Brain & Language 270 (2025) 105636

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

Neural substrates for the encoding of the contextual tonal alternation: An fNIRS study of Mandarin third-tone sandhi in word production

Xiaocong Chen [a,b,*,1], Tai Yuan [c,1], Yiya Chen [d], Fumo Huang [a], Caicai Zhang [a,b,**]

[a] Neurocognition of Language, Music and Learning Lab, Department of Language Science and Technology, The Hong Kong Polytechnic University, Hong Kong SAR, China
[b] Research Centre for Language, Cognition, and Neuroscience, Department of Language Science and Technology, The Hong Kong Polytechnic University, Hong Kong SAR, China
[c] Department of Neurobiology and Behavior, Stony Brook University, NY, USA
[d] Leiden University Center for Linguistics and Leiden Institute for Brain and Cognition, Netherlands

ARTICLE INFO

Keywords:
Mandarin T3 sandhi
Word production
fNIRS
Tonal alternation
Priming

ABSTRACT

Phonological alternations are common in speech, but the neurocognitive mechanisms for their encoding during word production remain unclear. Mandarin Tone 3 sandhi is an example of phonological alternation, whereby the Tone 3 (T3), a low-dipping tone, changes to a Tone 2 (T2)-like rising tone when followed by another T3. Previous research indicates that both the underlying tonal category and the surface tonal variant are activated during T3 sandhi word production, but the neural substrates of these sub-processes remain unclear. Using Mandarin T3 sandhi as a case study, we employed functional near-infrared spectroscopy (fNIRS) to better understand the neural bases of phonological alternations. Participants completed a phonologically-primed picture naming task, with different tonal relationships between monosyllabic primes and T3 sandhi words manipulated. Behaviorally, we replicated the facilitatory effects of T3 and T2 primes on the naming latencies of T3 sandhi words, confirming the activation of both underlying and surface tonal information. Compared to control primes, the fNIRS data revealed reduced activation in left temporal and bilateral frontal regions during T3 sandhi word production following T3 primes, indicating facilitation in retrieving the underlying tonal category and/or the wordform of T3 sandhi words, which may proceed to the downstream articulatory planning and execution of the context-specific tonal contour. Conversely, increased activation in left temporal regions but decreased activation in frontal regions was found during T3 sandhi word production following T2 primes, implying higher lexical-phonological competition in the wordform retrieval but facilitation in articulatory planning. Our findings offer implications for understanding the neural encoding of phonological alternations.

1. Introduction

Phonological alternation refers to the instance in which a morpheme or a word undergoes phonological changes in different morphological or phonological contexts. For example, the plural morpheme "-s" in English can be realized as [s] or [z] depending on the voicing of the preceding phonemes. Despite the abundance of phonological alternations in many languages, the neural substrates for the encoding of phonological alternations have been neglected in current neurocognitive models of speech production (e.g., Indefrey, 2011; Indefrey & Levelt, 2004) and remain largely unknown. The tone sandhi phenomenon in some tonal languages is a well-known example of phonological alternation in the suprasegmental aspect, whereby the canonical/citation form of a lexical tone (i.e., tonal pronunciation in isolation) changes to different tonal variants in different phonological contexts (M. Chen, 2000). For instance, in Standard Chinese with four lexical tones (T1: /55/,[2] high-

* Corresponding author at: Room HHB729, Department of Language Science and Technology, The Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR, China.
** Corresponding author at: Room HHB710, Department of Language Science and Technology, The Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR, China.
E-mail addresses: xiaocong.chen@polyu.edu.hk (X. Chen), caicai.zhang@polyu.edu.hk (C. Zhang).
[1] The first two authors contributed to this paper equally.
[2] The number here denotes the tonal values from Chao's (1948) five-point-scale tonal notation system, with 5 representing the highest pitch value and 1 representing the lowest pitch value.

https://doi.org/10.1016/j.bandl.2025.105636
Received 14 February 2025; Received in revised form 25 August 2025; Accepted 26 August 2025
Available online 4 September 2025

level; T2: /35/, high-rising; T3: /214/, low-dipping; T4: /51/, high-falling), a T3 changes from its citation form /214/ (e.g., yu3 雨,[3] 'rain') to a high-rising tone /35/ (resembling the Mandarin T2, yu2) when followed by another T3 (e.g., yu3-san3 雨伞, 'umbrella'). This is known as the T3 sandhi rule (e.g., Chao, 1948; Peng, 2000; Shih, 1986; Wang and Li, 1967; Yuan and Chen, 2014; C. Zhang and Peng, 2013; J. Zhang and Lai, 2010), which could result in perceptually indistinguishable word pairs in Mandarin (e.g., the T3 sandhi word *bai3-mi3百米 'a hundred meters'* sounds almost homophonous to the T2 initial word *bai2-mi3白米 'rice'*, Wang and Li, 1967). Despite an increasing number of recent studies investigating the neurocognitive mechanisms behind Mandarin T3 sandhi (e.g., Chang & Kuo, 2016; Chang et al., 2014; X. Chen et al., 2022; C. Zhang et al., 2015; J. Zhang et al., 2022), the neural substrates for real-time Mandarin T3 sandhi word production are still not well understood. The goal of the current study is thus to investigate the brain networks for the sub-processes of Mandarin T3 sandhi word production using functional near-infrared spectroscopy (fNIRS), a non-invasive neuroimaging technique with both good ecological validity and reasonable spatial resolution.

1.1. Processing mechanisms of the encoding of Mandarin T3 sandhi in production

In previous literature, it has been intensely debated whether the real-time production of Mandarin T3 sandhi words involves a computation mechanism (Chang & Kuo, 2020; Politzer-Ahles & Zhang, in press; C. Zhang et al., 2015; J. Zhang & Lai, 2010) or a lexical/allomorph storage mechanism (Hsieh, 1970, 1976; Tsay & Myers, 1996). The computation mechanism assumes that the specific surface sandhi tonal variants (known as surface forms in generative phonology, see M. Chen, 2000) were not stored but generated online via a tone-substitution phonological rule from a presumed underlying abstract tonal representation (known as underlying form in generative phonology, see M. Chen, 2000) during production (Chang & Kuo, 2020; J. Zhang & Lai, 2010). In contrast, the lexical or allomorph storage mechanism assumes that the sandhi tonal variants were pre-stored as part of lexical items or allomorphs and directly retrieved during production (Hsieh, 1970, 1976; Tsay & Myers, 1996). The debate surrounding the above processing mechanisms has been examined by previous studies investigating the productivity of Mandarin T3 sandhi using nonce probe tests (i.e. combining presented tonal syllables into non-word sequences). This line of research revealed high productivity of the T3 sandhi rule even in pseudowords/nonwords (Deng et al., 2003; Tang, Xu Rattanasone, et al., 2019; Xu, 1997; C. Zhang et al., 2015; C. Zhang & Peng, 2013; J. Zhang & Lai, 2010). The findings from productivity studies are interpreted as being more consistent with the computation mechanism, which assumes computation of the sandhi variant on the fly in nonce words as well as in real words.

The online computation account is also further supported by two recent electroencephalographic (EEG) studies (C. Zhang et al., 2015; J. Zhang et al., 2022). C. Zhang et al. (2015) revealed a larger P200 amplitude when participants produced disyllabic sequences involving T3 sandhi (i.e., T3-T3) compared to those involving non-sandhi (a matched sequence of T2-T3). Moreover, this effect was not modulated by whether the disyllabic sequences formed real words or nonwords. Similarly, J. Zhang et al. (2022) found that T3 sandhi words elicited larger positivity in the EEG signals than non-sandhi words during the planning of speech production, and this effect was not modulated by lexical frequency. The larger EEG amplitudes when producing T3 sandhi sequences in both studies have been claimed to indicate more effortful phonological processing by the researchers, thus supporting the existence of extra online computation during T3 sandhi production. Moreover, the lexicality status or lexical frequency did not modulate the above effects in neural signals, which is also more consistent with the computation account.

However, the nonce probe tests used in the above-mentioned studies required participants to explicitly combine tonal syllables into sequences, which could encourage participants to resort to the computation mechanism (Chien et al., 2020; J. Zhang et al., 2022). Therefore, findings from this line of research may not be generalizable to the real-time production of sandhi words in more naturalistic tasks such as picture naming. Instead, other researchers have proposed that multiple levels of phonological representations, including the underlying tonal category and surface tonal contour variant, could be stored and activated during the real-time process of T3 sandhi word production, which is known as the multivariant activation account (Y. Chen et al., 2011; Li & Chen, 2015). For example, using a picture-word interference task (i.e., naming pictures while ignoring distractors superimposed on pictures), Nixon et al. (2015) found that the overall naming onset latencies of T3 sandhi picture names (e.g., fu3 dao3辅导, 'tutor') were shortened by both T3 (only sharing the underlying tonal category with the sandhi T3, e.g., 辅, fu3) and T2 (only sharing the surface tonal contour with the sandhi T3, e.g., 服, fu2) distractors compared to control distractors (carrying T1 or T4, without tonal overlap, e.g., 付, fu4). This result suggested that the underlying abstract tonal category and the surface context-specific rising tonal variant were both activated during the production of disyllabic T3 sandhi words. Moreover, by manipulating the SOA (stimulus-onset asynchrony, i.e., time interval between the distractor and the picture), Nixon et al. (2015) further found that the facilitatory effect of the T3 distractors tended to be more evident at the early SOA (0 ms). In contrast, the facilitatory effect of the T2 distractors was observed across the early and late SOAs (0 ms and 83 ms). These results suggested different time courses for the activation of the underlying tonal category and surface tonal contour information during Mandarin T3 sandhi word production.

The activation of both the underlying tonal category and surface tonal contour information was further evidenced by a recent EEG study (X. Chen et al., 2022), which employed a phonologically primed picture naming task (i.e., naming pictures preceded by primes). X. Chen et al. (2022) found that the naming latencies of Mandarin T3 sandhi words were facilitated by both T3 and T2 primes compared to control primes, replicating the results of Nixon et al. (2015). More importantly, their EEG data revealed that T3 primes elicited larger positive amplitudes in the frontocentral electrodes than control primes in an early time window (350–500 ms after picture onset and 500–400 ms before speech response onset). In contrast, T2 primes elicited larger negative amplitudes in the left frontocentral electrodes than control primes in a late time window (250–100 ms before speech response onset). The EEG results further indicated that T3 sandhi word production underwent two stages, consisting of an early retrieval of the underlying abstract tonal category and later specification and motor preparation of the context-specific tonal contour information. However, Chen et al. (2022) did not perform source localization of the neural effects using high-density EEG recordings. Given the limited spatial resolution of EEG, the neural substrates underlying the two stages of Mandarin T3 sandhi word production were not fully addressed.

1.2. Previous studies on neural substrates of Mandarin T3 sandhi production

As far as we are aware, two previous functional magnetic resonance imaging (fMRI) studies (Chang & Kuo, 2016; Chang et al., 2014) have investigated the neural substrates for Mandarin T3 sandhi production. Chang et al. (2014) asked the participants to overtly produce quadrisyllabic sequences, consisting of either repeated tones (e.g., T1-T1-T1-T1, T2-T2-T2-T2, T3-T3-T3-T3, T4-T4-T4-T4) or mixed tones (e.g., T2-T4-T3-T1) in Mandarin. Compared to sequences of repeated tones without T3 sandhi (i.e., T1-T1-T1-T1, T2-T2-T2-T2, T4-T4-T4-T4),

[3] The number after the syllable stands for the tonal category: 1 = Tone 1, 2 = Tone 2, 3 = Tone 3, 4 = Tone 4.

producing sequences of mixed tones and repeated tones involving T3 sandhi (i.e., T3-T3-T3-T3) activated broadly distributed regions of the speech-production network, including the left supplementary motor area, bilateral insula and bilateral putamen. In contrast, higher activation of the right posterior inferior frontal gyrus (IFG) was only observed when producing sequences of repeated tones involving T3 sandhi. Similar results were obtained in Chang and Kuo (2016), which also revealed higher activation of the right IFG in the production of disyllabic tonal sequences involving T3 sandhi (T3-T3) than that of disyllabic tonal sequences without T3 sandhi. Moreover, higher activation of the right IFG was observed only in the overt production condition, not in the covert production condition. These results highlighted the important role of the right IFG in Mandarin T3 tone sandhi production. As the authors argued, increased activation of the right IFG could not be attributed to the inherent articulatory difficulty of T3, because Chang and Kuo (2016) did not find activation differences between the production of T3 and other lexical tones in isolation.

The findings of the above fMRI studies have shed light on the debate over the mechanisms for Mandarin tone sandhi. As argued by Chang and Kuo (2020), the temporal lobe is responsible for storing phonological representations, while the frontal lobe is involved in online phonological and articulatory processing. The increased activation of IFG in T3 sandhi sequences observed in fMRI studies (Chang et al., 2014; Chang & Kuo, 2016) probably reflected that Mandarin T3 sandhi requires greater efforts in online phonological and articulatory processing, supporting the online application of the tonal substitution rule. As noted in Chang & Kuo (2020), this appears to align more with the online computation account but runs counter to the lexical/allomorph storage selection account. However, the current fMRI findings do not provide clear evidence to adjudicate between the online computation of a rule change from T3 to T2 or online selection from the stored variants of T3, due to two limitations in the existing fMRI studies.

Firstly, like nonce probe tests, the tasks used in the fMRI studies (Chang et al., 2014; Chang & Kuo, 2016) involved an explicit combination of presented tonal syllables into sequences, which may not fully capture the online processes of real-time production of sandhi words in naturalistic settings. Thus, the debate over the mechanisms underlying the T3 sandhi is not fully settled.

Secondly, as discussed above, recent EEG and behavioral studies using priming or interference paradigms (X. Chen et al., 2022; Nixon et al., 2015) have suggested that the real-time T3 sandhi word planning consists of two-stage sub-processes: the retrieval of the underlying abstract tonal category and the preparation of the context-specific tonal contour. While informative, the task paradigm used in the fMRI studies did not directly tap into these sub-processes. It is unclear how to map the right IFG observed in the fMRI studies to these sub-processes and whether other regions (e.g., temporal regions) are also involved. Taken together, the functional neuroanatomy for the specific sub-processes of T3 sandhi word production warrants further investigation that combines the priming/interference paradigms and the neuroimaging techniques with better spatial resolution (such as fMRI or fNIRS), which is the aim of the current study.

1.3. Implications for cortical regions subserving the sub-processes of T3 sandhi word production from previous literature

Although the neural substrates for the two-stage sub-processes of the T3 sandhi word production remain unclear, several critical cortical regions could be identified based on previous literature (see discussion in Chang et al., 2014; Chang & Kuo, 2020). Regarding the early retrieval of the underlying tonal category, the superior temporal gyrus (STG) and the middle temporal gyrus (MTG) may play a role. According to the neurocognitive model of word production proposed by Indefrey and his colleagues (Indefrey, 2011; Indefrey & Levelt, 2004), the retrieval of phonological codes (abstract phonological representation) is subserved by the left posterior STG and left posterior MTG (also see reviews by De Zubicaray, 2023; Nozari, 2022). This proposal was supported by fMRI studies (De Zubicaray et al., 2002; De Zubicaray & Mcmahon, 2009), which found less activation of the left posterior MTG/STG when participants named pictures with phonologically related distractors than with phonologically unrelated distractors. Similarly, in the production model proposed by Hickok (2012), speech production involves a sensory-motor loop for the encoding of higher-level speech information (e.g., syllable). Within this loop, it is assumed that the auditory targets for speech gestures (akin to the context-free invariant phonological representation) are stored and activated in the auditory cortex (located in the temporal lobe). Thus, based on these models, the temporal lobe is thought to be responsible for storing the long-term phonological representations (also see Hickok & Poeppel, 2007; Nozari, 2022), which could include the underlying abstract tonal representation.

This view is further supported by previous neuroimaging studies on tonal perception. For instance, in an fMRI study, L. Zhang et al. (2011) found that the mid portion of the MTG in the left hemisphere showed higher activation to the across-category lexical tonal stimuli compared to within-category lexical tonal stimuli among native Mandarin speakers. Si et al. (2017) used the intracranial recordings and revealed larger responses to the across-category lexical tonal pairs than the within-category lexical tonal pairs in both STG and MTG. In two recent fMRI studies, Feng and his colleagues found that the acoustic-invariant tonal category representation could be reliably decoded in the anterior-to-middle portion of the left STG (Feng et al., 2018; Feng et al., 2021). All these studies indicate that the temporal lobe stores the abstract tonal representation and could play a role in retrieving the tonal category information during production.

The inferior parietal lobe (IPL), or more specifically, the supramarginal gyrus (SMG) (a subregion in the IPL), is another region that could be involved in the early retrieval of tonal category information. In some fMRI studies on word production, the IPL was found to show stronger activation in picture naming with the presence of phonologically related primes/distractors compared to unrelated ones (Abel et al., 2009, 2012; Arrigoni et al., 2024; Diaz et al., 2014). Currently, the role of the IPL in speech production remains debated. Gow (2012) and Dell et al. (2013) suggested that the left SMG stores the articulatorily organized word-form representations, so the left SMG was thought to play a role in the word-form retrieval in production (De Zubicaray, 2023). Nozari (2022) also argued that the left SMG serves as the phonological buffer and is involved in the storage and active maintenance of phonological codes. Moreover, in Hickok's (2012) production model, it is postulated that the Spt (the posterior Sylvian fissure at the parieto-temporal boundary, which naturally extends to the IPL/SMG) is responsible for transforming the auditory representations (akin to abstract phonological representations) to motor representations (i.e., articulation-based codes) (also see Hickok & Poeppel, 2007). The left IPL is thus considered to be the sensorimotor interface in speech production (Chu et al., 2023; W. Zhang et al., 2023), which mediates the mapping between sound and articulation (Gow, 2012). Despite some controversy over the role of the IPL, the IPL could be involved in the retrieval of phonological representations, including abstract tonal representations, during speech production.

Regarding the later encoding of the surface tonal contour, this process may recruit the prefrontal regions, or more specifically, the inferior frontal gyrus (IFG) and the middle frontal gyrus (MFG). According to the neurocognitive model of word production (Indefrey, 2011; Indefrey & Levelt, 2004), the left posterior IFG (Broca's area, including pars opercularis/BA44 and pars triangularis/BA45) is thought to be engaged in syllabification (i.e., combining phonemic sequences into syllables) and possibly prosodification (i.e., combining segmental contents with the prosodic frame such as stress) at later stages of phonological encoding or at the articulatory-phonetic encoding stage during word production. This is evidenced by a previous fMRI study (De Zubicaray & Mcmahon, 2009), which showed reduced activation of the left IFG when participants named pictures with phonologically-related distractors compared

to phonologically unrelated distractors. Other fMRI studies also showed that the left IFG was involved in the production of phonemic and/or syllabic sequences (Bohland & Guenther, 2006; Ghosh et al., 2008; Papoutsi et al., 2009; Rong et al., 2018; Segawaa et al., 2015). Another intracranial EEG study (Sahin et al., 2009) also showed that the left IFG was involved in the computation of the phonological alternation like the case of the voicing alternations for the English plural morpheme "-s". Moreover, the left posterior IFG (more specifically, the pars opercularis/BA44) and the left posterior MFG (BA6) are also proposed to store the motor programs in Hickok's (2012) production model and be engaged in the later stage of speech planning. This is evidenced by recent studies with intracranial neural recordings (Duraivel et al., 2024; Khanna et al., 2024), which showed that the IFG and MFG were associated with the planning of the abstract speech units (syllables and phonemes). In addition, in a recently proposed model, Hickok et al. (2023) also argued that the left posterior MFG is a core area for prosodic planning, including the planning of both lexical and phrasal prosody (such as lexical tone/stress and intonation). These studies suggest that the IFG and MFG may play a role in the later planning of the context-specific tonal alternation (i.e., the surface rising tonal contour). In addition to the prefrontal regions, based on the recent production model (Hickok et al., 2023), the motor-related cortical regions (e.g., precentral and postcentral gyri) could also be involved in the execution of the lexical tones, as evidenced by recent studies on tonal production with intracranial EEG recordings (Y. Liu et al., 2023; Lu et al., 2023).

Note that previous research shows that the planning of word production mostly involves the left-lateralized brain network (e.g., the involvement of left STG/MTG and IFG, see reviews in Indefrey, 2011; Indefrey & Levelt, 2004). However, research on tonal perception indicates that brain regions from the right hemisphere are often involved in tonal processing (B. Liang & Du, 2018; K. Yu et al., 2020). This pattern was also observed in studies on tonal production. For instance, L. Liu et al. (2006) conducted an fMRI study to investigate tonal production using the adaptation paradigm, which was based on the observation that repetitions of a stimulus feature could decrease the activation of brain areas associated with the processing of the target stimulus feature. They asked participants to repeat characters and pinyin sharing either vowels or lexical tones in Mandarin. Compared to the baseline task, they found that both vowel repetition and tonal repetition involved reduced activation of the bilateral IFG. However, compared to vowel repetition, tonal repetition led to higher activation of the right IFG. Moreover, while both vowel and tonal repetition activated left-lateralized brain regions, the brain activations of tonal repetition were less left-lateralized. Consistent with the above findings, Chang and her colleagues (Chang et al., 2014; Chang & Kuo, 2016) also found the higher activation of the right IFG in producing sequences with tone sandhi compared to those without tone sandhi. A recent fNIRS study (Guo & Chen, 2022) also found that lexical tonal information during the covert production of a set of tonal monosyllables was decoded with higher accuracy rates based on the signals from the speech-related areas of the right hemisphere. Thus, based on these findings, the tonal planning process during the T3 sandhi word production could also involve the activation of those corresponding right hemispheric regions (including STG, MTG, IPL/SMG, IFG, MFG, and motor-related regions).

1.4. The present study

As pointed out above, although recent EEG research like X. Chen et al. (2022) provided valuable insights into the temporal dynamics of Mandarin T3 sandhi word production, the limited spatial resolution of EEG leaves open questions regarding the specific brain regions contributing to the sub-processes of real-time T3 sandhi word production. Thus, in the current study, we aimed to probe the brain regions involved in those sub-processes during T3 sandhi word production. To address this question, we chose the fNIRS technique. This technique has higher spatial resolution than EEG and could provide more reliable spatial information about the neural substrates underlying T3 sandhi word production than the EEG technique used in previous studies (X. Chen et al., 2022; C. Zhang et al., 2015; J. Zhang et al., 2022). As an optical non-invasive neuroimaging technique, the fNIRS signals measure the cortical hemodynamic changes in response to the neural activities by leveraging differential absorption rates of the near-infrared light for the oxygenated and deoxygenated hemoglobin (Ferrari & Quaresima, 2012). Despite its comparatively lower spatial resolution than fMRI, fNIRS can still provide good spatial information on the brain activities of the surface cortical areas, which are the focus of interest in our current study. More importantly, compared to fMRI, fNIRS avoids interference from loud gradient noises and is more robust to motion artifacts, thereby offering better ecological validity for investigating the neural substrates of speech production (Hitomi et al., 2021).

To specifically tap into the encoding sub-processes of the underlying tonal category and the surface tonal contour information, we followed previous studies (X. Chen et al., 2022; Nixon et al., 2015) and used a phonologically primed picture naming task. In this task, the participants were asked to name pictures of disyllabic T3 sandhi words preceded by a monosyllabic prime. We manipulated the tonal overlap between the primes and the initial morpheme of the target words carrying T3 sandhi (with an overlap in either the underlying tonal category or the surface tonal contour information) (See details in the Method section below). For the behavioral data, we predicted that sharing either the underlying tonal category or the surface tonal contour would facilitate the naming onset latencies, as already observed in previous studies (X. Chen et al., 2022; Nixon et al., 2015). Regarding the fNIRS data, based on our earlier review in Section 1.3, we selected the following regions of interest (ROI) in both hemispheres for analysis: the bilateral temporal areas (STG and MTG), the bilateral frontal areas (IFG, posterior MFG, precentral and postcentral gyri), and the bilateral inferior parietal areas (SMG). As previously inferred, these specific regions may be involved in different subprocesses of tonal encoding during word production. However, this hypothesis requires further confirmation through the fNIRS technique, particularly in the context of Mandarin T3 sandhi word production, as it could provide more fine-grained spatial information. We thus hypothesize that the encoding of the underlying tonal category and the surface tonal variant information may recruit different regions in T3 sandhi word production. Specifically, as discussed earlier, the temporal areas and the inferior parietal area may be mainly engaged in the retrieval of the underlying tonal category, while the frontal areas may be involved in the planning and execution of the surface pitch contour of the sandhi tone.

2. Method

2.1. Participants

Fifty-two participants attended the experiment, but only the data from forty-eight participants were retained, as three participants did not satisfy the requirement of language background (acquiring other non-Mandarin dialects from South China before puberty) and the fNIRS data from another participant were not recorded normally due to the device breakdown. All the remaining participants were native Standard Mandarin speakers (24 females, mean age = 25.2 years, SD = 3.7, range = 19 – 34). They were born and raised in Mandarin-speaking regions in

Northern China (regions to the north of the Qinling-Huaihe Line) and reported that they acquired Standard Mandarin (Putonghua) before the age of seven.[4] They were right-handed or ambidextrous (2 participants), as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). All of them had normal or corrected-to-normal vision and reported no hearing, language, or other neurological deficits. The experimental protocols were approved by the Human Subjects Ethics Sub-committee of The Hong Kong Polytechnic University (PolyU HK) (Application number: HSEARS20230403004). Participants were provided with informed written consent and paid for their participation.

2.2. Design and stimuli

We adopted the same design as X. Chen et al. (2022), presenting the monosyllabic prime simultaneously in both visual and auditory forms. For the target stimuli, we selected 36 imageable disyllabic words containing Mandarin T3 sandhi (e.g., 雨伞, yu3-san3, 'umbrella') and the corresponding grayscale line drawings from the stimuli of X. Chen et al. (2022). The monosyllabic primes for each target word shared the same syllable as the first morpheme of the target word but corresponded to a different morpheme. We manipulated the tonal relatedness between the primes and the target word and included three prime conditions for each target word: a T3 prime (sharing the underlying tonal category, e.g., 语, yu3, 'language'), a T2 prime (sharing the surface tonal variant, e.g., 娱, yu2, 'entertainment'), and a control prime (carrying T1 or T4 and having no tonal overlap, e.g., 玉, yu4, 'jade'). Semantic or orthographic relatedness between the primes and the target word was avoided, and no polyphonic/heterophonic characters were used as primes. We also matched the three prime conditions in character frequency (log-transformed frequency per million taken from the Leiden Weibo corpus (Van Esch, 2012); $F(2, 105) < 0.001$, $p = 1.00$), visual complexity (number of strokes; $F(2, 105) < 0.001$, $p = 1.00$), and the tonal syllable frequencies (log-transformed raw token frequency of a syllable plus tone extracted from Da (2010); $F(2, 105) = 0.345$, $p = 0.709$) (also see Table S2 in Supplementary materials for further details). All the target words and primes are listed in Table S4 in Supplementary materials.

We included another 60 disyllabic imageable items as fillers (see Table S5 in Supplementary materials), which were also selected from X. Chen et al. (2022). The initial morpheme of the fillers did not share any syllable with the initial morpheme of the target words. Similar to the experimental items, the fillers were preceded by primes that were either phonologically related (with syllabic overlap) or unrelated (without syllabic and tonal overlap) to the first morpheme of the fillers. Among the 60 fillers, we included 12 T3 sandhi words, which did not appear in the experimental items, with four paired with two phonologically related primes and one unrelated prime, four paired with one phonologically related prime and two unrelated primes, and the remaining four paired with three unrelated primes. The purpose of this design was to avoid the situation that all the sandhi words in the experiment were only paired with phonologically related primes (as in target words). The remaining 48 filler words were non-sandhi words. To make the number of the items with phonologically related primes and that of the items with unrelated primes identical across the experiment, four of the non-sandhi words were paired with three phonologically related primes, four paired with two phonologically related primes and one unrelated prime, four paired with one phonologically related prime and two unrelated primes, and the remaining 36 were paired with three unrelated primes. Another eight non-sandhi disyllabic words, paired with one unrelated prime, were selected as practice items.

The auditory forms of all the primes were digitally recorded in a quiet room at a sampling rate of 48,000 Hz and 16 bits per sample. They were elicited from a 26-year-old male native speaker of Beijing Mandarin, who pronounced at least three tokens of each monosyllabic prime in isolation. The best token of each prime was selected by the first author and evaluated by the third author. The duration of all the auditory primes was then normalized to 400 ms in Audacity (Audacity Team, 2017) and the average intensity was scaled to 70 dB using a customized Praat script (Boersma & Weenink, 2022).

2.3. Apparatus and procedure

The participants were tested individually in a quiet laboratory room. They sat approximately 70 cm in front of an LCD computer screen with a refresh rate of 60 Hz and a resolution of 1920 × 1080 pixels. The software E-Prime 3.0 (Psychology Software Tools, Inc., Pittsburgh, PA) was used to present the stimuli and collect data. The audios for the primes were presented through a pair of ER3 disposable form eartips connected to the Chronos Box. The participants' oral responses were recorded via an Audio-Technica ATR-1100 dynamic microphone connected to the Chronos Box and saved as.wav audio files sampled at 48,000 Hz with 16 bits per sample.

Before the experiment, the participants were trained to familiarize themselves with the names of all the pictures. They were first provided with printed pictures and their corresponding names, and then completed picture-naming tests on a tablet computer. The participants proceeded to the fNIRS experiment only after correctly naming all the items.

Following picture familiarization and the completion of the preparation for fNIRS data acquisition, the participants first completed the practice block before proceeding to the experimental blocks. Given the purpose and nature of our design, we followed Hitomi et al. (2021) and adopted a rapid event-related design. This approach allowed us to include a sufficient number of trials for each prime condition within the limited duration of the experiment. The same procedure was applied to all the trials (Fig. 1). The participants first saw a black fixation against the grey background lasting 200 ms at the center of the screen. Then a visual Chinese character prime in black 56-point Song font appeared at the center of the screen for 400 ms, accompanied by its auditory form (with the same duration) played binaurally to the participants. This was followed by a blank screen lasting 800 ms. The picture (scaled to 450 × 450 pixels) was then presented at the center of the screen for 2000 ms and disappeared, followed by a jittered blank inter-trial interval (ITI) (4000, 6000, 8000 ms) before the next trial started. Participants were asked to name each picture as quickly and accurately as possible upon seeing the picture while ignoring the preceding prime. Moreover, they were asked to remain relaxed, speak softly, and minimize their body movements to reduce artifacts in the fNIRS signals. The onset of each audio recording of the participants' responses was synchronized with the picture presentation onset by E-prime.

We arranged the primes and the target words in three lists, with each target picture paired with only one type of prime in each list. Each list was further split into two blocks, creating six blocks in total. The probabilities of the three types of primes were equal in each block and the filler pictures with three repetitions were then evenly allocated across the six blocks. The order of the items within each block was pseudo-randomized using Easy-Optimize-X (Spunt, 2016) before the experiment. Participants completed all the six blocks and thus saw each picture three times in the experiment (paired with a different prime each time). This resulted in 288 trials, with 96 trials per block. The order of the three prime conditions for each target picture was also counterbalanced across participants. Short breaks were allowed between blocks and the whole experiment took roughly 2 h.

[4] Twenty-four participants reported that they could additionally speak a local Mandarin dialectal variety (all are Northern Mandarin varieties, with seven participants speaking Central Plains Mandarin, three participants speaking Northeastern Mandarin, seven participants speaking Jilu Mandarin, two participants speaking Lanyin Mandarin, and five participants speaking Jiaoliao Mandarin). Six participants also reported acquiring some Cantonese after puberty and using it in less than 35% of their daily interactions after settling in Hong Kong.

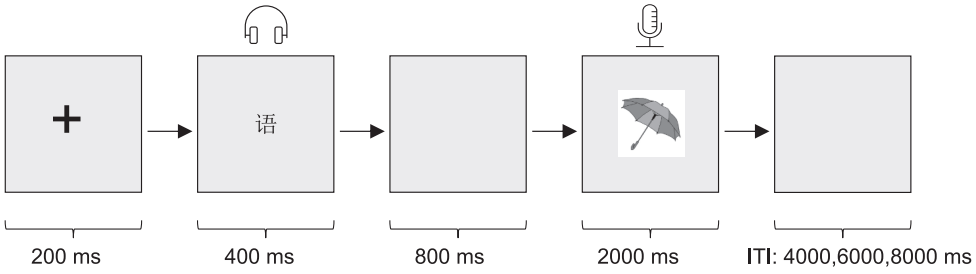

Fig. 1. Illustration of the experimental procedure in one trial.

fNIRS data acquisition

The fNIRS signals were concurrently recorded by the NIRSPort 2 continuous wave device (NIRx Medical Technologies, Glen Head, NY, USA) and the Aurora acquisition software at a sampling rate of 10.17 Hz, with two NIR light wavelengths (760 and 850 nm) emitted from the LED source optodes. A total of 31 optodes (consisting of 16 source emitters and 15 detectors, with a separation of approximately 3 cm) were positioned on the textile EasyCap (EasyCap, Herrsching, Germany) according to the 10–10 system, resulting in 44 standard channels. The montage of the optodes (see Fig. 2A) was pre-determined using fOLD (Zimeo Morais et al., 2018) in order to cover our ROIs across the bilateral frontal, temporal and inferior parietal regions. We also included eight short-distance detectors with a separation of 8 mm from the source to form eight short separation channels, which were evenly distributed across the scalp (see Fig. 2A). The short separation channels exclusively recorded the extracerebral signals (e.g., scalp and skull) irrelevant to the cortical activities, which could be employed for removing the confounding systemic physiological noises in the fNIRS data (see Section 2.4 below). The spatial localization of fNIRS channels was performed using NIRSite 2.0 (NIRx Medical Technologies, Glen Head, NY, USA) and NIRS_SPM (Ye et al., 2009). First, NIRSite 2.0 generated the scalp coordinates for optodes based on the ICBM-152 head model, with channel coordinates defined as the midpoints of source-detector pairs. The three-dimensional coordinates were then imported into NIRS_SPM, where spatial registration using the NFRI method aligned them to the ICBM-152 template, obtaining MNI-space coordinates for optodes and channels. Channels were further projected onto the cortical surface in NIRS_SPM (see Fig. 2B), and their anatomical coverages in the Automated Anatomical Labeling (AAL) atlas were extracted based on the cortical MNI coordinates (see Table S1 in Supplementary materials).

Before the experiment, each participant's head circumference was measured, and a proper cap was selected from the three available cap sizes (56 cm, 58 cm and 60 cm) accordingly. The cap was then placed on the participant's head, and the CZ point of the cap was aligned with the intersection of the nasion-inion line and the bilateral preauricular line marked on the scalp. To obtain better data quality, the participant's hair under each optode was pushed aside using wooden sticks and the optodes were then attached to the cap in the defined positions using spring holders. An overcap (62 cm) was used to block the environmental light. The signal quality was then evaluated using the Aurora calibration function and the relevant optodes were refitted if poor signals were detected. The experiment began only after the fNIRS signals of all the channels showed acceptable or good quality. During the experiment, event markers signaling the picture onset were sent via Chronos Box to the fNIRS device.

2.4. Data pre-processing

Behavioral data.

The participants' production accuracy for the target pictures was coded by the first author. Trials were regarded as errors when the audio recordings contained (a) incorrect responses (i.e., a different word other than the target word); (b) incomplete responses (i.e., only word segments) or no response; (c) pronunciation errors (segmental errors, or tonal errors including tone sandhi application errors); and (d) disfluencies (filler expressions such as uh, stuttering, repairs, long pause between the two syllables, coughing, etc.). Given the ceiling accuracy rates for all the three prime conditions (over 97 %, see Table S3 in Supplementary materials), we did not further analyze the accuracy data. Naming onset latencies were measured in the Praat software (Boersma & Weenink, 2022) using the audio recordings and speech response onsets were manually annotated by the first author based on characteristic acoustic features of onset sounds in spectrograms (see similar practices in Mantegna et al., 2025). For naming latency analysis and subsequent fNIRS analysis, we excluded trials with response errors (1.50 %) and with naming latencies shorter than 200 ms and longer than 2000 ms, as well as those beyond three standard deviations from the average latencies of each prime condition for each participant (1.58 %). The remaining naming onset latencies were log-transformed for further analysis.

fNIRS data.

Invalid trials (e.g., trials with errors or too short/long naming onset latencies) were removed. We excluded four standard channels located in the bilateral parietal regions (Channel 21, 22, 43 and 44) due to the malfunction of relevant source emitters. These channels mostly detected the activities from the bilateral angular gyri, which are primarily associated with lexico-semantic processing in word production (De Zubicaray, 2023; Nozari, 2022). Thus, these channels were less relevant to our interest in this study, as we focused on the phonological and phonetic encoding processes in production. For the remaining 40 standard channels and eight short separation channels, bad channels were identified using the following three methods. First, all channels were visually inspected by the second author using the software Homer (version 3.0, Huppert et al., 2009), and channels with flat waveforms, excessive motion, or a lack of cardiac/Mayer wave components were deemed bad. Second, channels with the Coefficient of Variation (CV) exceeding 25 % in either wavelength were also regarded as bad channels. Lastly, noisy channels were further identified using the hmrR_PruneChannels function from Homer.[5] All the identified bad channels (13.2 %) were excluded from further analysis.

The remaining data were then preprocessed using functions from the NIRS Brain AnalyzIR Toolbox (Santosa et al., 2018) and Homer 3 implemented in MATLAB. The raw light intensity data were first converted to the optical density signals. After conversion, motion artifacts were corrected using wavelet filtering (Molavi & Dumont, 2012), as recommended by previous fNIRS studies (Brigadoi et al., 2014; Hitomi et al., 2021). A Butterworth bandpass filter (0.01–0.5 Hz) was further applied to remove physiological noises such as heart rate and respiration. The optical density changes were then converted to the hemoglobin concentration changes using the modified Beer-Lambert law. The data were further downsampled to 0.5 Hz to reduce the computational

[5] The parameters of the hmrR_PruneChannels function were set as follows: Dynamic range of raw signal intensity: 1e-4 to 1; signal-to-noise threshold: 5; Source-detector distance range: 0 to 45 mm.

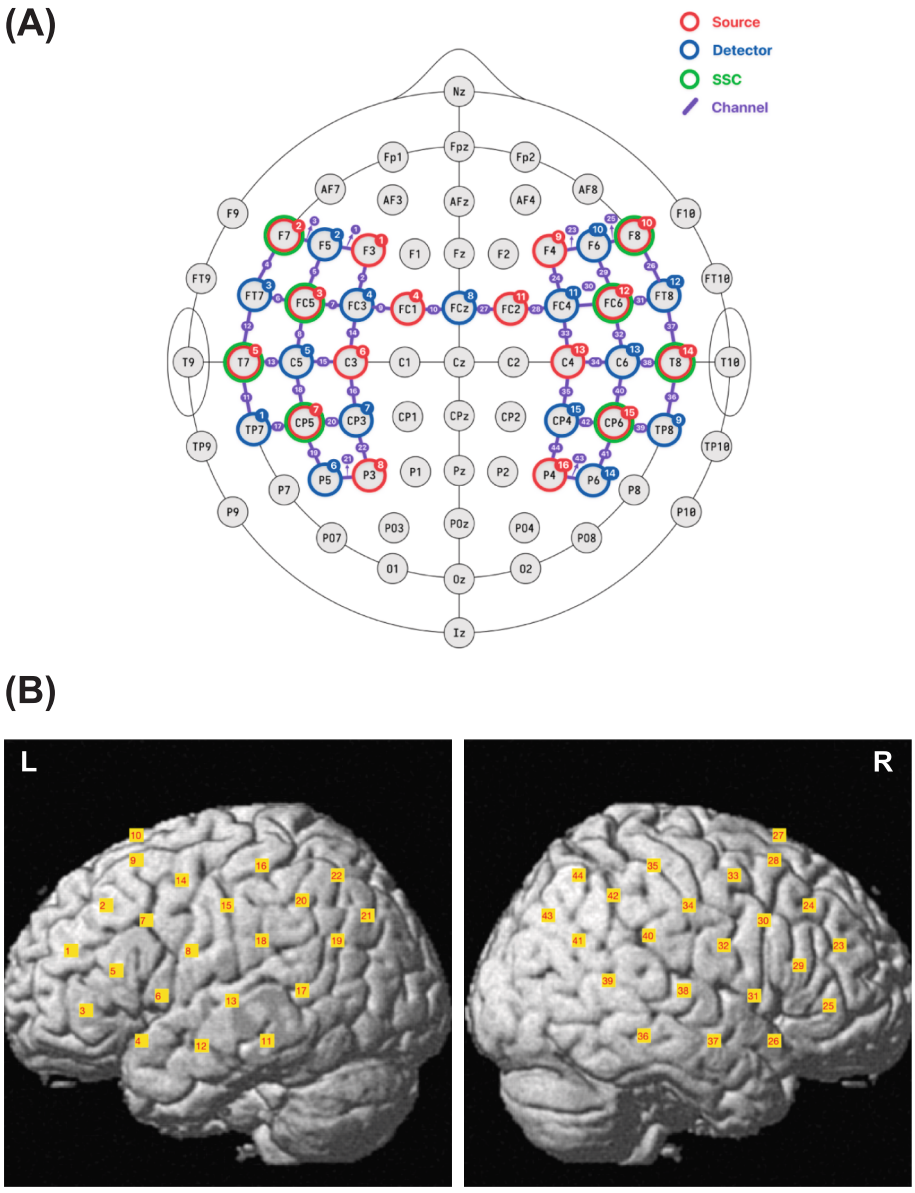

Fig. 2. (A) Montage of the optodes positioned according to 10–10 system, with sources in red circles and detectors in blue circles. 44 standard channels (in purple) were formed between each pair of source and detector. 8 short separation channels (SSC) were positioned around the sources in green circles. (B) Channel configuration rendered on the MNI template brain using NIRS-SPM. L: Left; R: Right. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

demands. Three types of measures were generated, including oxygenated hemoglobin (HbO), deoxygenated hemoglobin (HbR), and total hemoglobin (HbT) responses. As previous studies indicated that the HbO was a more sensitive measure of neural activities in overt picture naming (Hitomi et al., 2021; Moriai-Izawa et al., 2012), only the HbO data were further analyzed and reported in the current study.

The preprocessed HbO data were subjected to the two-level General Linear Modeling (GLM) analysis using the NIRS Brain AnalyzIR Toolbox. For the participant-level (first-level) analysis, a deconvolution method using the finite-impulse response (FIR) basis function was applied, as this method was reported to provide more accurate estimates of the hemoglobin signals for rapid event-related designs (Aarabi et al., 2017). The FIR basis function was constructed with eight time bins (with a bin width of 2 s), spanning from 0 to 16 s after the picture onset. We used an auto-regressive iterative robust least-squares (AR-IRLS) model to correct the autocorrelation in the error terms and the heteroscedasticity (Huppert, 2016; Santosa et al., 2018). To minimize the confounding systemic physiological noises from non-cortical layers, measurements from all the short separation channels with acceptable signal quality after preprocessing were added into the design matrix and statistically controlled (Santosa et al., 2020). The regression coefficients (beta weights) of each prime condition were obtained for each time bin and channel within each participant.

For the group-level (second-level) analysis, all the beta values of the eight time bins from all the participants were first fitted by a linear mixed-effect model using the NIRS Brain AnalyzIR Toolbox, with the prime condition as the fixed factor and the by-participant intercept as the random effect. We then performed post-hoc comparisons between conditions based on the fitted mixed-effect model using the paired *t*-test function implemented in NIRS Brain AnalyzIR Toolbox.

3. Results

3.1. Behavioral results

We conducted the linear mixed-effect model regression analysis on naming onset latencies using *lme4* package (Bates et al., 2015) in R (R Core Team, 2021)-. We included the Prime condition as the fixed-effect predictor. Following X. Chen et al. (2022), we focused on the comparisons between the T3 and the Control prime condition, as well as between the T2 and the Control prime condition. Thus, the Prime condition was dummy-coded, with the control prime condition serving as the reference level. The random effect structure was determined by first fitting a model with maximal random effect structure (Barr et al., 2013) and then simplifying the random effects until there were no singularity fitting and non-convergence problems. The final random effects included the by-participant and by-item intercepts. The p-values were estimated using the Satterthwaite approximation method by the lmerTest package (Kuznetsova et al., 2017).

The average naming onset latencies for the three prime conditions are shown in Fig. 3 (also see Table S3 in Supplementary materials). As shown in Table 1, the results of the mixed-effect model revealed that T3 and T2 primes both elicited significantly shorter naming onset latencies when producing T3 sandhi words than control primes.

3.2. fNIRS results

Given our primary interest, the planned post-hoc contrasts in t-tests were conducted only between T3 and Control prime condition, and between T2 and Control prime condition. The planned contrasts were restricted to the data from the second to the fifth time bins (i.e., the time window from 2 to 10 s after picture onset),[6] as this time window mainly captured the peak of the hemoglobin signal changes (also see a similar practice in Hitomi et al., 2021). P-values from multiple channels were adjusted using the False discovery rates (FDR) method (the Benjamini-Hochberg procedure).

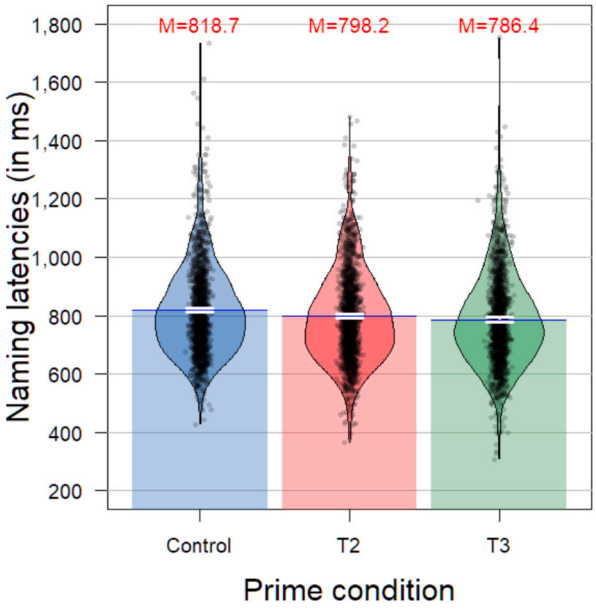

Fig. 3. The pirate plot showing the average naming onset latencies (in milliseconds) for the three prime conditions. The points represent the raw data points, the bars/blue lines represent the mean in each prime condition, the beans represent the density distribution for each prime condition, and the white bands in the middle denote the 95% confidence intervals of the mean. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Summary of the results of the final LMM for log-transformed naming onset latencies (Prime condition was dummy coded, with Control prime as the reference baseline).

Predictors	*β*	*SE*	*df*	*t*	*p*
intercept	6.692	0.020	82.1	338.191	<0.001
T3 prime	−0.044	0.005	4939	−8.797	<0.001
T2 prime	−0.028	0.005	4939	−5.688	<0.001

Note. The formula of the LMM is: lmer(logLatency ~ Prime+(1|Subject)+(1|Item). *SE*: standard error; *df*: degree of freedom.

The statistical results for the planned comparisons were summarized in Table 2. As shown in Table 2 (also see Fig. 4A and Fig. 5), compared to the Control prime condition, the T3 prime condition induced decreased activity in multiple regions encompassing both hemispheres during the production of T3 sandhi words, including the left temporal regions (anterior-to-mid portion of MTG/STG and posterior MTG, Channel 4, 11 and 12), left IFG (Channel 3 and 5), right IFG (Channel 29) and right postcentral gyrus (Channel 32). In contrast, the T2 prime condition induced reduced activity in the left IFG (Channel 5) and right MFG (Channel 24), but increased activity in the left temporal regions (the mid-to-posterior portion of MTG/STG, Channel 12 and 13) during the production of T3 sandhi words (also see Fig. 4B and Fig. 5).

Table 2
Results of channels with statistically significant differences in brain activation for post-hoc comparisons.

Channel	Anatomical Region	*df*	*t*	*FDR-corrected p*
T3 vs. Control Prime				
3	L inferior frontal gyrus (pars triangularis)	1128	−3.937	0.002
4	L superior temporal gyrus	1128	−7.125	<0.001
5	L inferior frontal gyrus (pars triangularis)	1128	−3.820	0.002
11	L middle temporal gyrus	1128	−3.494	0.006
12	L middle temporal gyrus	1128	−6.567	<0.001
29	R inferior frontal gyrus (pars triangularis)	1128	−3.674	0.004
32	R postcentral gyrus	1128	−5.205	<0.001
T2 vs. Control Prime				
5	L inferior frontal gyrus (pars triangularis)	1128	−3.085	0.046
12	L middle temporal gyrus	1128	3.615	0.009
13	L middle/superior temporal gyrus	1128	3.783	0.007
24	R middle frontal gyrus	1128	−3.902	0.007

Note. L: Left; R: Right; *df*: degree of freedom (error).

[6] Following Hitomi et al. (2021), we directly specified the time bins of interest (i.e., second to fifth time bins) in the post-hoc contrasts when performing the t-tests based on the group-level results of the fitted mixed-effect model. Thus, there was no need to average the beta values from different time bins. For further details, please see our scripts at https://osf.io/e2pka/.

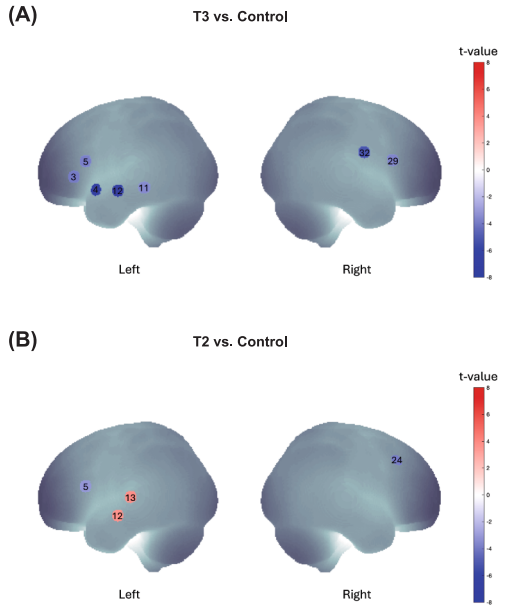

Fig. 4. (A) fNIRS channels showing significant differences in the beta weights between the T3 prime and the control prime conditions (FDR-corrected *ps* < 0.05) mapped on the cortical areas; (B) fNIRS channels showing significant differences of the beta weights between the T2 prime and the control prime conditions (FDR-corrected *ps* < 0.05) mapped on the cortical areas. The color denotes the t-statistics for the respective T3 and T2 prime conditions compared to the control prime condition, with positive values in red indicating increased neural activities compared to the control prime condition, and negative values in blue indicating decreased neural activities compared to the control prime condition. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. General discussion

4.1. Functional neuroanatomy of T3 sandhi word production

As expected, our behavioral data revealed facilitation effects on the naming onset latencies for both T3 and T2 primes against control primes. This finding fully replicated the behavioral results of previous studies (X. Chen et al., 2022; Nixon et al., 2015) and further supported the activation of both the underlying tonal category and surface tonal contour information during T3 sandhi word production.

More importantly, our fNIRS data revealed partially distinct brain activation patterns during the production of T3 sandhi words preceded by T3 and T2 primes, compared to control primes. For the T3 prime condition, we observed reduced activation in the left temporal regions (STG and MTG), including the left anterior/mid portion and posterior portion. One interpretation of this finding is that it may reflect the ease of retrieving the abstract phonological representation (more specifically, the underlying tonal category) during speech planning. As indicated by previous neuroimaging studies on speech perception (Joanisse et al., 2007; Liebenthal et al., 2005; Liebenthal et al., 2010), the left mid-to-posterior portions of the temporal regions (STG and MTG) are associated with the storage of high-level lexical-phonological representations. Previous research on tonal processing also suggests that the abstract tonal categorical representations are likely stored in the left anterior-to-mid portion of the temporal regions (Feng et al., 2018; Feng et al., 2021; Si et al., 2017; L. Zhang et al., 2011). Since the left temporal regions are considered to house the long-term abstract phonological representations (including the underlying tonal category), most production models have assumed that the left temporal regions (STG and MTG) are actively engaged in retrieving the abstract phonological representations (De Zubicaray, 2023; Hickok, 2012; Indefrey, 2011; Indefrey & Levelt, 2004; Nozari, 2022). Thus, it is possible that the T3 primes may facilitate the retrieval of the underlying tonal category, which presumably occurs during the phonological encoding stage, leading to reduced activity in the left temporal regions during the T3 sandhi word planning. This interpretation aligns with the findings of the previous EEG study by X. Chen et al. (2022), which indicated that the T3 prime benefited the early tonal retrieval process during the production of Mandarin T3 sandhi words.

Another alternative explanation is that the reduced activation in the left temporal regions following the presentation of T3 primes may reflect facilitation in the activation of the lexical entry for the target T3 sandhi words. Previous research has indicated that the left mid-to-posterior portion of the MTG is associated with lexical selection in production (see reviews in De Zubicaray, 2023; Nozari, 2022); other researchers have also proposed that the left posterior MTG may play a role in lexical representation and mediate the mapping between sound and meaning (Gow, 2012; Hickok & Poeppel, 2007). As suggested by recent priming studies on the recognition of Mandarin T3 sandhi words (Chien et al., 2016, 2021; Gao & Lin, 2025; Meng et al., 2021), the monosyllabic T3 primes could activate the wordforms of those disyllabic words beginning with the T3 (including the T3 sandhi words). This may lead to faster access to the target T3 sandhi words and the corresponding wordforms, thus resulting in reduced neural activities in the left temporal regions during production.

In addition, the T3 prime condition elicited reduced activation in the bilateral frontal regions (primarily the bilateral IFG) during T3 sandhi word production. As mentioned in the introduction, the left IFG (Broca's area) is argued to be engaged in downstream articulatory-phonetic encoding (De Zubicaray, 2023; Hickok, 2012; Indefrey, 2011; Indefrey & Levelt, 2004; Nozari, 2022), particularly in the planning of phonemic and syllabic sequences (Bohland & Guenther, 2006; Duraivel et al., 2024; Ghosh et al., 2008; Papoutsi et al., 2009; Rong et al., 2018; Sahin et al., 2009; Segawaa et al., 2015), or in the speech motor programming (Hickok, 2012). Previous clinical studies also reported deficits in tonal production among aphasic patients from tonal languages with lesions to the Broca's area (Gandour et al., 1996; J. Liang & Van Heuven, 2004), suggesting that the left IFG may be critical in the planning of lexical tones. In addition to the left IFG, the right IFG has also been found to be activated in tonal production (L. Liu et al., 2006), and more specifically, in the production of tonal sequences with Mandarin T3 sandhi (Chang & Kuo, 2016; Chang et al., 2014). Chang and her colleagues argued that the right IFG may be engaged in the online computation of the surface tonal contour for the T3 sandhi rule application (also see Chang & Kuo, 2020). Based on the above findings, the reduced activation of the frontal regions (primarily the bilateral IFG) in the T3 prime condition may suggest that the T3 primes incurred less effort in downstream articulatory planning and motor execution of the T3 sandhi during production, which was possibly due to the earlier facilitation in retrieving the underlying tonal category induced by the T3 primes.

In contrast, the T2 prime condition yielded stronger activation in the left temporal regions (STG and MTG) during subsequent T3 sandhi word production compared to the control prime condition, an opposite pattern to that of the T3 prime condition. We propose that the stronger activation of these regions following the presentation of T2 primes may reflect increased competition in the wordform or phonological code retrieval process. As shown by previous behavioral studies, native Chinese speakers exhibited more perceptual confusion between T2 and T3 compared to other tonal pairs (e.g., T1-T3 or T4-T3), relative to non-native speakers (A. Chen et al., 2016; Huang & Johnson, 2011). Similarly, EEG studies using the oddball paradigm (Chandrasekaran, Gandour, & Krishnan, 2007; Chandrasekaran, Krishnan, & Gandour, 2007; Li & Chen, 2015) revealed that compared to other tonal pairs (e.g., T1-T3 or T4-T3), the tonal contrast between the T2 and T3 elicited lower amplitudes or delayed latencies of mismatched negativity (MMN, an ERP component indexing pre-attentive processing of phonological contrasts), indicating less perceptual distinctiveness between T2 and T3. In

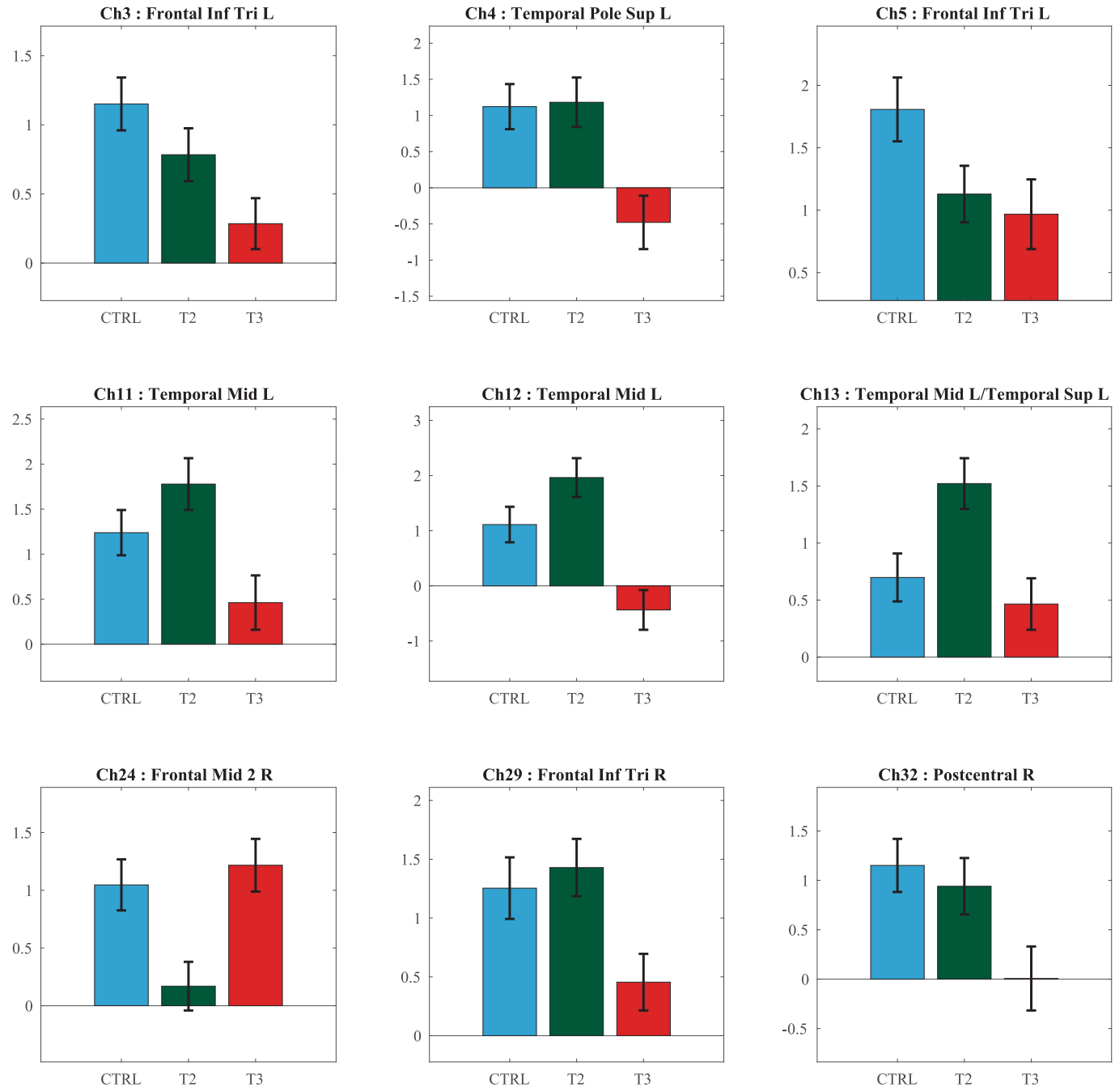

Fig. 5. Bar plots of beta weight values (averaged across the second to fifth time bins, i.e., 2–10 s after the picture onset) for the three prime conditions from those channels showing significant effects (see Fig. 4), with each channel labeled by the most representative brain regions based on the AAL atlas (see Table S1 in Supplementary materials). The error bars stand for the standard error of the mean. Ch: Channel; CTRL: Control primes; L: Left; R: Right; Sup: Superior; Mid: Middle; Inf: Inferior; Tri: Pars triangularis.

line with these findings, it is probable that the T2 primes may cause more phonological competition and stronger interference during the retrieval of the underlying T3 category than T1 or T4 primes. Moreover, given the high phonetic similarity between the sandhi T3 variant and the T2 (e.g., see the example of 百米 and 白米 in the introduction), the T2 primes would more strongly activate a cohort of both the disyllabic words beginning with the T2 and the T3 sandhi variant (see Meng et al., 2021). As indicated in previous priming studies (Dufour & Peereman, 2003; Radeau et al., 1989), primes with greater phonological similarity to the target word could activate more similar-sounding candidates, leading to stronger competition and greater interference in accessing the target word. Thus, the additional activation of disyllabic lexical candidates beginning with T2, which were more phonetically similar to the T3 sandhi words, may induce more interference to the access of the T3 sandhi words. In contrast, the T1 or T4 primes may more strongly activate disyllabic words beginning with T1 or T4, which bore less phonetic similarity to the T3 sandhi words and thus may induce less interference to the lexical retrieval process.[7] As shown by previous neuroimaging studies, stronger lexical-phonological competition may

[7] As suggested by an anonymous reviewer, future studies may also consider potential phonological neighborhood density differences, e.g., number of words carrying T2/T3 versus T1/T4, when explaining the T2 prime effect.

elicit more activation in the left temporal regions (STG and MTG) (Luthra et al., 2019; Minicucci et al., 2013; Okada & Hickok, 2006). In summary, the observed greater activation of the left temporal regions during production following the T2 prime presentation was probably due to higher lexical-phonological competition and larger interference induced by the T2 primes compared to the control primes (T1 or T4 primes).

Although T2 primes appeared to induce higher competition in the retrieval process, we found that the T2 prime condition yielded less activation in the left IFG during T3 sandhi word production, similar to what was observed in the T3 prime condition. As argued earlier, this may indicate that the T2 primes still facilitated downstream articulatory planning of the final surface tonal contour during T3 sandhi word production. Since T2 is highly similar to the surface tonal contour, it could help the speakers plan the final rising contour with ease at the phonetic encoding or the articulatory stage. This interpretation aligns with the EEG findings of X. Chen et al. (2022), which showed that the T2 prime primarily facilitated later motor planning before speech onset during the production of Mandarin T3 sandhi words. However, unlike the T3 prime condition, we did not observe reduced activation of the right IFG for the T2 primes compared to the control primes. One tentative explanation is that the T2 primes may induce higher lexical-phonological competition in the upstream retrieval process (as argued above) and thus require more cognitive control to resolve the competition. Some previous neuroimaging studies showed that the right IFG was often activated in tasks requiring inhibitory control (Aron et al., 2014; Cai et al., 2014; Hartwigsen et al., 2019; Levy & Wagner, 2011). Based on this finding, the right IFG may be recruited in resolving the intense lexical-phonological competition induced by the T2 primes, leading to null activation in this region for the T2 prime condition compared to the control prime condition.

Moreover, compared to the control prime condition, decreased activation was observed in the right MFG during production for the T2 prime condition but not for the T3 prime condition. The right MFG has been found to be involved in tonal perception (Gandour et al., 2004) and tonal production (Chang et al., 2014; Chang & Kuo, 2016). But according to a metalinguistic study into the involvement of the right hemispheric regions in language processing, it has been proposed that the right MFG is mainly associated with auditory selective attention (Vigneau et al., 2011). Based on this proposal, activations in this region possibly reflect attention to the discrepancy between the primes and the executed pitch contour of the T3 sandhi tone when producing the T3 sandhi words. Note that the pitch contour of the T2 primes is closest to the produced pitch contour of the T3 sandhi tone, whereas that of the control primes (T1/T4) is very different. Thus, the large discrepancy in pitch contour between the control primes and the T3 sandhi tone may boost the participants' awareness of the pitch contour difference and enhance their selective attention to the final pitch execution, which was reduced in the T2 prime condition. On the other hand, the pitch contour of the T3 primes was also different from that of the T3 sandhi tone, thus accounting for the null activation in this region when compared to the control primes.

Although we did not aim to test brain lateralization for tonal processing, our neural data indicate that regions from both hemispheres are likely involved in tone sandhi production. On the one hand, it has been indicated that pitch processing tends to engage the right hemispheric regions (Zatorre & Belin, 2001), which is evidenced by neuroimaging studies in non-speech pitch perception (e.g., Zatorre et al., 1992) or musical pitch production like singing (Brown et al., 2006; Özdemir et al., 2006). Similarly, previous fMRI studies on tonal perception (Gandour et al., 2004; Kwok et al., 2017, 2019; B. Liang & Du, 2018) and tonal production (Chang et al., 2014; Chang & Kuo, 2016; L. Liu et al., 2006) also showed the involvement of right hemispheric regions, aligning with our results. On the other hand, previous neurogaming studies on tonal perception (e.g., Gandour et al., 2004; B. Liang & Du, 2018) and tonal production (L. Liu et al., 2006) also revealed additional engagement of the left hemispheric regions in tonal processing for native tonal language speakers, which is also consistent with our current results. Moreover, our fNIRS data indicated that the retrieval of high-level lexical-phonological information (i.e., the abstract tonal category) appeared to be supported primarily by the left hemispheric regions (specifically, the temporal regions). In contrast, low-level articulatory planning and execution of the pitch contour seemed to involve bilateral activity in the frontal regions. This pattern is also consistent with previous neuroimaging studies on speech production, which showed that lexico-semantic and phonological wordform retrieval tend to engage left-lateralized cortical regions, whereas articulatory planning and motor execution tend to involve bilateral cortical regions (see reviews in Nozari, 2022).

However, the involvement of the right frontal regions observed in our study and previous studies on tonal production must be interpreted with caution. As mentioned earlier, the right frontal regions (e.g., right IFG and MFG) are also involved in domain-general executive control or attentional processes (Aron et al., 2014; Cai et al., 2014; Hartwigsen et al., 2019; Levy & Wagner, 2011; Vigneau et al., 2011). Thus, the (de) activation of the right frontal regions may alternatively reflect different degrees of cognitive control involved in different conditions rather than linguistic processes per se, an issue neglected in previous imaging studies in tonal production. For instance, L. Liu et al. (2006) found greater activation in the right IFG when participants produced Mandarin syllables with tonal overlap but no segmental overlap (tonal repetition), compared to when they produced Mandarin syllables with vowel overlap but no tonal overlap (vowel repetition). The authors interpreted this pattern as an indication of the important role of the right IFG in tonal production. However, a few previous studies (J. Chen et al., 2002; X. Chen & Zhang, 2025; Q. Zhang, 2008) have shown that greater interference may occur when producing words with tone-related but segmentally-unrelated primes, compared to unrelated primes, possibly due to a greater disruption in tone-to-syllable integration during production. Thus, the enhanced activation of the right IFG may simply be attributed to greater cognitive control in the tone repetition condition. Similarly, two other fMRI studies (Chang et al., 2014; Chang & Kuo, 2016) observed stronger activation of the right IFG when participants produced Mandarin syllabic sequences involving T3 sandhi than producing sequences without tone sandhi. This could also be alternatively explained by greater cognitive control engaged to suppress the dominant activation of the citation form of Mandarin T3 (e.g., triggered by the visual tone cue of T3 presented on the screen) in order to overtly execute the correct surface pitch contour during the production of sequences involving T3 sandhi. In our study, the phonologically primed task can be regarded as a variant of the picture-word interference task, and the distractors or the primes bearing different degrees of phonological overlap with the target words could induce different degrees of lexical-phonological competition during word planning (De Zubicaray et al., 2002). Thus, the involvement of the right frontal regions in this study may alternatively be attributed to various degrees of cognitive control induced by different types of primes. As such, the specific role of the right frontal regions in tonal production remains less clear and warrants further investigation in future research.

Unlike previous fMRI studies on Mandarin T3 sandhi (Chang et al., 2014; Chang & Kuo, 2016), we additionally observed certain left brain regions activated in the T3 sandhi word production other than the right IFG. This discrepancy may be attributed to different task paradigms and stimulus types between our study and previous fMRI studies. As discussed in the introduction, the priming technique used in this study allowed us to tap into the subtle sub-processes of T3 sandhi word production, in particular the early retrieval process. This may account for why we observed the involvement of the left temporal regions, which are thought to be mainly associated with the retrieval process. Moreover, our study used real T3 sandhi words as the stimuli, while the previous fMRI studies (Chang et al., 2014; Chang & Kuo, 2016) used non-word tonal sequences. This may also explain why we found the

additional involvement of left brain regions in the T3 sandhi word production (such as left IFG and left temporal regions, see Fig. 4), as tonal stimuli with more linguistic relevance may tend to more engage the left hemispheric regions (e.g., Zatorre & Gandour, 2008).

In our study, we did not observe activation differences in the IPL between different prime conditions. In some previous fMRI studies (Abel et al., 2009, 2012; Arrigoni et al., 2024; Diaz et al., 2014), the left IPL was found to be activated more strongly when participants named pictures with phonologically related distractors compared to phonologically unrelated distractors in the picture-word interference task. Note that the picture-word interference task resembles the phonologically primed task used in our study. The absence of the IPL involvement in our study may be because different types of phonological overlap between the primes and target words were manipulated between our study (focusing on tonal overlap) and those previous fMRI studies (focusing on segmental overlap).

However, the above discrepancy could also be due to some methodological issues when applying the fMRI technique in speech production. Note that those fMRI studies detecting the involvement of the IPL in the phonological facilitation effect employed the continuous imaging acquisition, with the simultaneous presence of gradient noises during speech production. As pointed out by previous research (De Zubicaray et al., 2002; De Zubicaray & Piai, 2019), the gradient noises could interfere with speech production and alter the brain activations revealed in the fMRI signals. Interestingly, our fNIRS results seemed to be more aligned with two earlier fMRI studies investigating the phonological facilitation effect in the picture-word interference task (De Zubicaray et al., 2002; De Zubicaray & Mcmahon, 2009), which also revealed reduced activation in the left temporal regions for the phonologically related condition. These two fMRI studies employed the sparse temporal sampling acquisition instead of continuous imaging, which acquired only a single fMRI volume imaging at a delayed interval after the picture onset to capture the peak hemodynamic changes. Using this acquisition method, participants could produce speech at the relative silent interval without the gradient noises, showing more resemblance to the situation as in our fNIRS study. Moreover, the modality of the distractors/primes may also lead to different activation patterns. Using auditory distractors, De Zubicaray & Mcmahon (2009) additionally observed reduced activities in the left IFG for the phonologically related condition. This pattern is similar to what we found in our study, which also used auditory primes. These observations underline the impact of choices of task design, neuroimaging techniques (fMRI and fNIRS), and the presence of noise on functional activation results, which calls for further research.

In sum, our fNIRS data supplemented the EEG and behavioral findings from X. Chen et al. (2022) and revealed distinct influences between the two types of primes on specific brain regions. These findings have implications for understanding the neural bases of phonological alternations in speech production, which we discuss below.

4.2. Theoretical implications

The derivation of phonological alternations has been a central topic in the field of phonology (Arndt-lappe and Ernestus, 2020) but unfortunately received less attention in current research on speech production (Bürki, 2018). As such, the neural substrates of phonological alternations remain poorly understood in the current neurocognitive model of speech production (Indefrey, 2011; Indefrey & Levelt, 2004). As indicated by our current results, for the phonological alternations like Mandarin T3 sandhi, the retrieval of the abstract phonological form shared for different variants may be subserved by the left temporal regions. In contrast, the articulatory planning and execution of the context-specific phonological/phonetic variants may be subserved by the frontal regions. This is consistent with the results of Sahin et al. (2009), which observed that the encoding of the phonological alternations for the English plural morpheme "-s" was reflected in the Broca's areas (left IFG) during production. However, given the existence of various types of phonological alternations (Bürki, 2018; Bürki & Gaskell, 2012; Scheer & Mathy, 2021), the neurocognitive mechanisms of other types of phonological alternations still require further investigation.

Regarding the debate over the processing mechanisms for Mandarin T3 sandhi production, we believe that the current data are insufficient to fully resolve the issue. Our behavioral results, together with findings from other previous studies (X. Chen et al., 2022; Nixon et al., 2015), suggested that both the underlying tonal category and surface tonal contour information were activated in the real-time planning process of Mandarin T3 sandhi word production. This ran counter to the lexical storage account (Hsieh, 1970, 1976), which only assumed the storage of the surface tonal contour in the lexicon and predicted no role of the underlying tonal category in word production. Instead, our behavioral data seemed to be more consistent with the multivariant activation account (Y. Chen et al., 2011; Nixon et al., 2015). However, our data can also be deemed to be consistent with the computation account, which also assumes that both the underlying tonal category and the surface tonal contour information are activated in production (though they may be activated in a serial manner). Thus, the current evidence remains equivocal with regard to the processing mechanism of real-time T3 sandhi words. As our neural data showed that both frontal and temporal regions were involved in the real-time planning process of Mandarin T3 sandhi word production, it is possible that the production of T3 sandhi words involves a combination of online computation and lexical retrieval mechanisms, rather than relying exclusively on one or the other. To inform the debate over the processing mechanisms for Mandarin T3 sandhi production, future research may need to further compare the neural activities during the production of T3 sandhi words with that of well-matched non-sandhi words (e.g., C. Zhang et al., 2015; J. Zhang et al., 2022) using naturalistic picture or word naming tasks.

Although tone sandhi is a phonological phenomenon specific to the Chinese tonal languages, our detected brain regions underlying the T3 sandhi word production could be largely explained by the current neurocognitive models of speech production (Hickok, 2012; Indefrey, 2011; Indefrey & Levelt, 2004). These production models, albeit developed based on evidence from non-tonal Indo-European languages, highlight distinct functional roles of frontal and temporal brain regions in speech production processes, which align closely with our current findings. It is possible that neural networks supporting speech production may overlap across languages, while some specificity for certain languages or types of speech sounds also exists. Indeed, some previous studies did suggest that lexical tonal processing may engage distinct brain sub-regions from syllabic or segmental processing (e.g., Feng et al., 2018; J. Yu et al., 2023). As our study did not directly compare the encoding of syllable/segments with tonal encoding, it remains unclear whether lexical tonal encoding in tonal languages recruits distinct brain sub-regions during real-time word production from syllabic or segmental encoding, which requires further research.

4.3. Limitations and future work

Together with previous fNIRS studies on speech production (Hitomi et al., 2021; Moriai-Izawa et al., 2012), our study demonstrated the value of the fNIRS technique in investigating overt speech production. In particular, our study showed that fNIRS can be successfully applied using the rapid event-related design to reveal the neural substrates underlying specific linguistic processes, with the application of proper data preprocessing methods (Aarabi et al., 2017; Huppert, 2016) and the use of short separation channels to remove the confounding systemic physiological artifacts (Santosa et al., 2020). That said, our study has certain limitations, given the inherent constraints of the fNIRS technique. First, the spatial resolution of fNIRS is not as high as that of fMRI, so it can only provide rough estimations of the brain regions for neural processing. More precise mapping to the underlying anatomical regions using fNIRS also requires the use of the 3D digitizer or further acquisition of the individual structural images using MRI, which was lacking in

our study. Second, previous research on speech production showed that the inner cortical regions (e.g., cingulate cortex and insula) and subcortical regions (e.g., basal ganglia) were also involved in speech production (De Zubicaray, 2023; Nozari, 2022), but our study was unable to probe the neural activities from these regions since fNIRS could only detect the neural activities from the surface cortical areas. Third, due to the poor temporal resolution of fNIRS, it was difficult to precisely determine which regions were active during each specific processing stage of production. Moreover, as pointed out by one anonymous reviewer, the hemodynamic changes were typically slow but the temporal interval between the prime and the target word in our study was short. This rendered it difficult to tease apart the neural activities of the prime processing from those of the picture naming per se in the current task paradigm. This is also an inherent problem in previous fMRI or fNIRS studies employing the picture-word interference naming paradigm (Abel et al., 2009, 2012; Diaz et al., 2014; De Zubicaray et al., 2002; De Zubicaray & Mcmahon, 2009; Hitomi et al., 2021), where the presentation of distractors showed temporal overlap with the picture presentation. Thus, future research employing techniques with both high temporal and spatial resolution, such as magnetoencephalography (MEG) or intracranial EEG, will better reveal a complete picture of the spatiotemporal dynamics underlying Mandarin T3 sandhi production process when using the priming or interference naming paradigm.

Future research could also further extend our current study in several ways. First, as shown by recent studies on the recognition of T3 sandhi words (Gao et al., 2021; Gao & Lin, 2025), the morphological structure and morpheme frequency may influence the phonological representation of the Mandarin T3 sandhi words, an issue requiring further exploration in production. Moreover, the application of Mandarin T3 sandhi is not only confined to disyllabic words, and the exertion of the T3 sandhi rule on sequences with three T3 syllables or more or on consecutive T3s across the word boundary is more complicated and subject to the influence of morphosyntactic structure and prosodic factors (e.g., speaking rate or pause) (M. Chen, 2000; Lai & Li, 2022; C. Liu & Chen, 2020; Tang, Yuen, et al., 2019;- Shih, 1986). But the underlying neural mechanisms for the T3 sandhi application beyond disyllabic words require future investigation. Lastly, there are different types of tonal alternations occurring in many tonal languages/dialects and recent studies on word recognition show that tonal alternations in other Chinese dialects may involve different processing mechanisms from the Mandarin T3 sandhi (Chang et al., 2019; X. Chen & Zhang, 2023; Chien et al., 2017; Lau & Wong, 2024; Yan et al., 2020, 2021), which also requires further exploration.

5. Conclusion

In conclusion, this study revealed facilitation effects of both T3 and T2 primes on naming onset latencies of Mandarin T3 sandhi words, suggesting activation of both the underlying tonal category and the context-specific tonal contour information in production. The fNIRS data indicated that the left temporal regions and the bilateral frontal regions were involved in supporting the T3 sandhi word production, with T3 and T2 primes inducing distinct neural activations compared to control primes. Further research should further elucidate the specific roles of different brain regions in Mandarin T3 sandhi production.

CRediT authorship contribution statement

Xiaocong Chen: Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Tai Yuan:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation. **Yiya Chen:** Writing – review & editing, Resources, Methodology, Conceptualization. **Fumo Huang:** Writing – review & editing, Data curation. **Caicai Zhang:** Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The first two authors contributed to this paper equally. This work was supported in part by the Postdoctoral Fellowships Scheme at the Department of Chinese and Bilingual Studies (CBS) of the Hong Kong Polytechnic University (HK PolyU) awarded to CXC (P0043384) and the HK PolyU CBS Departmental General Research Fund (P0046373) awarded to CCZ. The facilities used in this study were supported by the University Research Facility in Behavioral and Systems Neuroscience (UBSN) at the Hong Kong Polytechnic University. We thank Mr. Shixuan Wang for recording the audio materials, and Dr. Zebo Xu and Dr. Tommy L.H. Lam for technical assistance. The authors declare no conflict of interest.

Declaration of generative AI in the writing process.

When preparing the manuscript, the first author used the AI agents available at poe.com (mostly using GPT-4o, GPT4.5 and Clause 3.7) to rewrite some sentences and certain paragraphs in order to improve the clarity and readability of the texts. The authors reviewed and edited all the AI-generated content as needed and take full responsibility for the content of the published article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandl.2025.105636.

Data availability

Materials, data and codes are available at https://osf.io/e2pka/.

References

Aarabi, A., Osharina, V., & Wallois, F. (2017). Effect of confounding variables on hemodynamic response function estimation using averaging and deconvolution analysis: An event-related NIRS study. *NeuroImage, 155*, 25–49.

Abel, S., Dressel, K., Bitzer, R., Kümmerer, D., Mader, I., Weiller, C., & Huber, W. (2009). The separation of processing stages in a lexical interference fMRI-paradigm. *NeuroImage, 44*(3), 1113–1124.

Abel, S., Dressel, K., Weiller, C., & Huber, W. (2012). Enhancement and suppression in a lexical interference fMRI-paradigm. *Brain and Behavior, 2*(2), 109–127.

Arndt-lappe, S., & Ernestus, M. (2020). Morpho-phonological alternations: The role of lexical storage. In V. Pirrelli, I. Plag, & W. U. Dressler (Eds.), *Word knowledge and word usage* (pp. 192–227). De Gruyter.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: One decade on. *Trends in Cognitive Sciences, 18*(4), 177–185.

Arrigoni, E., Rappo, E., Papagno, C., Romero Lauro, L. J., & Pisoni, A. (2024). Neural Correlates of semantic interference and phonological facilitation in picture naming: A systematic review and coordinate-based Meta-analysis. *Neuropsychology Review, 35*, 35–53.

Audacity Team. (2017). Audacity (R): Free audio editor and recorder (2.3.3). Retrieved from https: //audacityteam.org/.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Boersma, P., & Weenink, D. (2022). Praat. *Retrieved from.* http://www.fon.hum.uva.nl/praat/download_win.html.

Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *NeuroImage, 32*(2), 821–841.

Brigadoi, S., Ceccherini, L., Cutini, S., Scarpa, F., Scatturin, P., Selb, J., Gagnon, L., Boas, D. A., & Cooper, R. J. (2014). Motion artifacts in functional near-infrared spectroscopy: A comparison of motion correction techniques applied to real cognitive data. *NeuroImage, 85*, 181–191.

Brown, S., Martinez, M. J., & Parsons, L. M. (2006). Music and language side by side in the brain: A PET study of the generation of melodies and sentences. *European Journal of Neuroscience, 23*(10), 2791–2803.

Bürki, A. (2018). Variation in the speech signal as a window into the cognitive architecture of language production. *Psychonomic Bulletin and Review, 25*(6), 1973–2004.
Bürki, A., & Gaskell, M. G. (2012). Lexical representation of schwa words: Two mackerels, but only one salami. *Journal of Experimental Psychology: Learning Memory and Cognition, 38*(3), 617–631.
Cai, W., Ryali, S., Chen, T., Li, C. S. R., & Menon, V. (2014). Dissociable roles of right inferior frontal cortex and anterior insula in inhibitory control: Evidence from intrinsic and task-related functional parcellation, connectivity, and response profile analyses across multiple datasets. *Journal of Neuroscience, 34*(44), 14652–14667.
Chandrasekaran, B., Gandour, J. T., & Krishnan, A. (2007). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity Bharath. *Restorative Neurology and Neuroscience, 25*(3–4), 195–210.
Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research, 1128*(1), 148–156.
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 One1, 11*(7), Article e0159835.
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 H. Liu, F. Tsao, & P. Li (Eds.), *Speech perception, production and acquisition: Multidisciplinary approaches in Chinese languages* (pp. 117–135). Springer Singapore.
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 One1, 9*(1), Article e83126.
Chang, C. H. C., Lin, T.-H., & Kuo, W.-J. (2019). Does phonological rule of tone substitution modulate mismatch negativity? *Journal of Neurolinguistics, 51*, 63–75.
Chao, Y.-R. (1948). *Mandarin primer: An intensive course in spoken chinese*. Harvard University Press.
Chen, A., Liu, L., & Kager, R. (2016). Cross-domain correlation in pitch perception, the influence of native language. *Language, Cognition and Neuroscience, 31*(6), 751–760.
Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word-form encoding in Mandarin chinese as assessed by the implicit priming task. *Journal of Memory and Language, 46*(4), 751–781.
Chen, M. Y. (2000). *Tone sandhi: Patterns across chinese dialects*. Cambridge University Press.
Chen, X., & Zhang, C. (2023). Encoding of contextual sandhi tonal variants in Chaoshan Min: An implicit priming study. In G. Peng, J. Kong, Z. Shen, & F. Wang (Eds.), *Inspirations from a Lofty Mountain— Festschrift in Honor of Professor William S-Y. Wang on his 90th Birthday* (pp. 216–242). Hong Kong: City: University of Hong Kong Press.
Chen, X., & Zhang, C. (2025). Setting the "tone" first and then integrating it into the syllable: An EEG investigation of the time course of lexical tone and syllable encoding in Mandarin word production. *Journal of Memory and Language, 140*, Article 104575.
Chen, X., Zhang, C., Chen, Y., Politzer-Ahles, S., Zeng, Y., & Zhang, J. (2022). Encoding category-level and context-specific phonological information at different stages: An EEG study of Mandarin third-tone sandhi word production. *Neuropsychologia, 175*, Article 108367.
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).
Chien, Y.-F., Sereno, J. A., & Zhang, J. (2016). Priming the representation of Mandarin tone 3 sandhi words. *Language, Cognition and Neuroscience, 31*(2), 179–189.
Chien, Y.-F., Sereno, J. A., & Zhang, J. (2017). What's in a word: Observing the contribution of underlying and surface representations. *Language and Speech, 60*(4), 643–657.
Chien, Y.-F., Yan, H., & Sereno, J. A. (2021). Investigating the lexical representation of Mandarin tone 3 phonological alternations. *Journal of Psycholinguistic Research, 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. *Frontiers in Psychology, 11*, 646.
Chu, Q., Ma, O., Hang, Y., & Tian, X. (2023). Dual-stream cortical pathways mediate sensory prediction. *Cerebral Cortex, 33*(14), 8890–8903.
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. (2023). The neural organisation of language production: Evidence from neuroimaging and neuromodulation. In R. Hartsuiker, & K. Strijkers (Eds.), *Current issues in the Psychology of Language*. Routledge.
De Zubicaray, G. I., & McMahon, K. L. (2009). Auditory context effects in picture naming investigated with event-related fMRI. *Cognitive, Affective and Behavioral Neuroscience, 9*(3), 260–269.
De Zubicaray, G. I., McMahon, K. L., Eastburn, M. M., & Wilson, S. J. (2002). Orthographic/phonological facilitation of naming responses in the picture-word task: An event-related fMRI study using overt vocal responding. *NeuroImage, 16*(4), 1084–1093.
De Zubicaray, G. I., & Piai, V. (2019). Investigating the Spatial and Temporal Components of Speech Production. In G. I. De Zubicaray, & N. O. Schiller (Eds.), *The Oxford handbook of neurolinguistics* (pp. 471–497). Oxford University Press.
Dell, G. S., Schwartz, M. F., Nozari, N., Faseyitan, O., & Branch Coslett, H. (2013). Voxel-based lesion-parameter mapping: Identifying the neural correlates of a computational model of word production. *Cognition, 128*(3), 380–396.
Deng, Y., Feng, L., & Peng, D. (2003). Research on articulatory rule of third tone sandhi of Standard chinese in different contexts. *Acta Psychologica Sinica, 35*(6), 719–725.
Diaz, M. T., Hogstrom, L. J., Zhuang, J., Voyvodic, J. T., Johnson, M. A., & Camblin, C. C. (2014). Written distractor words influence brain activity during overt picture naming. *Frontiers in Human Neuroscience, 8*, 167.
Dufour, S., & Peereman, R. (2003). Lexical competition in phonological priming: Assessing the role of phonological match and mismatch lengths between primes and targets. *Memory & Cognition, 31*(8), 1271–1283.
Duraivel, S., Rahimpour, S., Barth, K., Chiang, C., Wang, C., Harward, S. C., Lad, S. P., Sexton, D. P., Friedman, A. H., & Derek, G. (2024). *Neural mechanisms of the transition from planning to execution in speech production.* bioRxiv, 2024.10.07.617122.
Feng, G., Gan, Z., Llanos, F., Meng, D., Wang, S., Wong, P. C., & Chandrasekaran, B. (2021). A distributed dynamic brain network mediates linguistic tone representation and categorization. *Neuroimage, 224*, Article 117410.
Feng, G., Gan, Z., Wang, S., Wong, P. C., & Chandrasekaran, B. (2018). Task-general and acoustic-invariant neural representation of speech categories in the human brain. *Cerebral cortex, 28*(9), 3241–3254.
Ferrari, M., & Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *NeuroImage, 63*(2), 921–935.
Gandour, J., Potisuk, S., Ponglorpisit, S., Dechongkit, S., Khunadorn, F., & Boongird, P. (1996). Tonal coarticulation in Thai after unilateral brain damage. *Brain and Language, 52*(3), 505–535.
Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X., & Lowe, M. (2004). Hemispheric roles in the perception of speech prosody. *NeuroImage, 23*(1), 344–357.
Gao, F., & Lin, C. J. C. (2025). Incorporating Frequency Effects in the Lexical Access of Mandarin Tone 3 Sandhi. *Language and Speech, 68*(1), 204–228.
Gao, F., Lyu, S., & Lin, C. J. C. (2021). Processing Mandarin Tone 3 sandhi at the morphosyntactic interface: Reduplication and lexical compounds. *Frontiers in Psychology, 12*(7), Article 713665.
Ghosh, S. S., Tourville, J. A., & Guenther, F. H. (2008). A neuroimaging study of premotor lateralization and syllables. *Journal of Speech Language Hearing Research, 51*(5), 1183–1202.
Gow, D. W. (2012). The cortical organization of lexical knowledge: A dual lexicon model of spoken language processing. *Brain and Language, 121*(3), 273–288.
Guo, Z., & Chen, F. (2022). Decoding lexical tones and vowels in imagined tonal monosyllables using fNIRS signals. *Journal of Neural Engineering, 19*(6), Article 066007.
Hartwigsen, G., Neef, N. E., Camilleri, J. A., Margulies, D. S., & Eickhoff, S. B. (2019). Functional segregation of the right inferior frontal gyrus: Evidence from coactivation-based parcellation. *Cerebral Cortex, 29*(4), 1532–1546.
Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience, 13*(2), 135–145.
Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393–402.
Hickok, G., Venezia, J., & Teghipco, A. (2023). Beyond Broca: Neural architecture and evolution of a dual motor speech coordination system. *Brain, 146*(5), 1775–1790.
Hitomi, T., Gerrits, R., & Hartsuiker, R. J. (2021). Using functional near-infrared spectroscopy to study word production in the brain: A picture-word interference study. *Journal of Neurolinguistics, 57*, Article 100957.
Hsieh, H.-I. (1970). The psychological reality of tone sandhi rules in Taiwanese. In *In Papers from the 6th Annual Regional Meeting of the Chicago Linguistic Society* (pp. 489–503).
Hsieh, H.-I. (1976). On the unreality of some phonological rules. *Lingua, 38*(1), 1–19.
Huang, T., & Johnson, K. (2011). Language specificity in speech perception: Perception of mandarin tones by native and nonnative listeners. *Phonetica, 67*(4), 243–267.
Huppert, T. J. (2016). Commentary on the statistical properties of noise and its implication on general linear models in functional near-infrared spectroscopy. *Neurophotonics, 3*(1), Article 010401.
Huppert, T. J., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2009). HomER: A review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied Optics, 48*(10), D280–D298.
Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology, 2*, 255.
Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1–2), 101–144.
Joanisse, M. F., Zevin, J. D., & McCandliss, B. D. (2007). Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using fMRI and a short-interval habituation trial paradigm. *Cerebral Cortex, 17*(9), 2084–2093.
Khanna, A. R., Muñoz, W., Kim, Y. J., Kfir, Y., Paulk, A. C., Jamali, M., & Williams, Z. M. (2024). Single-neuronal elements of speech production in humans. *Nature, 626*(7999), 603–610.
Kuznetsova, A., Brockhof, P. B., & Christensen, R. H. B. (2017). lmerTest package : Tests in linear mixed effects. *Journal of Statistical Software, 82*(13), 1–26.
Kwok, V. P. Y., Dan, G., Yakpo, K., Matthews, S., Fox, P. T., Li, P., & Tan, L. H. (2017). A meta-analytic study of the neural systems for auditory processing of lexical tones. *Frontiers in Human Neuroscience, 11*, 375.
Kwok, V. P. Y., Matthews, S., Yakpo, K., & Tan, L. H. (2019). Neural correlates and functional connectivity of lexical tone processing in reading. *Brain and Language, 196*, Article 104662.
Lai, W., & Li, A. (2022). Integrating phonological and phonetic aspects of Mandarin Tone 3 sandhi in auditory sentence disambiguation. *Laboratory Phonology, 13*(1), 8.
Lau, J. C. Y., & Wong, P. C. M. (2024). Neural processing of Mandarin and Cantonese lexical tone alternations: Implications for the nature of phonological representations. *Laboratory Phonology, 15*(1), 1–46.

Levy, B. J., & Wagner, A. D. (2011). Cognitive control and right ventrolateral prefrontal cortex: Reflexive reorienting, motor inhibition, and action updating. *Annals of the New York Academy of Sciences, 1224*(1), 40–62.
Li, X., & Chen, Y. (2015). Representation and processing of lexical tone and tonal variants: Evidence from the mismatch negativity. *PLoS ONE, 10*(12), Article e0143097.
Liang, B., & Du, Y. (2018). The functional neuroanatomy of lexical tone perception: An activation likelihood estimation meta-analysis. *Frontiers in Neuroscience, 12*, 495.
Liang, J., & Van Heuven, V. J. (2004). Evidence for separate tonal and segmental tiers in the lexical specification of words: A case study of a brain-damaged chinese speaker. *Brain and Language, 91*(3), 282–293.
Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex, 15*(10), 1621–1631.
Liebenthal, E., Desai, R., Ellingson, M. M., Ramachandran, B., Desai, A., & Binder, J. R. (2010). Specialization along the left superior temporal sulcus for auditory categorization. *Cerebral Cortex, 20*(12), 2958–2970.
Liu, C. T., & Chen, L. M. (2020). Testing the applicability of third tone sandhi at the intonation boundary: The case of the monosyllabic topic. *Language and Linguistics, 21*(4), 636–651.
Liu, L., Peng, D., Ding, G., Jin, Z., Zhang, L., Li, K., & Chen, C. (2006). Dissociation in the neural basis underlying chinese tone and vowel production. *NeuroImage, 29*(2), 515–523.
Liu, Y., Zhao, Z., Xu, M., Yu, H., Zhu, Y., Zhang, J., Bu, L., Zhang, X., Lu, J., Li, Y., Ming, D., & Wu, J. (2023). Decoding and synthesizing tonal language speech from brain activity. *Science Advances, 9*(23), Article eadh0478.
Lu, J., Li, Y., Zhao, Z., Liu, Y., Zhu, Y., Mao, Y., Wu, J., & Chang, E. F. (2023). Neural control of lexical tone production in human laryngeal motor cortex. *Nature Communications, 14*(1), 6917.
Luthra, S., Guediche, S., Blumstein, S. E., & Myers, E. B. (2019). Neural substrates of subphonemic variation and lexical competition in spoken word recognition. *Language, Cognition and Neuroscience, 34*(2), 151–169.
Mantegna, F., Orpella, J., & Poeppel, D. (2025). Time-resolved hemispheric lateralization of audiomotor functional connectivity during covert speech production. *Cell Reports, 44*(1), Article 115137.
Meng, Y., Wynne, H., & Lahiri, A. (2021). Representation of "T3 sandhi" in mandarin: Significance of context. *Language, Cognition and Neuroscience, 36*(6), 791–808.
Minicucci, D., Guediche, S., & Blumstein, S. E. (2013). An fMRI examination of the effects of acoustic-phonetic and lexical competition on access to the lexical-semantic network. *Neuropsychologia, 51*(10), 1980–1988.
Molavi, B., & Dumont, G. A. (2012). Wavelet-based motion artifact removal for functional near-infrared spectroscopy. *Physiological Measurement, 33*(2), 259–270.
Moriai-Izawa, A., Dan, H., Dan, I., Sano, T., Oguro, K., Yokota, H., Tsuzuki, D., & Watanabe, E. (2012). Multichannel fNIRS assessment of overt and covert confrontation naming. *Brain and Language, 121*(3), 185–193.
Nixon, J. S., Chen, Y., & Schiller, N. O. (2015). Multi-level processing of phonetic variants in speech production and visual word processing: Evidence from Mandarin lexical tones. *Language, Cognition and Neuroscience, 30*(5), 491–505.
Nozari, N. (2022). Neural basis of word production. In A. Papafragou, J. C. Trueswell, & L. R. Gleitman (Eds.), *The Oxford handbook of the mental lexicon* (pp. 536–560). Oxford University Press.
Okada, K., & Hickok, G. (2006). Identification of lexical-phonological networks in the superior temporal sulcus using functional magnetic resonance imaging. *NeuroReport, 17*(12), 1293–1296.
Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113.
Özdemir, E., Norton, A., & Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *NeuroImage, 33*(2), 628–635.
Papoutsi, M., De Zwart, J. A., Jansma, J. M., Pickering, M. J., Bednar, J. A., & Horwitz, B. (2009). From phonemes to articulatory codes: An fMRI study of the role of broca's area in speech production. *Cerebral Cortex, 19*(9), 2156–2165.
Peng, S. (2000). Lexical versus phonological representations of Mandarin sandhi tones. In *Language acquisition and the lexicon: Papers in laboratory phonology V* (pp. 152–167). Cambridge University Press.
Politzer-Ahles, S., & Zhang, J. (in press). Evidence for the role of tone sandhi in Mandarin speech production. *Journal of Chinese Linguistics (Series No. 25): Studies on tonal aspects of languages*.
R Core Team. *R: A language and environment for statistical computing*. (3.6.3). R Foundation for Statistical Computing. https://www.r-project.org/.
Radeau, M., Morais, J., & Dewier, A. (1989). Phonological priming in spoken word recognition: Task effects. *Memory & Cognition, 17*(5), 525–535.
Rong, F., Lisette Isenberg, A., Sun, E., & Hickok, G. (2018). The neuroanatomy of speech sequencing at the syllable level. *PLoS ONE, 13*(10), Article e0196381.
Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D., & Halgren, E. (2009). Phonological information within Broca' s area. *Science, 326*(5951), 445–449.
Santosa, H., Zhai, X., Fishburn, F., & Huppert, T. (2018). *The NIRS brain AnalyzIR toolbox. Algorithms, 11*(5), 73.
Santosa, H., Zhai, X., Fishburn, F., Sparto, P. J., & Huppert, T. J. (2020). Quantitative comparison of correction techniques for removing systemic physiological signal in functional near-infrared spectroscopy studies. *Neurophotonics, 7*(03), Article 035009.
Scheer, T., & Mathy, F. (2021). Neglected factors bearing on reaction time in language production. *Cognitive Science, 45*(10), 13050.
Segawaa, J. A., Tourvillea, J. A., Beala, D. S., & Guenther, F. H. (2015). The neural correlates of speech motor sequence learning. *Journal of Cognitive Neuroscience, 27*(4), 819–831.
Shih, C.-L. (1986). *The prosodic domain of tone sandhi in chinese*. San Diego, San Diego, US: University of California. Unpublished PhD dissertation.
Si, X., Zhou, W., & Hong, B. (2017). Cooperative cortical network for categorical processing of chinese lexical tone. *Proceedings of the National Academy of Sciences of the United States of America, 114*(46), 12303–12308.
Spunt, R. P. (2016). Easy-Optimize-X. *Retrieved from*. https://github.com/spunt/easy-optimize-x.
Tang, P., Xu Rattanasone, N, Yuen, I., Gao, L., & Demuth, K. (2019). The development of abstract representations of tone sandhi. *Developmental Psychology, 55*(10), 2114–2122.
Tang, P., Yuen, I., Xu Rattanasone, N., Gao, L., & Demuth, K. (2019). The acquisition of phonological alternations: The case of the Mandarin tone sandhi process. *Applied Psycholinguistics, 40*(6), 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* (pp. 395–405).
Van Esch, D. (2012). Leiden Weibo Corpus. Retrieved from https://lwc.daanvanesch.nl/.
Vigneau, M., Beaucousin, V., Hervé, P. Y., Jobard, G., Petit, L., Crivello, F., Mellet, E., Zago, L., Mazoyer, B., & Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? *Insights from a meta-analysis. NeuroImage, 54*(1), 577–593.
Wang, W. S. Y., & Li, K.-P. (1967). Tone 3 in Pekinese. *Journal of Speech and Hearing Research, 10*(3), 629–636.
Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics, 25*(1), 61–83.
Yan, H., Chien, Y.-F., & Zhang, J. (2020). Priming the representation of left-dominant sandhi words: A Shanghai dialect case study. *Language and Speech, 63*(2), 362–380.
Yan, H., Chien, Y.-F., & Zhang, J. (2021). The representation of variable tone sandhi patterns in Shanghai Wu. *Laboratory Phonology, 12*(1), 264.
Ye, J. C., Tak, S., Jang, K. E., Jung, J., & Jang, J. (2009). NIRS-SPM: Statistical parametric mapping for near-infrared spectroscopy. *NeuroImage, 44*(2), 428–447.
Yu, J., Zou, Y., & Wu, Y. (2023). The neural mechanisms underlying the processing of consonant, vowel and tone during chinese typing: An fNIRS study. *Frontiers in Neuroscience, 17*, Article 1258480.
Yu, K., Wang, R., & Li, P. (2020). Native and nonnative processing of acoustic and phonological information of lexical tones in chinese: Behavioral and neural correlates. In H. Liu, F. Tsao, & P. Li (Eds.), *Speech perception, production and acquisition: Multidisciplinary approaches in Chinese languages* (pp. 79–99). Springer Nature.
Yuan, J., & Chen, Y. (2014). 3rd tone sandhi in Standard chinese: A corpus approach. *Journal of Chinese Linguistics, 42*(1), 218–236.
Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11*, 946–953.
Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of Phonetic and Pitch Discrimination in Speech Processing. *Science, 256*(5058), 846–849.
Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1493), 1087–1104.
Zhang, C., & Peng, G. (2013). In *Productivity of Mandarin Third Tone sandhi: A wug test. In Eastward Flows the Great River: Festschrift in Honor of Prof. William S-Y* (pp. 256–282). City University of Hong Kong Press.
Zhang, C., Xia, Q., & Peng, G. (2015). Mandarin third tone sandhi requires more effortful phonological encoding in speech production: Evidence from an ERP study. *Journal of Neurolinguistics, 33*, 149–162.
Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology, 27*(1), 153–201.
Zhang, J., Zhang, C., Politzer-Ahles, S., Pan, Z., & Huang, X. (2022). The neural encoding of productive phonological alternation in speech production: Evidence from Mandarin Tone 3 sandhi. *Journal of Neurolinguistics, 62*, Article 101060.
Zhang, L., Xi, J., Xu, G., Shu, H., Wang, X., & Li, P. (2011). Cortical dynamics of acoustic and phonological processing in speech perception. *PLoS ONE, 6*(6), Article e20963.
Zhang, Q. (2008). Phonological encoding in monosyllabic and bisyllabic Mandarin word production: Implicit priming paradigm study. *Acta Psychologica Sinica, 40*(03), 253–262.
Zhang, W., Yang, F., & Tian, X. (2023). Functional connectivity between parietal and temporal lobes mediates internal forward models during speech production. *Brain and Language, 240*, Article 105266.
Zimeo Morais, G. A., Balardin, J. B., & Sato, J. R. (2018). fNIRS Optodes' Location Decider (fOLD): A toolbox for probe arrangement guided by brain regions-of-interest. *Scientific Reports, 8*(1), 3341.