic plots of the two time windows based on the two clusters also confirmed these different scalp distributions respectively for the effect of the T3 and T2 primes (Fig. 4D).

3.2.2. LMM results

Stimulus-locked ERP: For the single-trial LMM analysis of the stimulus-locked ERPs, we first averaged the voltage amplitudes in the time window of 320–550 ms after the picture onset on a set of selected electrodes based on the LIMO results (FP1, AF3, AF4, FZ, F1, F3, F5, F2, F4, F6, FCZ, FC1, FC3, FC5, FC2, FC4, FC6, CZ, C1, C3, C5, T7, C2, C4, C6, T8, CPZ, CP1, CP3, CP5, CP2, CP4, CP6, PZ, P1, P3, P5, P2, P4, P6). The LMM construction procedure followed those described in the behavioral results section. Our results (see Table 3) showed that the T3 prime elicited a significantly larger positive amplitude than the control prime, but no significant difference was observed between the T2 prime and control prime. These results confirmed the findings of the mass univariate analysis. Further examination revealed a marginally significant main effect for the Weibo corpus frequency (χ^2 (1) = 3.438, p = .064), as high-corpus-frequency words tended to show more negativity than low-corpus-frequency words. However, the interaction between the Weibo corpus frequency and Prime type was not statistically significant (χ^2 (2) = 1.386, p = .500). As for the subjective frequency, neither significant main effect (χ^2 (1) = 0.643, p = .423) nor significant interaction with Prime type (χ^2 (2) = 1.778, p = .411) was found.

Response-locked ERP: For the single-trial LMM analysis of the response-locked ERPs, we focused on two time windows: 500–400 ms and 240–100 ms before the speech onset. For the first time window, we averaged the voltage amplitudes on the electrodes selected based on the LIMO results (FP1, AF3, AF4, FZ, F1, F3, F5, F7, F2, F4, F6, F8, FCZ, FC1, FC3, FC5, FC2, FC4, FC6, CZ, C1, C3, C5, C2, C4, CPZ, CP1, CP3). For the second time window, we averaged the voltage amplitudes on another set of electrodes selected based on the LIMO results (FP1, FPZ, FP2, AF3, AF4, F1, F3, F5, F7, F11, F2, F6, F8, F12, FC1, FC3, FC5, FC6, FT11, C3, C5, T7, CP5, TP7). Model construction and selection followed the same procedure previously described.

For the response-locked data in the -500–-400 ms time window before the speech onset, the model results (see Table 4) showed that only the T3 prime elicited a significantly larger positive amplitude than the control prime, whereas there was no significant difference between the T2 prime and control prime. These results converged with the results of the mass univariate analysis. Further examination revealed that there were no significant main effects for the Weibo corpus frequency (χ^2 (1) = 2.267, p = .132) or the subjective frequency (χ^2 (1) = 0.434, p = .510), nor were there any significant interactions between the Weibo corpus/subjective frequency and Prime type (Weibo corpus frequency: χ^2 (2) = 1.103, p = .576; subjective frequency: χ^2 (2) = 2.695, p = .260) in this time window.

Regarding the response-locked data in the 240–100 ms time window before the speech onset, our results (see Table 5) showed that compared to the control prime, the T2 prime elicited a significantly larger negative amplitude. In contrast, no significant difference was found between the T3 prime and the control prime. These results further confirmed the results from the mass univariate analysis. We did not observe statistical significance for the main effect of the Weibo corpus frequency (χ^2 (1) = 0.545, p = .461) or its interaction with the Prime type (χ^2 (2) = 2.165, p = .339). Moreover, there was no significant main effect for the subjective frequency (χ^2 (1) = 1.599, p = .206). But we observed a very weak interaction between the subjective frequency and Prime type (χ^2 (2) = 5.047, p = .080). Further exploration revealed that there seemed to be a larger difference between the control and T2 prime for the low-subjective-frequency words than for high-subjective-frequency words, suggesting a larger T2 priming effect on the low-subjective-frequency words than the high-subjective-frequency words. However, as this interaction effect was very weak, caution must be taken to avoid over-interpreting the result.

4. Discussion

4.1. Major findings and discussion

In this study, we found that the T3 and T2 prime yielded significant facilitation effects on the naming latencies compared to the control prime behaviorally. However, our EEG result showed that the effect of priming with a T3 prime elicited more positive amplitude primarily in the broad anterior and central scalp regions roughly between 320 and 550 ms after the picture onset, as well as roughly between 500 and 400 ms before the speech onset. In contrast, the effect of priming with a T2 prime elicited more negative amplitude primarily in the left anterior and central scalp regions roughly between 240 and 100 ms before the speech onset. The temporal differences for the T3 and T2 priming effects suggest that the encoding of the underlying form and the surface form occurred at different stages. Moreover, the polarity and topographic differences for the T3 and T2 priming effects also suggest that the facilitation effect on naming latencies by the T3 and T2 primes may involve different mechanisms.

Based on the meta-analysis by Indefrey and his colleagues (Indefrey and Levelt, 2004; Indefrey, 2011), they suppose that the picture naming task roughly takes 600 ms on average to complete, and the phonological encoding (including the retrieval of phonological codes and syllabification) starts from approximately 275 ms after the picture onset, and the phonetic encoding starts from 455 ms. In the current study, the effect of T3 prime starts approximately from 320 ms after the picture onset, which corresponds roughly to the phonological encoding stage. Note that we also observed the effect of T3 between 500 ms and 400 ms in the response-locked ERP. Although the response onset varied across trials and participants, there is reason to believe that the -500–-400 ms time window in the response-locked ERP showed temporally overlap with the later time window in the stimulus-locked ERP. If we consider the average naming latencies of those trials used for ERP analysis (c.a. 754 ms), -500–-400 ms before the speech onset roughly corresponded to 255–355 ms after the picture/stimuli onset, which could be considered to be the start of the phonological encoding stage and showed some overlap with the time window for the T3 priming effect in the stimulus-locked ERP (320–550 ms). The topographic distribution of these two clusters is also similar – both involved broad frontocentral electrode sites, suggesting that the two clusters in the stimulus- and response-locked ERPs may capture similar neural stages. Moreover, we found a high correlation (r = 0.749, p < .001) between the mean amplitude difference between the T3 and Control prime in the 320–550 ms time window stimulus-locked ERP, and the mean amplitude difference between the T3 and Control prime in the -500–-400 ms time window of the response-locked ERPs. Based on the temporal and topographic overlap and strong correlation, we reasoned that the cluster for the T3 prime effect in the response-locked ERP reflected the same stage of phonological encoding. However, this point should be further verified in future studies (e.g., via source localization based on high-density scalp EEG recordings). In general, it seems that the T3 prime mainly influenced the phonological encoding stage in T3 sandhi word production, and the underlying form of T3 was mainly retrieved and processed during this stage.

In contrast, the effect of the T2 prime on the ERP waveforms was only observed in the time window roughly between 240 and 100 ms before

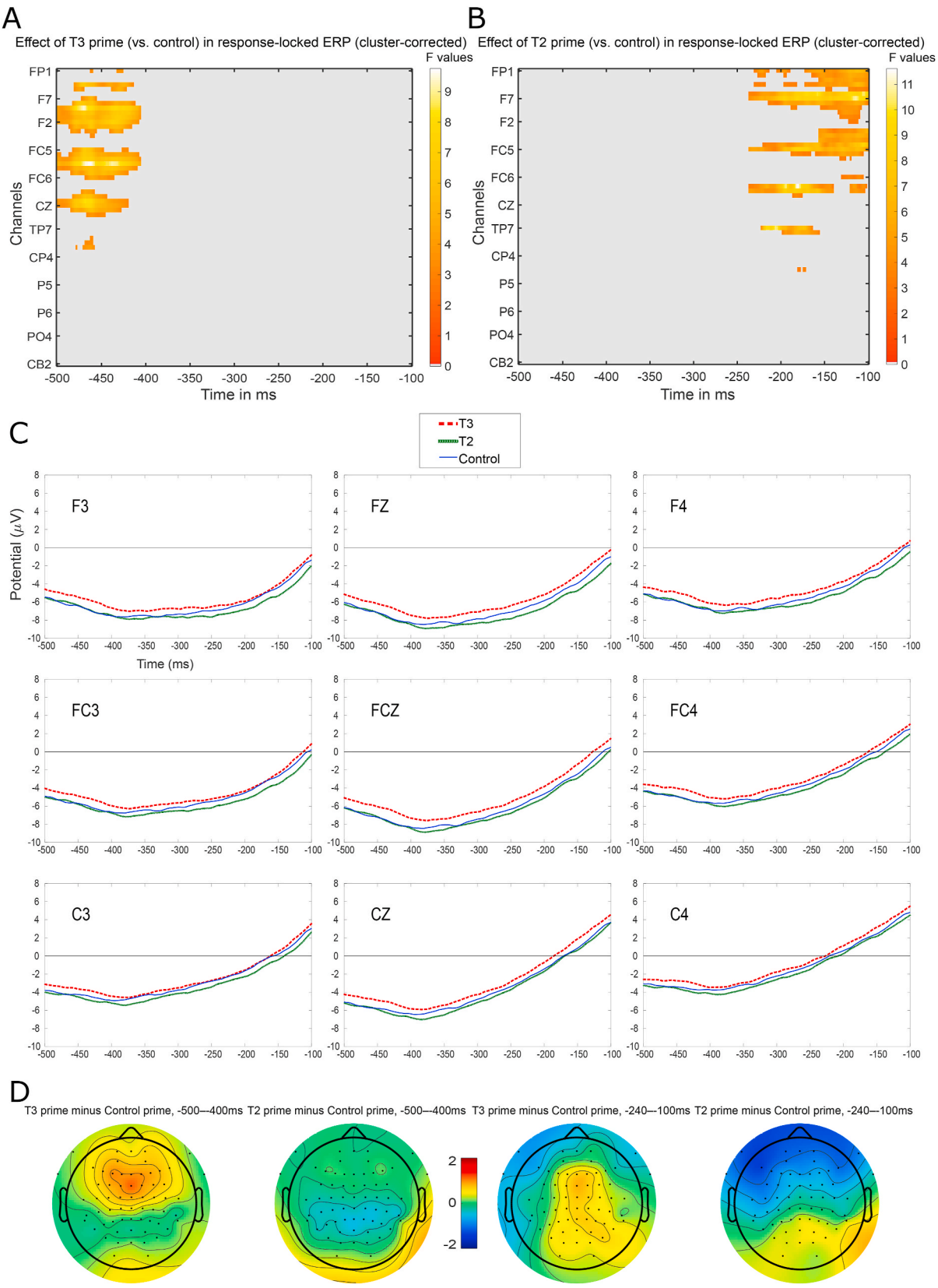


Fig. 4. Results of the mass univariate analysis for (A) the effect of T3 prime (compared to Control prime), and (B) the effect of T2 prime (compared to Control prime) in the response-locked ERP after cluster-correction for multiple comparisons, with the colored (orange) areas representing the electrodes/time points with F values exceeding the statistical significance threshold. (C) Grand-average response-locked ERPs for the three prime conditions in representative electrodes from fronto-central scalp regions. (D) Topographic maps of the difference between T3 and Control prime, and between T2 and Control prime in the -500–-400 ms time window, and in the -240–-100 ms time window of response-locked ERP. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Summary of the results of the final LMM for the stimulus-locked ERP data in the 320–550 ms time window after picture presentation onset (Prime type was dummy coded, with Control prime as the reference baseline).

Predictors	*β*	*SE*	*df*	*t*	*p*
Intercept	−2.674	0.874	53.0	−3.059	**.003**
T3 prime	0.847	0.240	9120.6	3.534	**<.001**
T2 prime	0.009	0.239	9120.7	0.037	.970

Note. The formula of the LMM is: lmer(Voltage ~ PrimeType+(1|Subject)+(1|Item); *p* values less than 0.05 were in bold.

Table 4
Summary of the results of the final LMM for the response-locked ERP data in the 500–400 ms time window before speech onset (Prime type was dummy coded, with Control prime as the reference baseline).

Predictors	*β*	*SE*	*df*	*t*	*p*
Intercept	−5.294	0.835	54.1	−6.340	**<.001**
T3 prime	0.763	0.255	9120.7	2.988	**.003**
T2 prime	−0.089	0.254	9120.9	−0.349	.727

Note. The formula of the LMM is: lmer(Voltage ~ PrimeType+(1 |Subject)+(1|Item); *p* values less than 0.05 were in bold.

Table 5
Summary of the results of the final LMM for the response-locked ERP data in the 240–100 ms time window before speech onset (Prime type was dummy coded, with Control prime as the reference baseline).

Predictors	*β*	*SE*	*df*	*t*	*p*
Intercept	−2.680	0.980	44.0	−2.734	**.009**
T3 prime	−0.100	0.298	9121.2	−0.334	.738
T2 prime	−0.941	0.297	9121.4	−3.169	**.002**

Note. The formula of the LMM is: lmer(Voltage ~ PrimeType+(1 |Subject)+(1|Item); *p* values less than 0.05 were in bold.

the speech onset close to the articulation. This time window could be attributed to the phonetic encoding or motor preparation stage. In particular, the T2 prime elicited more negative amplitude in the left anterior and central scalp regions. Based on the time window and channel locations, the effect of the T2 prime in the current study is probably linked to a left-lateralized slow-rising negative frontal component documented in previous EEG research on speech production (Riès et al., 2013, 2021), which is termed as the left-lateralized anterior negativity (LLAN) and thought to be associated with response preparation, motor planning and execution (Riès et al., 2021). Due to the pitch contour similarity between the T2 prime and the T3 sandhi variant (Nixon et al., 2015), the T2 prime probably brought benefits to the phonetic encoding or motor preparation stage, and the surface form was more likely to be encoded during this stage.

4.2. Discussion of online processing mechanisms for Mandarin T3 sandhi

Our results offer implications for the role of the underlying and surface form in the Mandarin T3 sandhi production. The priming effects observed for both the T3 and T2 prime in behavioral and ERP data are reminiscent of the results of Experiment 1 in Nixon et al. (2015) and suggest that both the underlying and the surface form are involved during T3 sandhi word production. This agrees with both the computation and multivariant account, but disconfirms the lexical storage account which assumes that only the surface form is involved in production.

More critically, regarding the time course of the activated underlying and surface forms in T3 sandhi word production, our ERP results reveal that Mandarin T3 sandhi word production probably undergoes a two-stage serial encoding process, consisting of the early retrieval of the abstract T3 category information (underlying form) and the later motor preparation for the context-specific tonal contour (surface form). This is compatible with the seriality assumption in the computation account, which predicts the facilitation effect of the T3 and T2 prime in different time windows. This is also consistent with the claim that the surface form is computed before the articulation (Politzer-Ahles and Zhang, in press) or before phonetic spell-out (J. Chen, 1999), but not compatible with the claim that the T3 sandhi is executed after the initiation of articulation (J. Chen et al., 2012). Furthermore, although our ERP results support the activation of the underlying and surface form in the T3 sandhi word production as specified in the multivariant activation account, our results do not seem to support the parallel activation of the underlying and surface form as suggested in Nixon et al. (2015), because it predicts that the effect of T2 prime should also be observed in the early ERP time window as the T3 prime, which is not the case according to our data. But if we assume that the underlying and surface form are serially activated in a hierarchical manner in different time windows during T3 sandhi word production, the multivariant activation account could also offer explanations for our data. That is, the abstract T3 category (underlying form) is first retrieved and the context-specific tonal contour (surface form) is only activated and selected at a later stage when the phonetic realization of T3 can be determined upon the following tonal information (as in the late selection hypothesis proposed by Miozzo and Caramazza, 1999). In this sense, the multivariant activation account is very similar to the serial processing described in the computation account.

However, it could still be argued that both underlying and surface form are activated in parallel, but the lack of the T2 prime effect in the early EEG time window may be because the T2 prime only activated the T2 tonal category, which could cause phonological conflicts in the retrieval of phonological codes. Alternatively, there is another similar account that assumes that the T3 non-sandhi and sandhi variants are stored as allomorphs for the same morpheme and both simultaneously activated in production (e.g., Tsay and Myers, 1996). It could be argued that only the T3 prime activated the corresponding T3 target allomorph but the T2 prime only activated a different T2 morpheme, so we only observed an early effect of the T3 prime and but not for T2. However, the above claims that the T2 prime could only activate the T2 tonal category or T2 morpheme are questionable. There is some evidence in recognition studies showing that T2 could also activate the T3 sandhi words (Meng et al., 2021, see also Speer and Xu, 2008, although their effect does not seem robust and remains to be replicated). Based on this finding, if we assumed that both the underlying and the surface form are activated in parallel, we would expect that the T2 prime could also activate the corresponding T3 underlying and surface form and show an effect as the T3 prime in the same ERP time window. However, the current study observed different time windows for the T3 and T2 prime, which is inconsistent with the assumption that both underlying and surface form are activated in parallel. Thus, our results may be better accommodated by the serial processing/activation view discussed earlier.

Taken together, our data do not unequivocally disentangle different theoretical accounts, but we may take our results to support serial processing, which involves the abstract T3 category in the early phonological encoding stage and the context-specific T2-like tonal contour in the later motor preparation stage. Further research, however, is needed to gather more evidence.

4.3. Discussion of the modulation of lexical frequency

In the current study, we also explored the effect of lexical frequency and its interaction with the prime type using two frequency norms: the Weibo corpus frequency and the subjective frequency. However, we only observed some very weak (marginally significant) effects, and these effects were not consistent across the two frequency norms. For example, regarding the subjective frequency, we observed a weak main effect for the naming latency data and a weak interaction effect with

prime type in the later time window of response-locked ERPs, whereas for the Weibo Corpus frequency, we observed a weak main effect in the stimulus-locked ERPs. Currently we do not have good explanations for these inconsistencies and speculate that these results could be due to the discrepancy between the two frequency norms, or due to the insufficient statistical power given the relatively small number of items (48 items) in the current experiment. Moreover, it should be noted that the repetition of targets and the presence of the phonologically related primes in the study design could potentially reduce the lexical frequency effect, making it difficult to interpret the weak main effect of lexical frequency or the weak interaction effect between prime type and lexical frequency. Moreover, as in Nixon et al. (2015), the frequency range of our stimuli may be restricted. Yuan and Chen (2014) only found the effect of lexical frequency on the acoustic realizations of T3 sandhi words when the extremely high-frequency words were included, and our stimuli did not cover the extremely high-frequency words due to the limited availability of qualified items. Thus, the modulation of lexical frequency was also inconclusive in the current study. Future studies with a larger number of items are required to better understand how lexical frequency modulates the T3 sandhi production.

4.4. Implications for other issues

Our current ERP data revealed different polarities for the T3 and T2 prime effects, with T3 eliciting more positive amplitudes and T2 eliciting more negative amplitudes. Moreover, the topographic distribution was also different, with the effect of T3 more broadly distributed in the frontocentral scalp regions and the effect of T2 located mostly in the left frontocentral scalp regions. We take both effects on the ERPs as an indication of facilitation, since we observed the global facilitation effect for both T3 and T2 primes on the naming latencies in the behavioral data.[11] Interestingly, a similar pattern of polarity and topography was also found in a recent ERP study investigating Cantonese word production (Wong et al., 2019), which showed that phonologically-related conditions elicited more bilateral positive amplitudes in the early time window but more left-lateralized negative amplitudes in the later time window compared to the phonologically-unrelated condition. However, since many previous studies did not yield consistent polarity effects for the phonological relatedness and those studies varied in terms of tasks and designs (see reviews in de Zubicaray and Piai, 2019; also see discussion in Q. Zhang and Damian, 2019), it remains unclear why the phonological relatedness led to different polarity effects, which awaits further investigation.

Our results offer some clues to the tonal representation for T3. The significant effect of the T3 prime using the half-third sandhi variant supports the presence of a shared underlying representation for all the T3 tonal variants (Chien et al., 2016, 2021), and cannot be explained by the surface representation view assuming the only storage of the surface form. However, the nature of the underlying representation is still debated, which has been argued to be the canonical form /214/ (as in the canonical representation view mentioned in Zhou and Marslen-Wilson, 1997), an abstract Low tone (e.g., Cheng, 1968), or an underspecified representation (e.g., Politzer-Ahles et al., 2016). Based on recent priming studies, it is probable that the underlying tone of T3 is represented as an abstract Low tone (Chien et al., 2020, 2021; Meng et al., 2021), but other studies have reported inconsistent findings (Politzer-Ahles et al., 2016; Zhou and Marslen-Wilson, 1997). As the current study was not designed to test the nature of the underlying representation, further research is still required.

Our EEG data also shed some light on lexical tonal encoding in Mandarin word production. To our knowledge, this study is the first EEG study to examine the encoding of lexical tones in overt Mandarin word production. Our results suggest that tonal encoding could start in the early phonological encoding stage, consistent with previous behavioral findings (Zhou and Zhuang, 2000). However, less is known about the relative encoding order of lexical tones and other phonological information such as segments and atonal syllables. Earlier ERP studies using meta-linguistic judgment tasks indicated that lexical tones may be encoded in a later time window than the onset segment but in a similar time window as the vowel segment in production (Q. Zhang and Damian, 2009; Q. Zhang and Zhu, 2011). However, the time window of tonal encoding in our study roughly corresponded to the time window of atonal syllable encoding reported in recent studies investigating overt Mandarin word production (e.g., Cai et al., 2020; Q. Zhang and Damian, 2019). As our study did not directly compare the encoding of tones, atonal syllables and segments, more research is needed to clarify these issues. Moreover, recent studies on speech errors seem to suggest that tonal encoding also involves a later tone-to-syllable mapping process linking the lexical tone to the syllabic frame (Alderete et al., 2019; Kember et al., 2015), but it is unknown how this could be reflected in the EEG signals, which awaits further research.

Our study also offers insight into the processing of phonological alternations in speech production. For productive phonological alternations like Mandarin T3 sandhi, our results suggest that an abstract underlying phonological representation could be first retrieved at the earlier phonological encoding stage, but the context-specific phonetic forms could be only derived or selected during the later phonetic encoding or motor preparation stage when the information of the following phonological context is available. However, since there are many different types of phonological alternations (Bürki, 2018; Bürki et al., 2010, 2011, 2014; Bürki and Gaskell, 2012), it remains to be investigated how the current results generalize to other types of phonological alternations. Even regarding tone sandhi, some recent research suggests that different processing mechanisms may be involved in the tone sandhi processing in other Chinese dialects (Chang et al., 2019; Chien et al., 2017; Yan et al., 2020, 2021). Moreover, the application of tone sandhi is also subject to other factors such as morpho-syntactic structure and prosodic structure (Chen, 2000), and recent research suggests that the processing of disyllabic Mandarin T3 sandhi words with different morphological structures (e.g., lexical compounds vs. reduplication) may also differ (Gao et al., 2021). Future research should elucidate the processing of different types of phonological alternations and the interaction with other high-level linguistic factors including syntax and semantics and delineate the conditions or contexts that will affect the processing mechanisms.

5. Conclusion

In summary, our experiment revealed facilitation effects of both the T3 and T2 prime on the production of T3 sandhi words in a phonologically-primed picture naming study, but their facilitation effects were observed in different ERP time windows. The results are taken to suggest that the abstract category-level (underlying form) and context-specific (surface form) phonological information are encoded at different stages of Mandarin T3 sandhi word production, with the former possibly encoded at the phonological encoding stage and the latter at the phonetic encoding or motor preparation stage.

Credit author statement

Xiaocong Chen: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Visualization, Writing – original draft, Writing – review & editing, **Caicai Zhang**: Conceptualization, Methodology, Resources, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing, **Yiya Chen**: Conceptualization, Methodology, Resources, Funding acquisition,Writing – review & editing **Stephen Politzer-Ahles**: Conceptualization, Methodology, Writing – review & editing **Yuyu Zeng**: Conceptualization, Methodology, Writing – review & editing, **Jie Zhang**: Conceptualization,

[11] But see Appendix 3 in the supplementary materials.

Methodology, Resources, Funding acquisition,Writing – review & editing.

Acknowledgments

This work was supported in part by the Postdoctoral Fellowships Scheme at the Department of Chinese and Bilingual Studies of the Hong Kong Polytechnic University awarded to CXC, the Departmental Reward Scheme for Research Publications in Indexed Journals awarded to CCZ, the Netherlands Organization for Scientific Research (VI.C.181.040) currently awarded to YC, and the National Science Foundation grant (BCS-1826547) awarded to JZ. We thank Mr. Jiayuan Fang for recording the audio materials, and Yiran Shan for helping measure the naming latencies, and Albert Chau and Xinyi Ye for their assistance in data collection. We also thank the two anonymous reviewers for their constructive comments on an earlier version of the manuscript. The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuropsychologia.2022.108367.

References

Alderete, J., Chan, Q., Yeung, H.-H., 2019. Tone slips in Cantonese: evidence for early phonological encoding. Cognition 191, 103952.
Arndt-lappe, S., Ernestus, M., 2020. Morpho-phonological alternations : the role of lexical storage. In: Pirrelli, V., Plag, I., Dressler, W.U. (Eds.), Word Knowledge and Word Usage: A Cross-Disciplinary Guide to the Mental Lexicon. Mouton De Gruyter, Berlin, pp. 192–227.
Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. J. Mem. Lang. 68, 255–278.
Bates, D., Kliegl, R., Vasishth, S., Baayen, H., 2015a. Parsimonious Mixed Models arXiv preprint arXiv:1506.04967.
Bates, D., Mächler, M., Bolker, B., Walker, S., 2015b. Fitting linear mixed-effects models using lme4. J. Stat. Software 67, 1–48.
Audacity Team, 2017. Audacity(R): Free Audio Editor and Recorder [Computer application], Version 2.3.3. https://audacityteam.org/.
Baayen, R.H., 2008. Analyzing Linguistic Data: A Practical Introduction to Statistics Using R. Cambridge University Press, Cambridge. https://doi.org/10.1558/sols.v2i3.471.
Boersma, P., Weenink, D., 2020. Praat, Version 6.1.15. http://www.fon.hum.uva.nl/praat/download_win.html.
Bürki, A., 2017. Electrophysiological characterization of facilitation and interference in the picture-word interference paradigm. Psychophysiology 54, 1370–1392.
Bürki, A., 2018. Variation in the speech signal as a window into the cognitive architecture of language production. Psychon. Bull. Rev. 25, 1973–2004.
Bürki, A., Alario, F.X., Frauenfelder, U.H., 2011. Lexical representation of phonological variants: evidence from pseudohomophone effects in different regiolects. J. Mem. Lang. 64, 424–442.
Bürki, A., Ernestus, M., Frauenfelder, U.H., 2010. Is there only one "fenêtre" in the production lexicon? On-line evidence on the nature of phonological representations of pronunciation variants for French schwa words. J. Mem. Lang. 62, 421–437.
Bürki, A., Gaskell, M.G., 2012. Lexical representation of schwa words: two mackerels, but only one salami. J. Exp. Psychol. Learn. Mem. Cognit. 38, 617–631.
Bürki, A., Laganaro, M., Alario, F.X., 2014. Phonologically driven variability: the case of determiners. J. Exp. Psychol. Learn. Mem. Cognit. 40, 1348–1362.
Cai, X., Yin, Y., Zhang, Q., 2020. The roles of syllables and phonemes during phonological encoding in Chinese spoken word production: a topographic ERP study. Neuropsychology 140, 107382.
Chang, C.H.-C., Kuo, W.-J., 2020. Neural processing of tone sandhi in production and perception: the case of Mandarin Tone 3 sandhi. In: Liu, H., Tsao, F., Li, P. (Eds.), Speech Perception, Production and Acquisition: Multidisciplinary Approaches in Chinese Languages. Springer, Singapore, pp. 117–135.
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, e0159835.
Chang, C.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, e83126.
Chang, C.H.-C., Lin, T.-H., Kuo, W.-J., 2019. Does phonological rule of tone substitution modulate mismatch negativity? J. Neurolinguistics 51, 63–75.
Chao, Y.-R., 1948. Mandarin Primer: an Intensive Course in Spoken Chinese. Harvard University Press, Cambridge, MA.
Chen, X., 2012. On the Role of the Underlying Tonal Representation in the Phonological Encoding Stage of Tone Sandhi Production of the Mandarin Third Tone. Guangdong University of Foreign Studies, Guangzhou, China. Unpublished MA thesis.
Chen, X., Zhang, C., Chen, Y., 2021. Different Time-Course of Activating Tonal Alternations in the Production of Mandarin Tone 3 Sandhi Words: Evidence from Reaction Time Survival Analysis. Poster presented at the 13th Society for the Neurobiology of Language Annual Meeting (SNL 2021).
Chen, J.-Y., 1999. The representation and processing of tone in Mandarin Chinese: evidence from slips of the tongue. Appl. Psycholinguist. 20, 289–301.
Chen, J.-Y., Chen, T.-M., Dai, Y., 2012. Cognitive mechanisms and locus of tone sandhi during Chinese spoken word production. In: Proceedings of the 3rd International Symposium on Tonal Aspects of Languages (TAL 2012). Nanjing, China, S2-01.
Chen, J.-Y., Chen, T.-M., Dell, G.S., 2002. Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. J. Mem. Lang. 46, 751–781.
Chen, M.Y., 2000. Tone Sandhi: Patterns across Chinese Dialects. Cambridge University Press, Cambridge.
Chen, Y., Shen, R., Schiller, N.O., 2011. Representation of allophonic tone sandhi variants. In: Proceedings of Psycholinguistics Representation of Tone. Satellite Workshop to ICPhS, pp. 38–41. Hong Kong.
Cheng, C.-C., 1968. English stresses and Chinese tones in Chinese sentences. Phonetica 18, 77–88.
Chien, Y.-F., Sereno, J.A., Zhang, J., 2016. Priming the representation of Mandarin tone 3 sandhi words. Lang. Cognit. Neurosci. 31, 179–189.
Chien, Y.-F., Sereno, J.A., Zhang, J., 2017. What's in a word: observing the contribution of underlying and surface representations. Lang. Speech 60, 643–657.
Chien, Y.-F., Yan, H., Sereno, J.A., 2021. Investigating the lexical representation of Mandarin Tone 3 phonological alternations. J. Psycholinguist. Res. 50, 777–796.
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. Front. Psychol. 11, 646.
Chomsky, N., Halle, M., 1968. The Sound Pattern of English. Harper and Row, New York.
Crowley, K.E., Colrain, I.M., 2004. A review of the evidence for P2 being an independent component process: age, sleep and modality. Clin. Neurophysiol. 115, 732–744.
Da, J., 2010. Chinese Text Computing: Syllable Frequencies with Tones. Retrieved from. https://lingua.mtsu.edu/chinese-computing/phonology/syllabletone.php.
de Zubicaray, G.I., Piai, V., 2019. Investigating the spatial and temporal components of speech production. In: de Zubicaray, G.I., Schiller, N.O. (Eds.), The Oxford Handbook of Neurolinguistics. Oxford Univerisity Press, New York, pp. 471–497.
Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134, 9–21.
Deng, Y., Feng, L., Peng, D., 2003. Research on articulatory rule of third tone sandhi of Standard Chinese in different contexts [in Chinese]. Acta Psychol. Sin. 35, 719–725.
Duñabeitia, J.A., Crepaldi, D., Meyer, A.S., New, B., Pliatsikas, C., Smolka, E., Brysbaert, M., 2018. MultiPic: a standardized set of 750 drawings with norms for six European languages. Q. J. Exp. Psychol. 71, 808–816.
Feng, C., Yue, Y., Zhang, Q., 2019. Syllables are retrieved before segments in the spoken production of Mandarin Chinese: an ERP study. Sci. Rep. 9, 11773.
Frömer, R., Maier, M., Rahman, R.A., 2018. Group-level EEG-processing pipeline for flexible single trial-based analyses including linear mixed models. Front. Neurosci. 12, 48.
Ganushchak, L.Y., Christoffels, I.K., Schiller, N.O., 2011. The use of electroencephalography in language production research: a review. Front. Psychol. 2, 208.
Gao, F., Lyu, S., Lin, C.-J.C., 2021. Processing Mandarin Tone 3 sandhi at the morphosyntactic interface: reduplication and lexical compounds. Front. Psychol. 12, 713665.
Hsieh, H.-I., 1970. The psychological reality of tone sandhi rules in Taiwanese. In: Papers from the 6th Meeting of the Chicago Linguistic Society. The Chicago Linguistic Society, Chicago, pp. 489–503.
Hsieh, H.-I., 1976. On the unreality of some phonological rules. Lingua 38, 1–19.
Huang, T., Johnson, K., 2011. Language specificity in speech perception: perception of Mandarin tones by native and nonnative listeners. Phonetica 67, 243–267.
Huang, X., Yang, J.C., Zhang, Q., Guo, C., 2014. The time course of spoken word recognition in Mandarin Chinese: a unimodal ERP study. Neuropsychology 63, 165–174.
Imai, S., Walley, A.C., Flege, J.E., 2005. Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. J. Acoust. Soc. Am. 117, 896–907.
Indefrey, P., Levelt, W.J.M., 2004. The spatial and temporal signatures of word production components. Cognition 92, 101–144.
Indefrey, P., 2011. The spatial and temporal signatures of word production components: a critical update. Front. Psychol. 2, 255.
Jeong, H., van den Hoven, E., Madec, S., Bürki, A., 2021. Behavioral and brain responses highlight the role of usage in the preparation of multiword utterances for production. J. Cognit. Neurosci. 33, 2231–2264.
Kember, H., Croot, K., Patrick, E., 2015. Phonological encoding in Mandarin Chinese: evidence from tongue twisters. Lang. Speech 58, 417–440.
Kim, K.H., Kim, J.H., Yoon, J., Jung, K.Y., 2008. Influence of task difficulty on the features of event-related potential during visual oddball task. Neurosci. Lett. 445, 179–183.
Kuo, Y., Xu, Y., Yip, M., 2007. The phonetics and phonology of apparent cases of iterative tonal change in Standard Chinese. In: Gussenhoven, C., Riad, T. (Eds.), Tones and Tunes (Volume 2): Experimental Studies in Word and Sentence Prosody. Mouton de Gruyter, Berlin, pp. 211–237.
Kuperman, V., van Dyke, J.A., 2013. Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. J. Exp. Psychol. Hum. Percept. Perform. 39, 802–823.

Kuznetsova, A., Brockhof, P.B., Christensen, R.H.B., 2017. lmerTest package : tests in linear mixed effects. J. Stat. Software 82, 1–26.
Laganaro, M., 2014. ERP topographic analyses from concept to articulation in word production studies. Front. Psychol. 5, 493.
Laganaro, M., Perret, C., 2011. Comparing electrophysiological correlates of word production in immediate and delayed naming through the analysis of word age of acquisition effects. Brain Topogr. 24, 19–29.
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 Lang. 120, 209–216.
Li, X., Chen, Y., 2015. Representation and processing of lexical tone and tonal variants: evidence from the mismatch negativity. PLoS One 10, e0143097.
Lopez-Calderon, J., Luck, S.J., 2014. ERPLAB: an open-source toolbox for the analysis of event-related potentials. Front. Hum. Neurosci. 8, 213.
Lorenz, A., Zwitserlood, P., Bürki, A., Regel, S., Ouyang, G., Abdel Rahman, R., 2021. Morphological facilitation and semantic interference in compound production: an ERP study. Cognition 209, 104518.
Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164, 177–190.
Meng, Y., Wynne, H., Lahiri, A., 2021. Representation of "T3 sandhi" in Mandarin: significance of context. Lang. Cognit. Neurosci. 36, 791–808.
Miozzo, M., Caramazza, A., 1999. The selection of determiners in noun phrase production. J. Exp. Psychol. Learn. Mem. Cognit. 25, 907–922.
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. Lang. Cognit. Neurosci. 30, 491–505.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychology 9, 97–113.
Palmer, J., Kreutz-Delgado, K., Makeig, S., 2011. AMICA: an Adaptive Mixture of Independent Component Analyzers with Shared Components. Swartz Center for Computational Neuroscience, San Diego, CA. Technical Report.
Pellet Cheneval, P., Bonnans, C., Laganaro, M., 2018. Does facilitation by phonological cuing in picture naming depend on the modality of the cue? Aphasiology 32, 204–232.
Peng, S.-H., 2000. Lexical versus "phonological" representations of Mandarin sandhi tones. In: Broe, M.B., Pierrehumbert, J. (Eds.), Papers in Laboratory Phonology V: Acquisition and the Lexicon. Cambridge University Press, Cambridge, pp. 152–167.
Pernet, C.R., Latinus, M., Nichols, T.E., Rousselet, G.A., 2015. Cluster-based computational methods for mass univariate analyses of event-related brain potentials/fields: a simulation study. J. Neurosci. Methods 250, 85–93.
Pernet, Cyril R., Chauveau, N., Gaspar, C., Rousselet, G.A., 2011. Limo EEG: a toolbox for hierarchical linear modeling of electroencephalographic data. Comput. Intell. Neurosci. 2011, 831409.
Politzer-Ahles, S., Schluter, K., Wu, K., Almeida, D., 2016. Asymmetries in the perception of Mandarin tones: evidence from mismatch negativity. J. Exp. Psychol. Hum. Percept. Perform. 42, 1547–1570.
Politzer-Ahles, S., Zhang, J., (in press). Evidence for the Role of Tone Sandhi in Mandarin Speech. J. Chinese Linguist. (Series No. 25): Studies on Tonal Aspects of Languages.
Protopapas, A., 2007. CheckVocal: a program to facilitate checking the accuracy and response time of vocal responses from DMDX. Behav. Res. Methods 39, 859–862.
Qu, Q., Feng, C., Hou, F., Damian, M.F., 2020. Syllables and phonemes as planning units in Mandarin Chinese spoken word production: evidence from ERPs. Neuropsychology 146, 107559.
R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from URL http://www.R- project.org/.
Riès, S., Janssen, N., Burle, B., Alario, F.X., 2013. Response-locked brain dynamics of word production. PLoS One 8, e58197.
Riès, S., Pinet, S., Nozari, N.B., Knight, R.T., 2021. Characterizing multi-word speech production using event-related potentials. Psychophysiology 58, e13788.
Roux, F., Armstrong, B.C., Carreiras, M., 2017. Chronset: an automated tool for detecting speech onset. Behav. Res. Methods 49, 1864–1881.
Schriefers, H., Meyer, A.S., Levelt, W.J.M., 1990. Exploring the time course of lexical access in language production: picture-word interference studies. J. Mem. Lang. 29, 86–102.
Shih, C.-L., 1986. The Prosodic Domain of Tone Sandhi in Chinese. University of California, San Diego, San Diego, US. Unpublished PhD dissertation.
Sivagnanam, S., Majumdar, A., Yoshimoto, K., Astakhov, V., Bandrowski, A., Martone, M. E., Carnevale, N.T., 2013. Introducing the neuroscience gateway. In: Proceedings of the 5th International Workshop on Science Gateways (IWSG'13), vol. 993. CEUR-WS.org.
Speer, S.R., Xu, L., 2008. Processing lexical tone in third-tone sandhi. Talk presented in Laboratory Phonology 11, 131–132. Wellington, New Zealand.
Strijkers, K., Costa, A., Thierry, G., 2010. Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. Cerebr. Cortex 20, 912–928.
Sun, C.C., Hendrix, P., Ma, J., Baayen, R.H., 2018. Chinese lexical database (CLD): a large-scale lexical database for simplified Mandarin Chinese. Behav. Res. Methods 50, 2606–2629.
Tang, P., Yuen, I., Rattanasone, N.X., Gao, L., Demuth, K., 2019. The acquisition of phonological alternations: the case of the Mandarin tone sandhi process. Appl. Psycholinguist. 40, 1495–1526.
Tsay, J., Myers, J., 1996. Taiwanese tone sandhi as allomorph selection. In: Proceedings of the Twenty-Second Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on the Role of Learnability in Grammatical Theory. The Berkeley Linguistics Society, Berkeley, CA, pp. 395–405.
Van Esch, D., 2012. Leiden Weibo Corpus. Retrieved from. http://lwc.daanvanesch.nl.
Wang, W.S.-Y., Li, K.-P., 1967. Tone 3 in pekinese. J. Speech Hear. Res. 10, 629–637.
Winkler, I., Debener, S., Müller, K., Tangermann, M., 2015. On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. In: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, Milan, Italy, pp. 4101–4105.
Wong, A.W.-K., Chen, H.-C., 2008. Processing segmental and prosodic information in Cantonese word production. J. Exp. Psychol. Learn. Mem. Cognit. 34, 1172–1190.
Wong, A.W.-K., Chiu, H.-C., Wang, J., Wong, S.-S., Chen, H.-C., 2019. Electrophysiological evidence for the time course of syllabic and sub-syllabic encoding in Cantonese spoken word production. Lang. Cognit. Neurosci. 34, 677–688.
Xu, Y., 1997. Contextual tonal variations in Mandarin. J. Phonetics 25, 61–83.
Yan, H., Chien, Y.-F., Zhang, J., 2020. Priming the representation of left-dominant sandhi words: a Shanghai dialect case study. Lang. Speech 63, 362–380.
Yan, H., Chien, Y.-F., Zhang, J., 2021. The representation of variable tone sandhi patterns in Shanghai Wu. Lab. Phonol. 12, 1–24.
Yuan, J., Chen, Y., 2014. 3rd tone sandhi in Standard Chinese: a corpus approach. J. Chin. Ling. 42, 218–236.
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. City University of Hong University Press, Hong Kong, pp. 256–282.
Zhang, C., Xia, Q., Peng, G., 2015. Mandarin third tone sandhi requires more effortful phonological encoding in speech production: evidence from an ERP study. J. Neurolinguistics 33, 149–162.
Zhang, J., Lai, Y., 2010. Testing the role of phonetic knowledge in Mandarin tone sandhi. Phonology 27, 153–201.
Zhang, J., Zhang, C., Politzer-Ahles, S., Pan, Z., Huang, X., Wang, C., Peng, G., Zeng, Y., 2022. The neural encoding of productive phonological alternation in speech production: evidence from Mandarin Tone 3 sandhi. J. Neurolinguistics 62, 101060.
Zhang, Q., Damian, M.F., 2009. The time course of semantic and orthographic encoding in Chinese word production: an event-related potential study. Brain Res. 1273, 92–105.
Zhang, Q., Damian, M.F., 2019. Syllables constitute proximate units for Mandarin speakers: electrophysiological evidence from a masked priming task. Psychophysiology 56, e13317.
Zhang, Q., Feng, C., Zhu, X., Wang, C., 2016. Transforming semantic interference into facilitation in a picture-word interference task. Appl. Psycholinguist. 37, 1025–1049.
Zhang, Q., Zhu, X., 2011. The temporal and spatial features of segmental and suprasegmental encoding during implicit picture naming: an event-related potential study. Neuropsychology 49, 3813–3825.
Zhou, X., Marslen-Wilson, W.D., 1997. The abstractness of phonological representation in the Chinese mental lexicon. In: Huang, H. (Ed.), Cognitive Processing of Chinese and Related Asian Languages. The Chinese University Press, Hong Kong, pp. 3–26.
Zhou, X., Zhuang, J., 2000. Lexical tone in the speech production of Chinese words. In: Proceedings of 6th International Conference on Spoken Language Processing (ICSLP 2000), vol. 2. ISCA, Beijing, China, pp. 45–50.