

Neural processing of Mandarin and Cantonese lexical tone alternations: implications for the nature of phonological representations

Joseph C. Y. Lau, Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong S.A.R., China; Brain and Mind Institute, The Chinese University of Hong Kong, Hong Kong S.A.R., China; Roxelyn and Richard Pepper Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL, U.S.A, josephcyau@northwestern.edu

Patrick C. M. Wong*, Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong S.A.R., China; Brain and Mind Institute, The Chinese University of Hong Kong, Hong Kong S.A.R., China, p.wong@cuhk.edu.hk

*Corresponding author

This study provides neurocognitive evidence to shed further insight into the architecture of phonological representations. We tap into these representations by focusing on the neural processing of phonological alternations. Traditional generative accounts postulate abstract representations which are transformed by phonological processes into different alternants. Recent psycholinguistically-based frameworks posit that different alternants are listed in parallel in the abstract representation. Exemplar-based theories posit that alternation is a manifestation of frequency effects among concrete phonetic forms stored in memory. However, it is unclear the extent to which phonological representations of different levels of abstractedness are neurocognitive realities, beyond formal entities for linguistic analysis. This study tests the hypothesis that phonological representations are neurocognitive primitives that modulate speech processing and lexical access. We examined the processing of two surface-similar, but distributionally distinct lexical tone alternation patterns in Mandarin and Cantonese. With a cross-modal priming paradigm, differential neurophysiological components (LPC vs. N400) associated with the processing of alternation violations were identified cross-linguistically. Results support our hypothesis, suggesting that cross-linguistically distinct abstract phonological representations differentially modulate phonotactic detection, lexical access, and phonological restructuring processes. Results also lend theoretical insight into the multiform nature of the phonological representation, which encompasses both abstract and concrete levels of representations.



1. Introduction

Mental representations, as internal cognitive symbols that represent abstract core properties of sensory signals, are a fundamental building block of human cognition (Thagard, 2005). Mental representations permit us to learn from prior experience and to infer meaning from novel sensory signals (G. A. Miller, 1990), thus enabling intricate cognitive capacities such as learning, memory, and communication (Fodor & Pylyshyn, 1988; Jackendoff, 1995; Perszyk & Waxman, 2018). Speech sounds, as an essential type of sensory input for hearing individuals, are no exception. Decades of linguistic research have focused on theorizing the architecture of mental representations of speech sounds, i.e., phonological representations (J. Anderson & Jones, 1974; Jones, 1957; Pierrehumbert, 2016). Based on theories on the abstract symbolic systems of how speech sounds and their systematic variations are represented as phonological representations in memory (i.e., the lexicon) (McQueen et al., 2006), linguists were able to advance our understanding of speech and language as a hallmark of human communication. They do so by examining how phonological representations guide the perception of invariant meaning from highly variable acoustic signals (Kuhl, 1991; Pierrehumbert, 2016), the generation of novel but patterned sound sequences in language production (Chomsky & Halle, 1968; Prince & Smolensky, 2004), and the acquisition of language in infancy (Kuhl et al., 2008). The present study aims to advance current understandings of the architecture of phonological representations, by examining how neural speech processing is modulated by structures of phonological representations that vary cross-linguistically.

Ever since the earliest theories of the phoneme (Trubetzkoy, 1939), phonological representations have been conceptualized not only as a shorthand for formal analysis, but also as neurocognitively real computational primitives that are actively involved in speech and language processing and production. In cognitive neuroscience, ever since the first demonstration of abstract properties of phonological representations in neural speech processing in Näätänen's pioneer mismatch negativity study (Näätänen et al., 1997), an abundance of neurolinguistic research has demonstrated the neurocognitive reality of phonological representations from phonological features (e.g., Eulitz & Lahiri, 2004; Mesgarani et al., 2014), segments (e.g., Domahs et al., 2009; Ulbrich et al., 2016; White & Chiu, 2017; Wiese et al., 2017), and prosody (e.g., Domahs et al., 2008, 2013; Honbolygó & Csépe, 2013).

However, one of the most important notions in phonological theories (Archangeli & Pulleyblank, 2014; Chomsky & Halle, 1968; Lahiri & Reetz, 2010; Pierrehumbert, 2001), i.e., the level of abstractness that speech sounds are represented in the mind and the brain, has received less attention in neurolinguistic work. Traditional generative theories (e.g., Chomsky & Halle, 1968; Prince & Smolensky, 2004) including underspecification theories (Archangeli, 1988; Lahiri & Reetz, 2010) postulate that phonological representations are abstract, distinct from the phonetic realization of sounds known as surface representation. On the other hand,

exemplar-based theories suggest that speech sounds are represented by traces of the concrete phonetic forms stored in memory (Bybee, 2002; Johnson, 2007; Pierrehumbert, 2001).

A core issue in this quest to understand the nature of phonological representations in phonological theory is the phenomenon of *phonological alternation* (Anttila, 2002). Phonological alternation refers to the realization of morphemes into perceptually distinct speech sounds as conditioned by the interaction among phonological, morphological, and lexical surroundings (Anttila, 2002). Phonological alternation is a core issue in the study of phonology not only because the architecture of phonological representations (e.g., the level of representational abstractedness) is postulated differently across phonological theories, but also as the theorized architecture of phonological representations changes, such changes entail a different understanding of how phonological alternation is realized in the phonetic form, e.g., the nature of the phonological grammar.

Earlier generative phonological theories (Archangeli, 1988; Chomsky & Halle, 1968) suggest that there exist abstract representational primitives known as the underlying representation (UR) that can be distant from their corresponding concrete phonetic forms known as surface representations (or SR) which may alternate in different morphophonological contexts. The theorized roles of the grammar are to transform and specify features of the abstract UR into concrete pronounceable forms (i.e., the SR) according to the morphophonological context. The more recent emergent phonological theory (Archangeli & Pulleyblank, 2014) posits that different allomorphs of a morpheme are listed in parallel and linked with each other in the mental lexicon; derivation of listed allomorphs into SR involves domain-general cognitive combinatorial processes that combine each allomorph with the morphophonological context as candidate SRs, which are then evaluated for their phonological well-formedness. Without positing a distinction between UR and SR, exemplar-based phonological models (Bybee, 2002; Johnson, 2007; Pierrehumbert, 2001) postulate that instead of morphemes represented in isolation, multimorpheme sequences are directly stored in memory, such that alternation as a phenomenon represents the manifestation of frequency effects of the multimorpheme sequences (i.e., which forms of morphemes appear more with each other) and articulatory mechanisms that modify the quality of speech sounds in morphological combinations.

While the long-standing debate on these theoretical models also pertains to issues beyond phonological alternation, the present study focuses on phonological alternation with an aim to advance the current understanding of the architecture of phonological representations. Specifically, we seek to examine the extent to which the abstract phonological representations are more than formal entities, but are in fact neurocognitive primitives that impact speech processing and lexical access. Specifically, we examine the extent to which neural speech processing of phonological alternations is modulated by structures of phonological representations that vary cross-linguistically. Following the spirit of modern phonological frameworks (Mascaró, 2007;

Pierrehumbert, 2016), we hypothesized that phonological representations are multiform in nature, i.e., that phonological representations may take abstract and concrete forms, which are determined by the grammar based on factors including lexical distributional properties of the phonological alternation. We predicted that the processing of surface-similar phonological alternation patterns subserved by different phonological representational structures are neurocognitively distinct.

1.1. The Present Study: The neural processing of surface-similar lexical tone alternation patterns cross-linguistically

The present study focuses on alternations of lexical tones, which are a type of phonemic-level representations that are suprasegmental in nature (Liang & Du, 2018). Lexical tones contrast lexical meaning using pitch patterns that vary in pitch level and contour (e.g., in Mandarin, the syllable /ma/ means ‘mother’ when produced with a high-level pitch pattern, but means ‘horse’ when produced with a dipping pitch pattern) (M. Yip, 2002). We examine the alternation of lexical tones found in Mandarin and Cantonese.

In Mandarin (in the standard Putonghua variety spoken in China), the third tone (mT3) undergoes alternation according to different neighboring lexical tone contexts—a phenomenon known as *tone sandhi*. An mT3 is realized as a rising tone (mTR) when it precedes another mT3, as the citation form (dipping tone, i.e., mTD) before a pause, and as a low falling tone (i.e., mTL) elsewhere (see Table 1).

Neighboring tone context	Example	Surface Tone Pattern
mT1 + /mT3/	<i>zhōng shǐ</i> ‘Chinese history’	mT1 + [mTD]
/mT3/ + mT3	<i>shǐ dǎng</i> ‘historical record’	[mTR] + mTD
/mT3/ + mT1	<i>shǐ shū</i> ‘history book’	[mTL] + mT1

Table 1: Mandarin T3 Sandhi. mT1: Mandarin Tone 1, high-level tone. Romanization of Mandarin is in *Pinyin*.

A surface-similar tone alternation pattern can also be observed in Cantonese. This alternation, known as *pinjam* (i.e., ‘changed tone’), involves syllables which carry a tone from the low register (Tone 4 [low-falling, henceforth *cT4*], Tone 5 [low-rising], and Tone 6 [low-level]) to be realized with a rising tone (*cTR*) (M. Yip, 2002).¹ The lexico-morphological environment where *pinjam* applies is restrictive, with many conditioning environments that manifest differently as a function

¹ Historically, there is another type of *pinjam* that operates on Cantonese high tones (i.e., the alternation between high-falling and high-level tones), but this alternation is not relevant to the phonologies of most modern Cantonese speakers since high-falling and high-level tones are no longer distinctive in modern-day Cantonese (Yu, 2007), and is thus outside of the scope of this study.

of lexical class and the part of speech (Alderete et al., 2022; Kam, 1977). Importantly, although pinjam is often formally analyzed as a result of autosegmental realignment of tone targets due to morphological processes (Alderete et al., 2022; M. J. Yip, 1980; Yu, 2007), the productivity of pinjam as an active linguistic process is unclear (Yu, 2007), with reports as early as in the 1970s suggesting that the relationship between the derived forms and their alleged base are not tacitly recognized (Kam, 1977). For the purpose of cross-linguistic comparison between surface-similar phonological patterns, the current study focuses on one of these lexical environments where pinjam for a low tone occurs *before another low tone*, i.e., name compounding (see **Table 2**).

Honorific Suffix	Example (<i>can4</i> ‘Chan’ [surname])	Surface Tone Pattern
<i>saang1</i>	<i>can4 saang1</i> ‘Mr. Chan’	cTL + cT1
<i>soe4</i>	<i>can4 soe4</i> ‘Officer Chan’	cTR + cTL

Table 2: Cantonese pinjam: cT1: Cantonese Tone 1, high-level tone; cTL: Cantonese low-falling tone.

At the surface level, such tone alternation patterns are very similar between Mandarin tone sandhi and this specific class of Cantonese pinjam (i.e., a Low + Low tone sequence realized as a Rising + Low tone sequence). Despite the surface level similarities, it should be emphasized that unlike in Mandarin, this tone alternation pattern and its specific phonological condition is only restricted to a specific class of lexical items in Cantonese in name + honorific compounds. Unlike the phonologically ill-formed Mandarin mTD + mTD sequence which triggers tone sandhi across most lexical and phonological contexts, all lexical tone combinations are permissible in Cantonese outside the restricted lexical environments where pinjam takes place.

Capitalizing on such similar surface patterns yet crucially different underlying lexical distributions, the present study examines the extent to which the phonological representations subserving Mandarin tone sandhi and Cantonese pinjam, which are similar in the surface form, differ in their shape due to the different lexical distributional properties. The phonological representation subserving Mandarin tone sandhi has been extensively investigated experimentally. In particular, neural Event-Related Potential (ERP) paradigms tapping into automatic pre-attentive auditory discrimination (J. C. Lau et al., 2019; Li & Chen, 2015; Politzer-Ahles et al., 2016), priming (Chien et al., 2016; Meng, Wynne, & Lahiri, 2021), visual word processing (Nixon et al., 2015), and speech elicitation tasks (Y. Chen et al., 2011; Zhang & Lai, 2010) have reported mixed results on the shape of the phonological representation subserving Mandarin tone sandhi. Results interpreted to support a traditional single representation (Chien et al., 2016; Meng, Wynne, & Lahiri, 2021; Zhang & Lai, 2010), listed allomorphy (Y. Chen et al., 2011; Li & Chen, 2015; Nixon et al., 2015), and underspecification (Politzer-Ahles et al., 2016) were all

identified. Such mixed results were potentially due to the different degree of acoustic control of stimuli (Politzer-Ahles et al., 2016), lack of stimulus variability (Li & Chen, 2015), and the use of auditory stimulation (Zhang & Lai, 2010), and primes (Chien et al., 2016; Meng, Wynne, & Lahiri, 2021) that were not designed to tease apart all potential forms of phonological representations.

Recent research has revisited the issue of Mandarin tone sandhi by focusing on understanding the time course and nature of both production (X. Chen et al., 2022; Zhang et al., 2022) and perceptual processes (Zeng et al., 2021) related to tone sandhi. This series of studies has provided novel and critical insight into the timescale of speech production (with the UR and SR involved in different stages of production) (X. Chen et al., 2022; Zhang et al., 2022) and perception of Mandarin tone sandhi (with SR-UR mapping concurring before the second syllable of a disyllabic stimulus, indicative of incremental processing of alternation). However, the theoretical question of the shape and architecture of the UR is not the primary focus of these studies.

Therefore, the extent to which the speech processing of tone alternation may vary as a function of different underlying representational structures and corresponding factors, such as lexical distribution and grammatical productivity of the alternation, remains an important question. For example, Cantonese pinjam, which offers an interesting contrastive model with critical distributional and productivity differences, has not been investigated psycholinguistically or neurolinguistically to our knowledge.

1.2. Event-Related Potential Cross-Model Priming Study: Empirical Background and Linking Theories

To this end, the visual-to-auditory cross-modal priming paradigm provides an optimal approach to interrogate the shape of phonological representations subserving lexical tone alternation. Priming refers to an implicit memory effect wherein the response to a stimulus (the target) is influenced by another previously presented stimulus (the prime) (Meyer & Schvaneveldt, 1971). Converging evidence in the behavioral experimental literature suggests that judgment of spoken words (auditory targets) can be influenced by word stimuli presented in orthographic forms (visual primes) (Jakimik et al., 1985; Seidenberg & Tanenhaus, 1979; Ventura et al., 2004), hence a cross-modal (visual to auditory) priming effect.

In the context of phonological studies, an ample body of work has established that in a cross-modal priming paradigm, phonological representations stored in the mental lexicon are part of the memory trace elicited by the orthographic prime. Specifically, both behavioral (Lahiri & Reetz, 2002) and ERP evidence (Friedrich et al., 2008; Lahiri & Reetz, 2010) suggest that an abstract level of representation corresponding to the UR is tapped in priming paradigms, with a broader ERP literature suggesting that the abstract UR can be directly tapped in speech processing experiments (e.g., Eulitz & Lahiri, 2004; Zeng et al., 2021). In priming paradigms, phonological properties of more fine-grained SRs elicited by the auditory target are mapped and compared with the more abstract UR elicited by the prime — when there is a mismatch

between SR and UR, ERP components are elicited and behavioral response is delayed; if there is no mismatch between SR and UR, no associated ERP components or behavioral differences will be elicited (Friedrich et al., 2008; Lahiri & Reetz, 2002, 2010).

In general, two types of ERP responses are associated with the processing of phonological violations in cross-modal priming paradigms, reflecting a mismatch between the UR and SR, namely, the N400 and Late Positive Complex (LPC) (Bohn et al., 2013; Domahs et al., 2008, 2013, 2014; Henrich et al., 2014; Molczanow et al., 2013).

The N400 is a negative ERP component that usually peaks at around 300–500 ms post-stimulus, and is most robust in centro-parietal electrode positions (E. F. Lau et al., 2008). N400 has been identified as a neural marker for lexico-semantic integration in language processing (Kutas & Federmeier, 2000). A similar N400 effect was also found in cross-modal priming paradigms, wherein N400 was elicited when the visual orthographic prime was not phonologically related to the auditory target, compared to when the prime and target were phonologically related (J. E. Anderson & Holcomb, 1995; Holcomb, 1993; Holcomb et al., 2005; Kiyonaga et al., 2007). The presence of N400 effects in cross-modal priming experiments suggests that the properties of the lexicon as tapped into by N400-related mechanisms are multidimensional in nature (i.e., not only semantic, but also phonological).

The N400 component is often accompanied by the LPC (e.g., Curran et al., 1993; Karayanidis et al., 1991; Woodward et al., 1993), a posterior-distributed ERP component that peaks at a later 400–1200 ms post-stimulus time window (Friedman & Johnson, 2000). While LPC has generally been understood as an index to memory recollection (Friedman & Johnson, 2000), in the context of language processing experiments, it has been associated with the level of mental effect in conscious semantic understanding (Juottonen et al., 1996), lexical judgment (Domahs et al., 2009; Finnigan et al., 2002), as well as lexical “patching” in erroneous contexts (Daltrozzo et al., 2012).

In the context of phonological processing, a body of phonological studies has found that phonotactically illegal phonological sequences (Domahs et al., 2009) as well as phonological violations that impact lexical integration (Domahs et al., 2013) both elicit the N400. The N400 may be accompanied by the LPC, in cases where restructuring (i.e., “patching”) of such phonological violations to achieve lexical access is possible (Bohn et al., 2013; Domahs et al., 2008, 2013, 2014; Henrich et al., 2014; Molczanow et al., 2013).

1.3. Predictions

With the empirically supported assumption that phonological representations stored in the mental lexicon can be tapped into using a cross-modal priming paradigm (Domahs et al., 2013; Friedrich et al., 2008; Lahiri & Reetz, 2010, 2002), the current study examines the structure of phonological representations subserving a set of surface-similar, but distributionally distinct tone alternation patterns in Mandarin tone sandhi and Cantonese pinjam.

We examined three potential forms of phonological representations of such tone alternation patterns, namely a low tone appearing as a rising tone when it precedes another low tone (also summarized in **Table 3**).

1. *Single underlying representation (Single UR)*: The Single UR account posits that the surface Rising tone of the first syllable is abstractly represented as a Low tone in the underlying representation (Chien et al., 2016). The Low tone is transformed into a Rising tone by a phonological process when it precedes another low tone.
2. *Listed Allomorphy (Listed UR)*: The Listed UR account postulates that both variants of the tone (i.e., Rising and Low) are *listed* in parallel in the underlying representation (Y. Chen et al., 2011). Alternation is achieved with phonological processes that generate and evaluate potential combinations based on both variants (i.e., Rising + Low vs. *Low + Low), to elect the more phonologically well-formed phonological sequence to surface (Mascaró, 2007).
3. *Single surface representation (Single SR)*: The Single SR account posits that the surface phonological pattern of the lexical item (Rising + Low sequence) is directly stored in the mental lexicon, without postulating an abstract form that deviates from its surface form (Chien et al., 2016; X. Zhou & Marslen-Wilson, 1997).

Using a cross-modal priming paradigm performed in Mandarin (Experiment 1) and Cantonese (Experiment 2), the present study compares behavioral and ERP responses to three types of phonological violations of a disyllabic auditory target:

1. *Underapplication* of tone alternation: lack of tone sandhi or pinjam in phonological and lexical contexts requiring tone alternation, i.e., [*Low + Low],
2. *Overapplication* of tone alternation: tone sandhi or pinjam in phonological and lexical contexts that do not require tone alternation, i.e., [*Low + High], and
3. *Plain violations*: pronunciation with the wrong tone category, i.e., [*High + Low] and [*High + High], with the correct pronunciation form of the visual prime as baseline.

As the control condition, it was predicted that in both Mandarin and Cantonese experiments, *plain violations* would elicit a N400 response neurophysiologically, with a lexical violation where the auditory target mismatches the tone category of the prime. Behavioral judgment would be the quickest, with the lexical violation being the most straightforward to detect.

Single UR, *Listed UR*, and *Single SR* yielded different predictions on the underapplication and overapplication conditions (schematized in **Table 3**):

1. *Single UR*: In the *underapplication condition* [*Low + Low], the N400 is expected due to the presence of phonological violations of two adjacent Low tones. Yet, with the alternating tone underlyingly a /Low/ tone, lexical access will not be hindered. However, additional

resources required to restructure the deviant phonological form will manifest as the LPC and a slower behavioral response. In contrast, since the Single UR does not contain a Rising tone, the *overapplication* condition [*Rising + High] will lead to a lexical mismatch with the prime, manifesting as an N400 only, no different from plain violations, with comparable behavioral response.

2. *Listed UR*: We posit that both variants of the alternating surface tones stored in the UR will be co-elicited by the prime. Therefore, neither *underapplication* nor *overapplication* forms will lead to a target-prime mismatch, such that lexical access will not be hindered in either condition. However, additional resources are required to restructure the incorrect allomorph into the corrected listed phonological form, as reflected by the LPC predicted for both conditions. Behavioral response would be slower due to this restructuring process. The N400 would also be elicited in both *underapplication* and *overapplication* conditions, reflecting phonological violations in the wrong derivation of the correct allomorphic form in both conditions.
3. *Single SR*: The N400 response would be elicited for both *underapplication* and *overapplication* conditions, with a mismatch between the auditory target and the form of phonological representation elicited by the prime (which does not undergo tone alternation transformation), no different from plain violations. Likewise, behavioral responses would not be different from plain violations.

Condition	Response	Single UR /L/	Listed UR /L~R/	Single SR /R + L/ (multisyllabic chunks)
Underapplication [*L + L]	Behavioral RT	Slower	Slower	Fast
	N400	Yes	Yes	Yes
	LPC	Yes	Yes	No
Overapplication [*R + H]	Behavioral RT	Fast	Slower	Fast
	N400	Yes	Yes	Yes
	LPC	No	Yes	No
Plain Violation [*H + L]	Behavioral RT	Fast	Fast	Fast
	N400	Yes	Yes	Yes
	LPC	No	No	No

Table 3: Competing hypotheses and predictions: Hypothesized phonological representational shapes of Mandarin tone sandhi and Cantonese *pinjam* in Single Underlying Representation (UR), Listed UR, and Single Surface Representation (SR) accounts, and their corresponding predicted results on behavioral response time (RT) and ERP experiments. L: Low tone (note that in Mandarin the L tone in the second syllable is realized as a dipping tone); R: Rising tone; H: High tone.

2. Methods

2.1. Experiment 1: Mandarin tone sandhi

2.1.1. Participants

16 native speakers of Mandarin (10 females) with an average age of 23.25 years ($SD = 1.69$) who reported no neurological or developmental conditions participated in the study. Participants self-reported normal hearing in both ears. All participants were born and raised in Northern areas of Mainland China and reported to only speak the Putonghua variety of Mandarin as their native language.

2.1.2. Stimuli

The stimuli consisted of 20 disyllabic Mandarin Chinese names (see Table S1, *Supplementary Materials*). Each name was composed of a monosyllabic family name (e.g., *mǎ* /ma3/) compounded with an honorific morpheme. A total of 10 monosyllabic family names all carrying mT3 were included. Each of the 10 family names was compounded with two types of honorific suffixes, namely *zǒng* /tsoŋ3/ ‘general manager’, and *gōng* /gong1/ ‘engineer’, resulting in the 20 disyllabic Mandarin Chinese names. The two honorific suffixes carry different lexical tones (mT3 and mT1 respectively), which in combination with the mT3 family names, results in the family names in the first syllable realizing with different surface tones due to tone sandhi. For example, the family name *lǐ* with a mT3 is realized as a mTR when compounded with the mT3 suffix *zǒng* (i.e., *sandhi names*), but realized as a mTL when compounded with the mT1 *gōng* (i.e., *non-sandhi names*).

For each disyllabic item for the sandhi name and non-sandhi name stimuli, a total of three versions were produced from recordings by a male, phonetically trained native speaker of Beijing Mandarin.

For sandhi names, the first version was the *correct* pronunciation that involved a *correct application* of the sandhi rule (i.e., with a mTR+mTD surface tone). The second and third versions were *violation* conditions wherein the pronunciation of the first syllable deviated from the correct form in terms of surface tone realization. The second version of sandhi names was with an *underapplication* of the sandhi rule (i.e., with a *mTL+mTD combination). The third version of sandhi names was a control condition with a *plain violation* of the tone in the first syllable (i.e., with a *mT1+mTD combination). Likewise, for non-sandhi names, the first version was the correct pronunciation that involved a *correct non-application* of the sandhi rule (i.e., with a mTL+mT1 surface tone). The second version of non-sandhi names was with an *overapplication* of the sandhi rule (i.e., with a *mTR+mT1 combination). The third version of non-sandhi names was a control condition with a *plain violation (2)* of the tone in the first syllable (i.e., with a *mT1+mT1 combination). The 10 family names were selected such that when the syllables were

realized as a mTR (which overlaps with mT2) or mT1, none of them were permissible family names.

These disyllabic stimuli were produced by combining two separate recordings, each with a single syllable; the two syllables were extracted from separate recordings uttered in phonotactically legal disyllabic contexts. For each family name, the same recordings for the three tone variants (mTL, mTR, and mT1) were used across both suffix type conditions. The same recording for each of the two suffixes was also used across all trials in their respective conditions. For each disyllabic word, the first syllable was duration-normalized in Praat to 390 ms, and the second syllable to 440 ms. Intensity was normalized to 70 db for all stimuli. Together, all stimuli were maximally controlled acoustically across conditions. Fundamental frequency (F0) properties of all stimuli are presented in **Figure 1**.

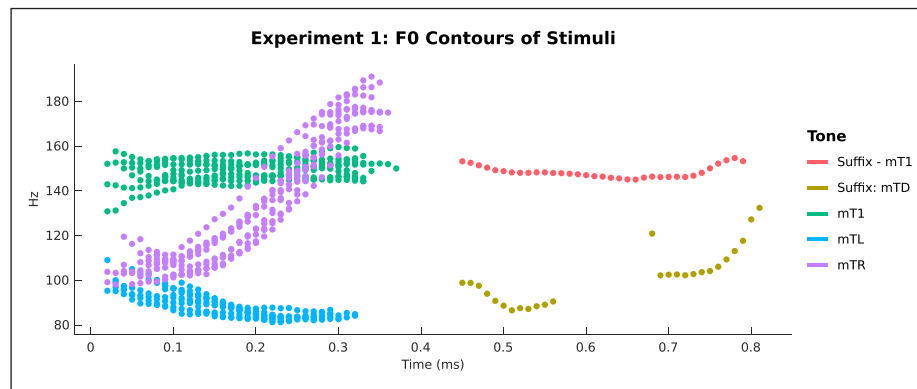


Figure 1: Experiment 1: F0 properties of stimuli. Note that the lack of f0 values detected in portions of mTD corresponds to creakiness typical in the tone.

2.1.3. Design

The current experiment utilized a masked cross-modal priming paradigm adopted for ERP testing by Kiyonaga et al. (2007). The paradigm was presented to the participant as a phonological judgment task: they were instructed to judge whether the auditory stimulus presented (the auditory target) was how they would pronounce the word which was immediately previously presented on the screen in Chinese characters (the visual prime). The Chinese characters always represented the correct form of the auditory target.

While most cross-modal priming tasks with orthographic primes involve alphabets (Bohn et al., 2013; Domahs et al., 2008, 2013, 2014; Henrich et al., 2014; Molczanow et al., 2013), logographic characters of Chinese are utilized in the cross-modal priming design of the current study. Although phonological information is not transparent for Chinese characters, we assumed that the Chinese characters could be a prime to an auditory target since phonological

information is elicited in visual processing of Chinese characters (Kuo et al., 2004; Tan et al., 2005; Wu et al., 2012; W. Zhou et al., 2018), and Chinese characters have also been used to prime speech production (You et al., 2012).

Each trial began with a blank screen for 2000 ms. Then, a fixation point was presented in the middle of the screen for 500 ms. A visual mask (two hash marks) was then presented and remained for 500 ms, and then replaced by the prime word displayed in Chinese characters. After 100 ms, the prime was immediately replaced by the visual mask. The auditory target was presented 13 ms after this second presentation of the visual mask. ERPs were time-locked to the onset of the auditory target presentation. After the presentation of the auditory target (all of which were 830 ms) was a 2000 ms time window whereby the participant gave their behavioral response. This response window was terminated once a response was recorded. The backward mask remained on the screen until the termination of the response window. The flow of each trial is presented in **Figure 2**.

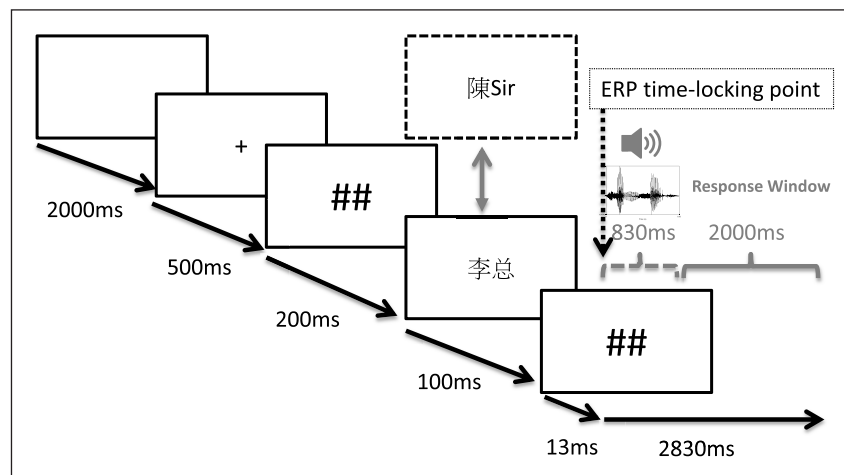


Figure 2: Experimental design: This figure visualizes one trial of the experimental paradigm, the sequence of which is identical for both Experiments 1 and 2. A total of 480 of these trials were presented in both experiments. In this figure, the example of the visual prime for Experiment 2 is denoted in the dotted rectangle above the slide presenting the visual prime in Experiment 1.

For each participant, 480 experimental trials were presented. Before ERP testing, the participant was instructed to sit in a comfortable chair. Auditory stimuli were delivered at 75 dB SPL to the participant's right ear through insert earphones (ER-3a, Etymotic Research, Elk Grove Village, USA). Visual stimuli were presented on a 19" LCD monitor with a 640 × 480 resolution (Dell P1913S) as white characters on a black background at size 60. The position of the screen was adjusted for each participant such that the distance between the screen and the participant's eyes was 70 cm, and the angle between the eyes and the center of the screen was approximately 25 degrees.

The stimuli presentation paradigms were arranged and presented with the stimulus presentation system E-Prime (Psychology Software Tools Inc., USA). Instructions for the phonological judgment task were delivered by an introductory text (in Chinese characters) preceding the experiment. The participant was instructed to respond by pressing corresponding buttons on a respond pad with their left and right thumbs indicated as “yes” or “no” by green and red color labels respectively on the respond pad. The positions of the “yes” and “no” buttons (and hence the green and red labels on the response pad) were counterbalanced across participants.

To familiarize the participant with the task as well as the positions of the buttons, each participant received 16 practice trials before the experimental trials. The 16 stimuli in the practice trials were randomly drawn from the list of stimuli. In the experimental trials, breaks were provided after every 120 trials. Each experimental session lasted approximately 30 minutes.

Behavioral responses, time locked to the onset of the auditory target, were recorded. Accuracy and RT (time-locked to the onset of the auditory stimulus) metrics were then calculated for each trial online and averaged across all trials by condition for each participant.

2.1.4. Electrophysiological recording

Electrophysiological responses were scalp-recorded using electroencephalography (EEG) in a sound-attenuated and RF-shielded booth located at The Chinese University of Hong Kong. EEGs were recorded using 34 Ag/AgCl electrodes mounted on an infracerebral electrode cap (Easycap) according to the International 10-20 locations, and connected to a SynAmps2 Neuroscan Inc. system (Compumedics Ltd., USA). The CPz electrode functioned as the online reference and Fpz electrode served as the ground electrode. Contact impedance was maintained below 10 k Ω whenever possible. EEG was acquired with Curry 7 (Compumedics Ltd., USA) and digitized online at a 1 kHz sampling rate.

2.1.5. Data analysis

Each participant’s raw EEGs were analyzed offline using EEGLAB (Delorme & Makeig, 2004) with the ERPLAB plugin (Lopez-Calderon & Luck, 2014). Raw EEGs were first band-pass filtered from 0.1 to 30 Hz with a roll-off of 12 dB per octave. EEGs were decomposed using the extended infomax independent component analysis (ICA) algorithm supplied by the EEGLAB. ICA was also implemented using ADJUST, an algorithm that automatically identifies and rejects stereotyped temporal and spatial artifacts (Mognon et al., 2011). Components that contained stereotyped oculomotor or motor activity, as suggested by ADJUST, were visually confirmed and removed from the dataset. All electrodes were then re-referenced offline to the averaged mastoid (TP9 + TP10). ERPs were then computed for each condition and electrode. Since there were twice as many trials in each correct condition relative to each of the two violation conditions, only half of the total number of correct trials were randomly drawn and averaged as the ERP waves.

Epochs began 200 ms before stimulus onset (of the auditory target) and 1500 ms post-stimulus onset, and were baseline corrected to a $-200-0$ ms window. Epochs with activities greater than $\pm 100 \mu\text{V}$ were considered artifacts and rejected. Difference waves were then computed by subtracting the ERP waves associated to the violation conditions from those from the correct condition (i.e., [1] *underapplication-correct application*; [2] *plain violation-correct application*; [3] *overapplication-correct non-application*).

Given the relatively low number of trials for each condition (60 trials), an a priori decision was made to use a jackknife-based analysis method on the ERPs to reduce potential Type I and II errors due to the low signal-to-noise ratio (SNR) of individual ERP waves (Luck, 2014; J. Miller et al., 1998; Ulrich & Miller, 2001). With a total of 16 participants, 16 leave-one-out grand averages were computed, and jackknife mean amplitude and fractional peak latency (50%) were measured from the 16 leave-one-out grand averages.

Analyses focused on three electrodes (Pz, P3, P4); further analyses were performed based on the average of these three electrodes in each condition, focused on two time windows: an early window (400–800 ms) and a late window (800–1200 ms). These electrode and time windows were selected a priori, according to the literature suggesting a central-parietal scalp distribution of N400 and LPC (Late Positive Complex) responses (Friedman & Johnson, 2000; E. F. Lau et al., 2008), confirmed during pilot runs of this experiment. The pilot results showed N400 and LPC-like components with a centro-parietal distribution at approximately 400–800 ms and 800–1200 ms respectively after collapsing all conditions. As such, fractional peak latency was quantified from *negative* fractional peaks at the 400–800 ms window, and from *positive* fractional peaks at the 800–1200 ms window.

2.1.6. Statistical analysis

Together, this design constitutes a 2 (CONTEXT: sandhi vs. non-sandhi names) \times 2 (VIOLATION TYPE: phonological violations vs. plain violations) repeated measures design, as summarized in **Table 4**.

A series of linear mixed effect models (LMMs) were fit to each behavioral and ERP metric, with CONTEXT, VIOLATION TYPE and their interaction as fixed factors.

We focused on averaged RT from all correctly rejected trials as a behavioral metric because of a ceiling effect identified in the accuracy measure. Before subsequent statistical analysis, RT (in ms) was log-transformed.

Individual participants were fit as a random factor for the model on RT for the model on ERP metrics; the number of jackknife samples (from 1 to 16) was fit as a random factor in the ERP analysis. Separate LMMs were fit on different time windows and for different ERP metrics.

CONTEXT	VIOLATION TYPE		
	Correct	Phonological Violation	Plain Violation
Sandhi/Pinjam Names: Mandarin: -zǒng Cantonese: -soe4	<i>Correct Application:</i> (mTR + mTD / cTR + cTL) 120 trials	<i>Underapplication:</i> (*mTL + mTD / cTL + cTL) 60 trials	<i>Plain Violation:</i> (*mT1 + mTD / cT1 + cTL) 60 trials
Non-Sandhi/Non-Pinjam Names: Mandarin: -gōng Cantonese: -saang1	<i>Correct Non-Application:</i> (mTL + mT1 / cTL + cT1) 120 trials	<i>Overapplication:</i> (*mTR + mT1 / cTR + cT1) 60 trials	<i>Plain Violation 2</i> (*mT1 + mT1 / cT1 + cT1) 60 trials
			Total: 480 Trials

Table 4: Experimental Design: Shared by Experiments 1 and 2.

Since analyses on ERP metrics were based on jackknife ERP samples, and the jackknife variance was artificially low, the F values were corrected by dividing the uncorrected F values by the square of the total number of participants minus one (Luck, 2014; J. Miller et al., 1998; Ulrich & Miller, 2001). Corrected p values of the fixed factor and contrasts were then computed using the corrected F value accordingly.

In case that the CONTEXT \times VIOLATION TYPE interaction was significant, three planned contrasts were examined, namely on (1) *Underapplication* versus *Plain violation*, (2) *Overapplication* versus *Plain violation 2*, and (3) *Underapplication* versus *Overapplication* conditions. p values of the three planned contrasts were adjusted for the jackknife method and for multiple comparison via the Holm-Bonferroni method.

2.2. Results

Given the experimental design outlined in **Table 4**, we identified both behavioral and ERP differences in the processing of Mandarin tone sandhi, partly consistent with the set of predictions outline in **Table 3**.

2.2.1. Behavioral results

Figure 3 shows the mean behavioral RT for all conditions. The linear mixed effects model (LMM) on RT (see **Table 5**) revealed main effects of both VIOLATION TYPE ($p < .001$) and CONTEXT ($p < .001$), suggesting a slower RT in phonological violations as compared to plain violations, as well as a slower RT when processing stimulus types sharing a suffix that carries a mT3.

The VIOLATION TYPE \times CONTEXT interaction ($p < .001$) was also significant. Planned comparisons revealed that the RT of the *Underapplication* condition was significantly higher than the *Plain violation* condition ($p < 0.001$, Holm-adjusted). RT in the *Overapplication* condition was also higher than that in the *Plain violation 2* condition ($p < 0.001$, Holm-adjusted). RT of the *Underapplication* condition was also higher than the *Overapplication* condition ($p < 0.001$, Holm-adjusted).

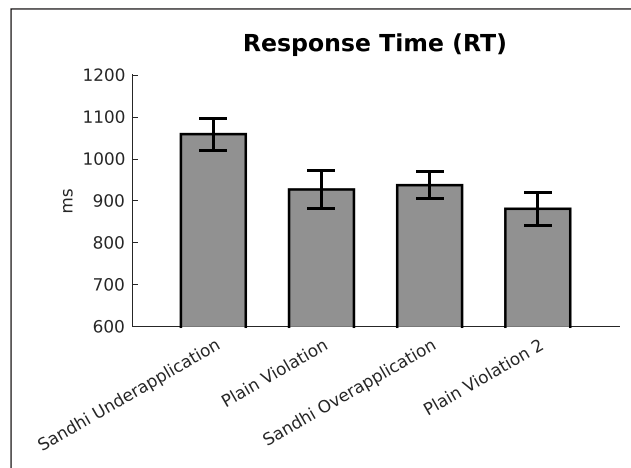


Figure 3: Results: Behavioral response time in Mandarin tone sandhi judgment. Error bars represent ± 1 standard error of the mean.

2.2.2. ERP results

The grand averaged ERP difference waves and topographical maps of scalp voltage of the difference waves at the two time windows of analysis are presented in **Figure 4**. In the 400–800 ms window, negative components with centro-parietal distributions can be observed in the *Underapplication* and *Overapplication* conditions, but not in the *Plain violation* and *Plain violation 2* conditions. In the 800–1200 ms window, sizeable positive components with centro-parietal distributions could be observed in the *Underapplication* and *Overapplication* conditions, but not in the *Plain violation* and *Plain violation 2* conditions.

Figure 5 (Panels A and B) presents the mean amplitude and fractional peak latency of all conditions in the 800–1200 ms window.

Dependent Variable	Factor	<i>F</i>	<i>p</i>
Behavioral Response Time (RT)	Intercept	30621	< 0.001
	CONTEXT	42.786	< 0.001
	VIOLATION TYPE	61.218	< 0.001
	CONTEXT × VIOLATION TYPE	7.318	0.009
Mean Amplitude: 800–1200 ms	Intercept	8.236	0.006
	CONTEXT	0.737	0.394
	VIOLATION TYPE	12.450	0.001
	CONTEXT × VIOLATION TYPE	1.858	0.178
Fractional Positive Peak Latency: 800–1200 ms	Intercept	52.448	< 0.001
	CONTEXT	0.012	0.914
	VIOLATION TYPE	3.277	0.075
	CONTEXT × VIOLATION TYPE	0.280	0.599
Mean Amplitude: 400–800 ms	Intercept	1.704	0.197
	CONTEXT	1.177	0.282
	VIOLATION TYPE	0.095	0.759
	CONTEXT × VIOLATION TYPE	0.208	0.650
Fractional Negative Peak Latency: 400–800 ms	Intercept	16.399	< 0.001
	CONTEXT	0.000	0.999
	VIOLATION TYPE	3.250	0.076
	CONTEXT × VIOLATION TYPE	0.134	0.715
*Mean Amplitude: 400–1200 ms	Intercept	1.768	0.189
	CONTEXT	0.001	0.971
	VIOLATION TYPE	7.163	0.010
	CONTEXT × VIOLATION TYPE	0.711	0.402
*Peak Amplitude: 400–1200 ms	Intercept	9.139	0.004
	CONTEXT	0.724	0.398
	VIOLATION TYPE	1.358	0.249
	CONTEXT × VIOLATION TYPE	0.441	0.509
*Fractional Negative Peak Latency: 400–1200 ms	Intercept	30.849	< 0.001
	CONTEXT	0.000	0.996
	VIOLATION TYPE	10.570	0.002
	CONTEXT × VIOLATION TYPE	0.638	0.428

Table 5: Linear mixed-effect model results for Experiment 1. *F* and *p* values have been adjusted for the jackknife method for ERP metrics. Significant factors are in bold. *Post-hoc analyses where more stringent alpha levels (.025 for two N400-related latency measures and .013 for three N400-related amplitude measures) apply.

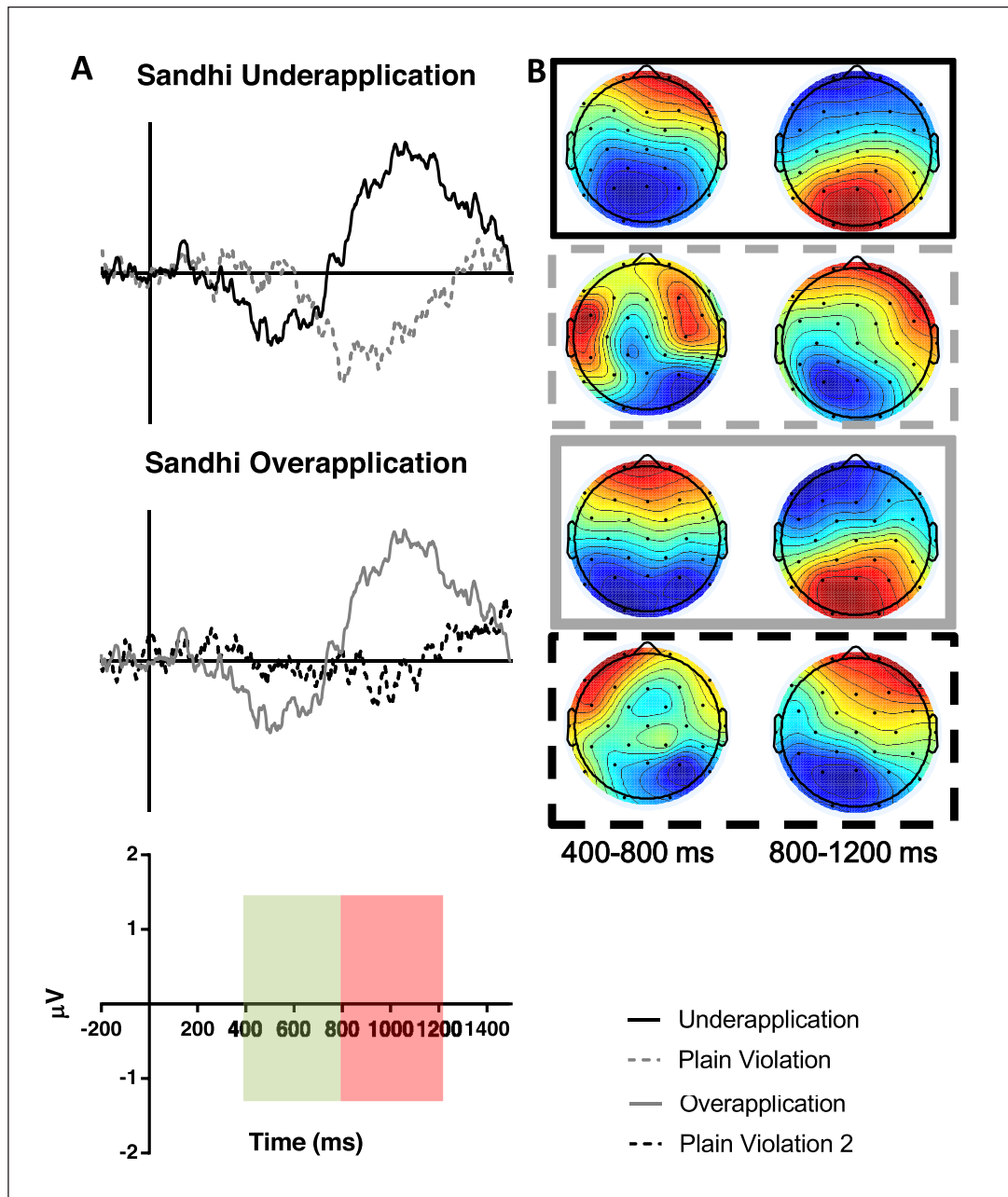


Figure 4: Results: Brain event-related potentials: Panel A presents the ERP difference waves (average of P3, Pz, and P4 channels) associated with *sandhi underapplication* (top) and *overapplication* (bottom), each against the respective plain violation conditions. Panel B presents the topographical maps of scalp voltage (mean amplitude) of the difference waves averaged per channel across 400–800 ms (left column) and 800–1200 ms (right column) time windows, of *sandhi underapplication* (1st row), *overapplication* (3rd row), and their associated plain violation conditions (2nd and 4th rows).

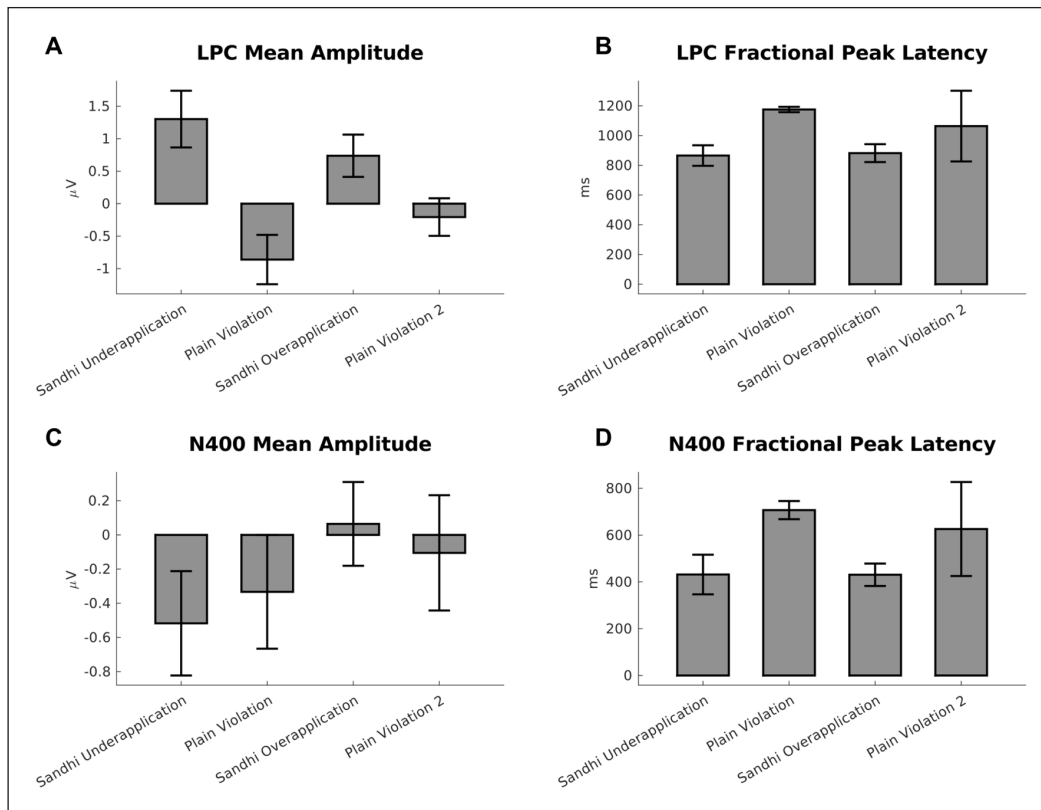


Figure 5: Results: Measurements of mean amplitude (Panels A and C) and fractional peak latency (Panels B and D) of the average of P3, Pz, and P4 channels for all conditions from the 400–800 ms (Panels C and D) and 800–1200 ms (Panels A and B) time windows. Error bars represent ± 1 standard error of the mean with a jackknife method.

Results of the linear mixed effect models (LMMs) (jackknife-corrected) are presented in **Table 5**. The LMM on mean amplitude of the 800–1200 ms window revealed a main effect of VIOLATION TYPE ($p < .001$), suggesting that phonological violations in tone sandhi (both underapplications and overapplications) elicited higher mean amplitude at this time window. In contrast, the main effect of CONTEXT ($p = .394$) and the CONTEXT \times VIOLATION TYPE interaction ($p = .178$) were not significant, suggesting that mean amplitude at this time window did not differ across underapplication and overapplication conditions.

The LMM on fractional peak latency of the 800–1200 ms window did not reveal any significant main effect or interaction.

Figure 5 (Panels C and D) presents the mean amplitude and fractional peak latency of all conditions in the 400–800 ms window. LMMs on mean amplitude and fractional peak latency of this time window did not reveal any main effect or interaction.

2.2.3. Post-hoc alternate analysis on the earlier negative component

For the 400–800 ms window, negative components can be observed in *Underapplication* and *Overapplication* conditions, but not for the *Plain violation* and *Plain violation 2* conditions. However, post-hoc visual inspections on the ERP waves revealed sizeable negative components which peaked at a later time window for the *Plain violation* and *Plain violation 2* conditions. These components peaked at around 1000 ms, well beyond the a priori 400–800 ms time window. As a result, post-hoc analyses on mean amplitude and fractional negative peak latency were computed on a larger time window that encompassed 400–1200 ms.

Results of the LMMs of this post-hoc analysis focusing on the 400–1200 ms window are summarized in **Table 5**.

The LMM on mean amplitude of the 400–1200 ms window revealed a main effect of VIOLATION TYPE. However, it must be noted that the mean amplitude measure in the 400–1200 ms window may not be appropriate since for the *Underapplication* and *Overapplication* conditions, this window also encompassed the time window for the late positive component. As a result, local *peak* amplitude only for *negative peaks* in the 400–1200 ms window were also measured, despite the fact the peak amplitude may not be a reliable measurement particularly incorporated with jackknife analysis. Indeed, just considering negative peaks, the peak amplitude of the 800–1200 ms window did not reveal any significant main effect or interaction. **Figure 6** presents the mean amplitude (Panel A) and negative peak amplitude (Panel B) of all conditions in the 400–1200 ms window.

Yet importantly, the intercept of the LMM on negative peak amplitude (400–1200 ms) was significant (see **Table 5**). This confirms the visual observation of the presence of N400-like negative components at the larger 400–1200 ms time windows, which did not vary in amplitude as a function of CONTEXT or VIOLATION TYPE.

Figure 6 (Panel C) presents the fractional peak latency of all conditions in the 400–1200 ms window. Results of the LMM on this window revealed a main effect of VIOLATION TYPE ($p = .01$), even considering a more stringent Bonferroni-adjusted alpha of .025 due to an additional model on N400 latency. This main effect suggests that phonological violations in tone sandhi (both underapplications and overapplications) elicited an earlier negative peak latency mean amplitude at this time window, as compared to plain violations. In contrast, the main effect of CONTEXT ($p = .996$) and the CONTEXT \times VIOLATION TYPE interaction ($p = .428$) were not significant, suggesting that negative peak latency at this time window did not differ across underapplication and overapplication conditions.

2.3. Experiment 2: Cantonese Pinjam

2.3.1. Participants

16 native speakers of Hong Kong Cantonese (nine females) with an average age of 21 (SD = 1.15) participated in this experiment. As identical to Experiment 1, participants reported to have no neurological, hearing, or developmental conditions.

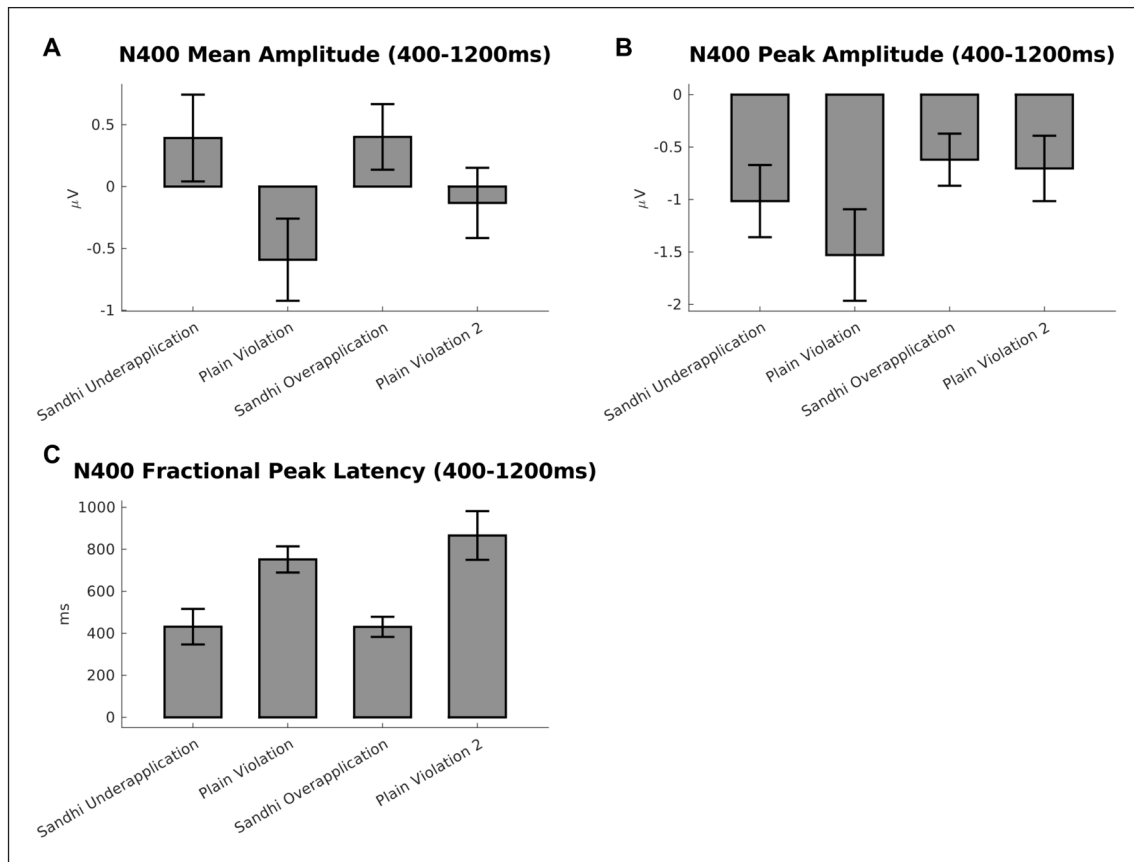


Figure 6: Results: Measurements of mean amplitude (Panel A), peak amplitude (Panel B), and fractional peak latency (Panel C) of the average of P3, Pz, and P4 channels for all conditions from the 400–1200 ms time windows. Error bars represent ± 1 standard error of the mean with a jackknife method. *** $p < 0.001$ in planned contrast (Holm-Bonferroni corrected).

2.3.2. Stimuli

The design of Experiment 2 closely followed that of Experiment 1. The stimuli consisted of 20 disyllabic Cantonese names (see Table S2, *Supplementary Materials*). Each name was composed of a monosyllabic family name (e.g., *can4* /ts^hɛn4/) compounded with an honorific morpheme. A total of 10 monosyllabic family names all carrying cT4 were included. Each of the 10 family names was compounded with two types of honorific suffixes, namely *soe4* /sɔ̃:4/ ‘teacher, or officer’, and *saang1* /sa:ŋ1/ ‘mister’, resulting in the 20 disyllabic Cantonese names.

The two honorific suffixes carry different lexical tones (cT4 and cT1 respectively), which in combination with the T4 family names, results in the family names in the first syllable realizing with different tones. The 10 family names chosen here were among the cT4 family names which must undergo pinjam. For example, the family name *can4* with a cT4 is realized as a cTR when compounded with the cT4 suffix *soe4* (i.e., *pinjam names*). When compounded with the cT1 suffix *saang1*, the cT4 remained as a cTL (i.e., *non-pinjam names*).

Dependent Variable	Factor	<i>F</i>	<i>p</i>
Behavioral Response Time (RT)	Intercept	1872	< 0.001
	CONTEXT	18.314	< 0.001
	VIOLATION TYPE	46.030	< 0.001
	CONTEXT × VIOLATION TYPE	0.653	0.422
Mean Amplitude: 800–1200 ms	Intercept	0.090	0.765
	CONTEXT	0.326	0.570
	VIOLATION TYPE	1.741	0.192
	CONTEXT × VIOLATION TYPE	0.006	0.940
Fractional Positive Peak Latency: 800–1200 ms	Intercept	83.107	< 0.001
	CONTEXT	0.379	0.541
	VIOLATION TYPE	0.077	0.782
	CONTEXT × VIOLATION TYPE	0.057	0.812
Mean Amplitude: 400–800 ms	Intercept	4.951	0.030
	CONTEXT	0.246	0.622
	VIOLATION TYPE	0.005	0.943
	CONTEXT × VIOLATION TYPE	0.304	0.584
Fractional Negative Peak Latency: 400–800 ms	Intercept	19.538	< 0.001
	CONTEXT	0.055	0.816
	VIOLATION TYPE	1.162	0.285
	CONTEXT × VIOLATION TYPE	0.053	0.818
*Mean Amplitude: 400–1200 ms	Intercept	1.671	0.201
	CONTEXT	0.332	0.566
	VIOLATION TYPE	0.516	0.475
	CONTEXT × VIOLATION TYPE	0.058	0.810
*Peak Amplitude: 400–1200 ms	Intercept	9.780	0.003
	CONTEXT	0.344	0.560
	VIOLATION TYPE	0.043	0.836
	CONTEXT × VIOLATION TYPE	0.013	0.909
*Fractional Negative Peak Latency: 400–1200 ms	Intercept	19.618	< 0.001
	CONTEXT	0.041	0.839
	VIOLATION TYPE	1.164	0.285
	CONTEXT × VIOLATION TYPE	0.151	0.699

Table 6: Linear mixed-effect model results for Experiment 2. *F* and *p* values have been adjusted for the jackknife method for ERP metrics. Significant factors are in bold. *Post-hoc analyses where more stringent alpha levels (.025 and .013) apply.

Like Experiment 1, for each disyllabic *pinjam name* and *non-pinjam name* stimuli, a total of three versions were produced from recordings recorded by a male, phonetically trained native speaker of Cantonese. For *pinjam names*, the first version was the correct pronunciation that involves a *correct application* of pinjam (i.e., with a cTR + cTL tone sequence). The second and third versions were violation conditions wherein the pronunciation of the noun deviated from the correct form in terms of surface tone realization. The second version of *pinjam names* was with an *underapplication* of pinjam (i.e., with a *cTL + cTL combination). The third version of *pinjam names* was a control condition with a *plain violation* of the tone in the first syllable (i.e., with a *cT1 + cTL combination). Likewise, for *non-pinjam names*, the first version was the correct pronunciation that involves a *correct non-application* of pinjam (i.e., with a cTL + cT1 surface tone). The second version of *non-pinjam names* was with an *overapplication* of pinjam (i.e., with a *cTR + cT1 combination). The third version of *non-pinjam names* was a control condition with a *plain violation (2)* of the tone in the first syllable (i.e., with a *cT1 + cT1 combination). The 10 family names were selected such that when the first syllables were realized as a cTR (which overlaps with cT2) or cT1, none of them was permissible family names.

All stimuli recording, production, and acoustic normalization procedures were identical to Experiment 1, ensuring all stimuli were maximally controlled acoustically across conditions. F0 properties of all stimuli are presented in Figure 7.

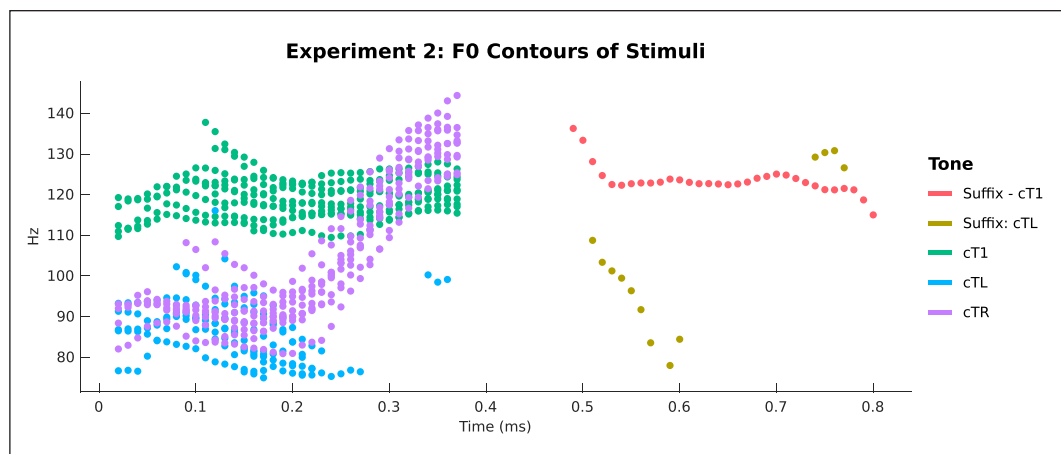


Figure 7: Experiment 2: F0 properties of stimuli. Note that the lack of consistent f0 values detected in portions of cTL corresponds to creakiness typical in the tone.

2.3.3. Design, procedures, and electrophysiological recording parameters

The experimental design, procedures, and electrophysiological recording parameters were identical to those in Experiment 1.

2.3.4. Data and statistical analyses

Procedures of data and statistical analyses were identical to those in Experiment 1, with a design that constitutes a 2 (CONTEXT: pinjam vs. non-pinjam names) \times 2 (VIOLATION TYPE: phonological violations (i.e., over- and underapplications) vs. plain violations) repeated measures design, as summarized in Table 4. Post-hoc analyses on N400 mean amplitude, peak amplitude, and fractional peak latency at the 400–1200 ms time window were also conducted.

2.4. Results

Given the experimental design outlined in Table 4 and the set of predictions outline in Table 3, results revealed inconsistencies across behavioral and ERP differences in the processing of Cantonese pinjam.

2.4.1. Behavioral results

Figure 8 shows the mean behavioral RT all conditions. The linear mixed-effect model (LMM) on RT (see Table 6) revealed main effects of both VIOLATION TYPE ($p < .001$) and CONTEXT ($p < .001$), suggesting a slower RT in phonological violations as compared to plain violations, as well as a slower RT when processing stimulus types sharing a suffix that carries an cT4. The VIOLATION TYPE \times CONTEXT interaction was not significant ($p = .290$).

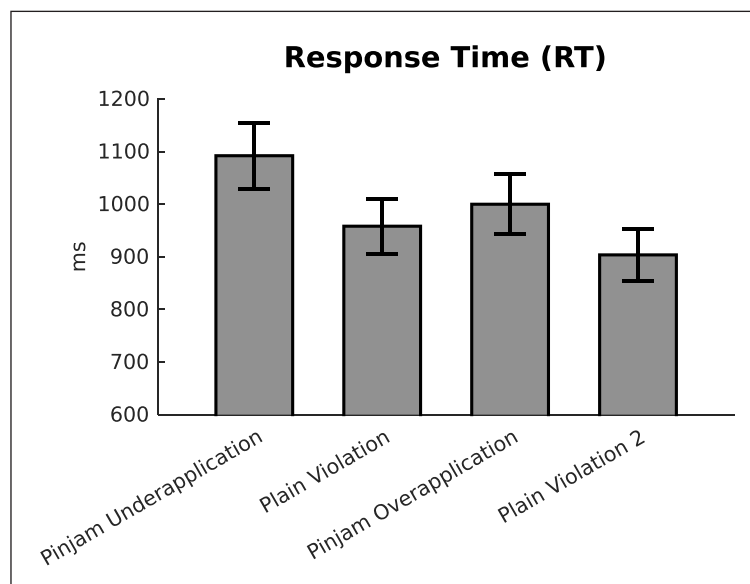


Figure 8: Results: Behavioral response time in Cantonese pinjam judgment. Error bars represent ± 1 standard error of the mean.

2.4.2. ERP results

The grand averaged ERP difference waves and topographical maps of scalp voltage of the difference waves at the two time windows of analysis are presented in **Figure 9**. Linear mixed-effect model (LMM) results are summarized in **Table 6**.

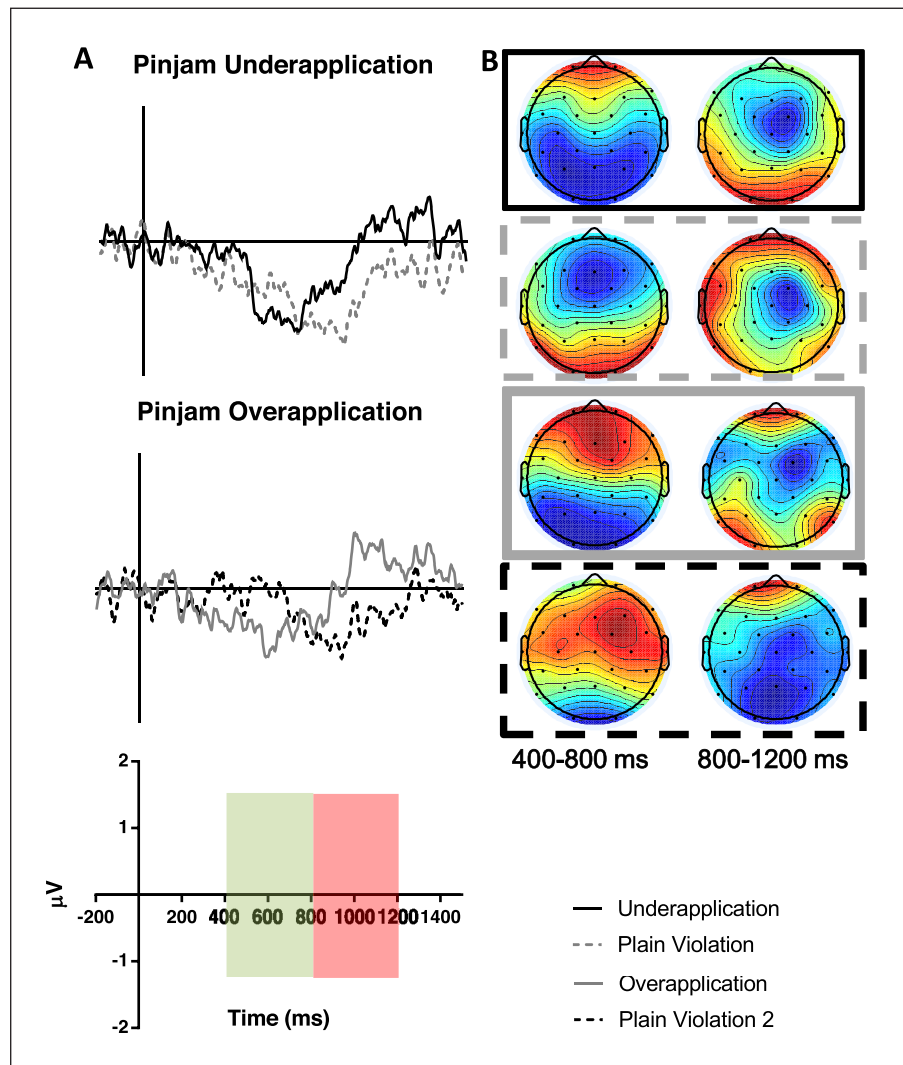


Figure 9: Results: Brain event-related potentials: Panel A presents the ERP difference waves (average of P3, Pz, and P4 channels) associated with *pinjam underapplication* (top) and *overapplication* (bottom), each against the respective plain violation conditions. Panel B presents the topographical maps of scalp voltage (mean amplitude) of the difference waves of all conditions averaged per channel across 400–800 ms (left column) and 800–1200 ms (right column) time windows, of *pinjam underapplication* (1st row), *overapplication* (3rd row), and their associated plain violation conditions (2nd and 4th rows).

For the 400–800 ms window, negative components are observed in all conditions. **Figure 10** (Panels C and D) presents the mean amplitude and fractional peak latency of all conditions in the 400–800 ms window. LMMs on mean amplitude and fractional peak latency of this time window did not reveal any main effect or interaction.

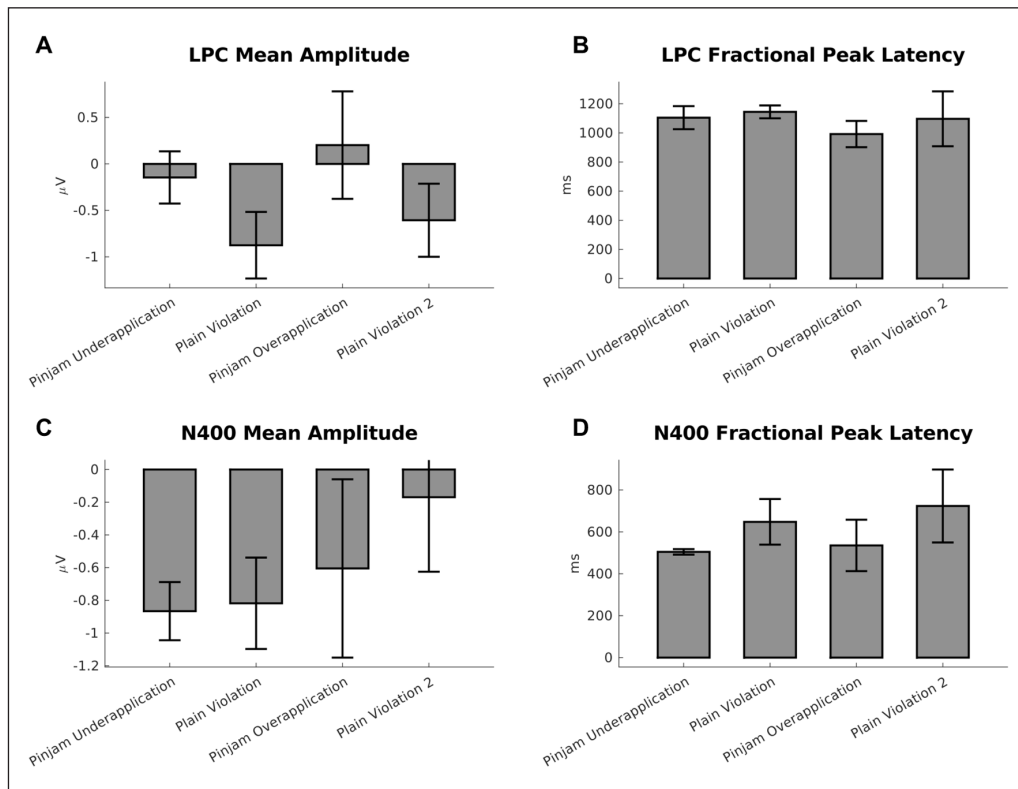


Figure 10: Results: Measurements of mean amplitude (Panels A and C) and fractional peak latency (Panels B and D) of the average of P3, Pz, and P4 channels for all conditions from the 400–800 ms (Panels C and D) and 800–1200 ms (Panels A and B) time windows. Error bars represent ± 1 standard error of the mean with a jackknife method.

For the 800–1200 ms window, no clear components are observed for any condition. **Figure 10** (Panels A and B) presents the mean amplitude of all conditions in the 800–1200 ms window. LMMs on mean amplitude and fractional peak latency of this time window did not reveal any main effect or interaction either.

LMMs in the post-hoc analyses on mean amplitude, peak amplitude, or fractional peak latency of the 400–1200 ms window (**Figure 11**) did not reveal any main effect or interaction. Like Experiment 1, the intercept of the post-hoc LMM on negative peak amplitude (400–1200 ms) was significant (see **Table 6**). Together with the significant intercept of the mean amplitude (400–800 ms) model, results confirm the visual observation of the presence of N400-like negative

components at these time windows that did not vary as a function of CONTEXT or VIOLATION TYPE.

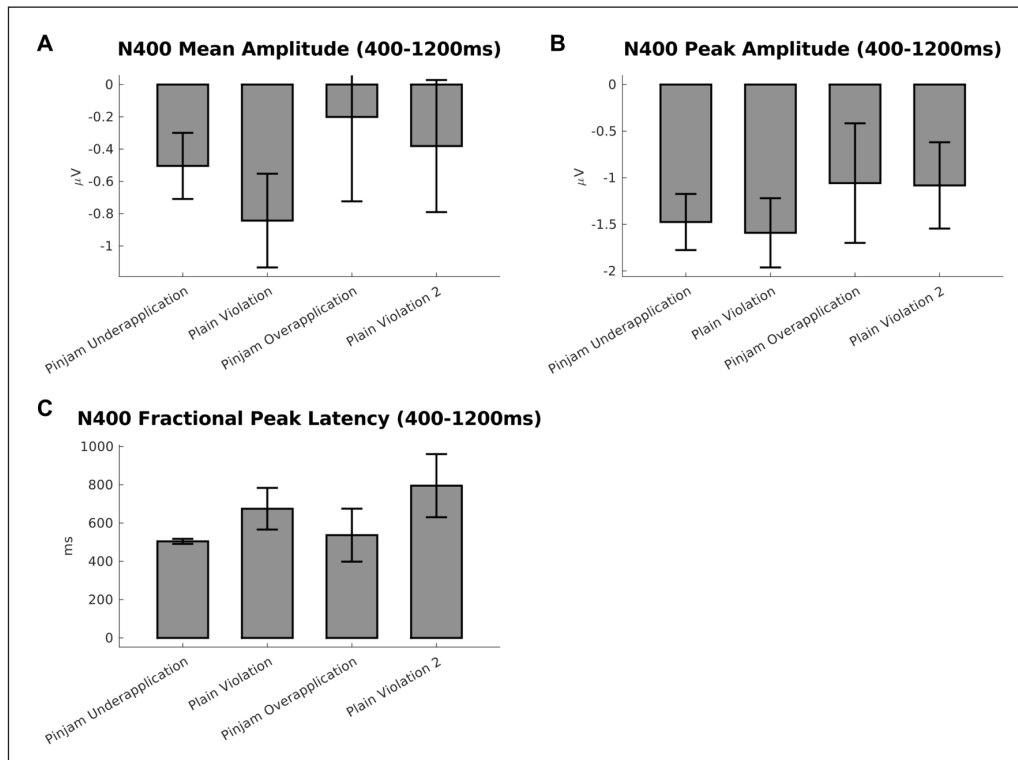


Figure 11: Results: Measurements of mean amplitude (Panel A), peak amplitude (Panel B), and fractional peak latency (Panel C) of the average of P3, Pz, and P4 channels for all conditions from the 400–1200 ms time windows. Error bars represent ± 1 standard error of the mean with a jackknife method.

3. Discussion

In this study, we examined the processing of a tone alternation pattern that is similar in surface, but with distinct lexical distributions across Mandarin and Cantonese, using a cross-modal priming paradigm. Behavioral and ERP responses in priming paradigms reflect abstract properties of mental representations (Jakimik et al., 1985; Seidenberg & Tanenhaus, 1979; Ventura et al., 2004), including phonological representations (Bohn et al., 2013; Domahs et al., 2008, 2013, 2014; Henrich et al., 2014; Molczanow et al., 2013). Results of Experiments 1 and 2, ERP responses in particular, suggest cross-linguistic differences in the processing of a surface-similar lexical tone alternation pattern.

At the behavioral level, Experiments 1 and 2 exhibited largely similar response patterns in the phonological judgment task embedded in the cross-modal priming paradigm. A main effect

of VIOLATION TYPE in both experiments 1 and 2 suggests that rejections of underapplication and overapplications of tone alternation (mT3 sandhi in Experiment 1; pinjam in Experiment 2) were both slower (in terms of RT) than plain violations. This similarity is striking, suggesting that despite their well-established lexical distributional differences and productivity of the alternation phenomenon, Mandarin tone sandhi and Cantonese pinjam appear to be subserved by the same type of phonological representation.

According to our prediction, both tone alternation patterns are subserved by Listed UR, converging with previous behavioral studies using visual-to-visual priming (Nixon et al., 2015) and auditory-to-auditory priming (Chien et al., 2016) to confirm the presence of priming effects among surface variants in a lexical tone alternation. Specifically, these results suggest that the elicitation of surface variants of lexical tone may facilitate behavioral responses involving its alternating surface variants relative to a contrastive tone category. However, the significant VIOLATION TYPE \times CONTEXT interaction identified only in Experiment 1 was not consistent with any of our predictions. Specifically, rejection of tone sandhi overapplications was faster than underapplications, but was different (i.e., slower) than plain violations. Although the behavioral RT results do not align exactly with any set of our predictions, the presence of the VIOLATION TYPE \times CONTEXT interaction for Mandarin sandhi in Experiment 1, and its lack thereof for pinjam in Experiment 2, suggest that for Mandarin sandhi, the underlying mTL is more relevant during the priming process, or that for pinjam, the surface cTR is more relevant.²

ERP results yielded specific patterns consistent with our competing hypotheses, providing a clearer picture. Crucially, notably different ERP patterns were found across Experiments 1 and 2. Experiment 1 found that the processing of both tone sandhi underapplications and overapplications both elicited an earlier N400 response and an additional LPC response compared to plain violations. The N400 and LPC responses in the underapplication and overapplication conditions were comparable (no difference in amplitude or latency). In contrast, it was found in Experiment 2 that both pinjam underapplications and overapplications only elicited N400 but not LPC responses. The amplitude and latency of the N400 responses across all conditions (underapplication, overapplication, and plain violations) were not different. While not fully compatible with the behavioral results (i.e., the lack of ERP differences despite behavioral differences in Experiment 2), these ERP patterns from Experiments 1 and 2 reflected that the processing of similarly surface-patterned violations in tone alternation involved different neurocognitive underpinnings across Mandarin and Cantonese. We interpret such neurocognitive differences to reflect differences in underlying phonological presentations that subserve the tone alternation patterns across the two languages.

² We thank an anonymous reviewer for this interpretation of RT results.

In the cross-modal priming paradigm, a mismatch between representations elicited by the prime and the target is known to elicit the N400 and the Late Positive Complex (LPC) ERP responses, relative to a baseline where the prime and target match with each other (J. E. Anderson & Holcomb, 1995; Holcomb, 1993; Holcomb et al., 2005; Kiyonaga et al., 2007). The N400 is an ERP component identified as a neural-marker for lexico-semantic integration in language processing, most commonly elicited by lexico-semantic violations (Kutas & Federmeier, 2000). The N400 is generated by a cortical semantic processing network involving mid temporal inferior frontal brain areas (Kutas & Federmeier, 2000; E. F. Lau et al., 2008), and peaks negatively at around 300–500 ms post-stimulus. The N400 component is often accompanied by the LPC (Curran et al., 1993; Karayanidis et al., 1991; Woodward et al., 1993), a posterior-distributed ERP component that peaks at a later 400–800 ms post-stimulus time window (Friedman & Johnson, 2000). LPC, generated in the lateral parietal cortex (Rugg & Curran, 2007), reflects the levels of effort or confidence in lexico-semantic judgment (Domahs et al., 2009; Finnigan et al., 2002), lexical “patching” in erroneous contexts (Daltrozzo et al., 2012), and memory recollection (Friedman & Johnson, 2000). In the context of phonological cross-modal priming, a collection of neurolinguistic studies found that given orthographic primes, violations of phonological properties of the auditory target affect lexical integration (as reflected in N400) and restructuring (“patching”) of such phonological violations (as reflected in LPC) (Bohn et al., 2013; Domahs et al., 2008, 2013, 2014; Friedrich et al., 2008; Henrich et al., 2014; Molczanow et al., 2013). Specifically, phonologically ill-formed violations in the target which are semantically recoverable, but require mental phonological restructuring to match with the prime, will elicit the LPC (Domahs et al., 2008). Meanwhile, violations of the phonological form which hinder lexical access, but are otherwise phonologically well-formed, will only elicit the N400 (Domahs et al., 2013).

In Experiment 1, the underapplication condition, as compared to the plain violation control, elicited a robust LPC response that followed an earlier N400 response. This suggests that although the violation form is phonologically ill-formed, hindering lexical access initially, the violation was deemed recoverable phonologically. This prompted a subsequent phonological restructuring process that recovered the correct phonological form to facilitate lexical access, as reflected by the LPC (Domahs et al., 2008). This violation was deemed recoverable phonologically potentially because the underapplication violation contains a low tone, which either matches 1) Single UR form, or 2) one of the listed allomorphs in the Listed UR.

Results of the overapplication condition provide deeper insight on which of the three forms subserve the alternating mT3 in the phonological representation. A priori, it was predicted that if the tone alternation pattern is subserved by a single phonological representation, overapplication violations would only elicit an N400 response comparable to plain violations because the auditory target would not match the form of phonological representation elicited by the prime (i.e., a Low

tone). While an N400 was indeed elicited in the overapplication condition in Experiment 1, this N400 was accompanied by a robust LPC, consistent with the predictions of the Listed UR account, and with previous behavioral and ERP experiments (Y. Chen et al., 2011; Li & Chen, 2015; Nixon et al., 2015). In cross-modal priming studies embedding a phonological judgment task, the LPC is interpreted as an index to the level of effort required to restructure violations in the auditory target to match with the primed lexical item to achieve behavioral judgment (Domahs et al., 2008, 2009, 2013, 2014). In addition, the literature on the bilingual lexicon suggested that the LPC may reflect the amount of cognitive resources required to inhibit coactivated mental representations in language judgment tasks (P. Chen et al., 2017). Therefore, the LPC in the overapplication condition may reflect that all alternation forms listed in the phonological representation were coactivated by the visual prime, regardless of its surface phonological context. Specifically, since mTR and mTD/mTL were all coactivated by the visual prime, the mTR in the overapplication violation can be traced back to its mT3 lexical representation, and was therefore deemed recoverable. Excess mental effort was then needed to restructure the violation into the corrected listed phonological form (i.e., mTL), while inhibiting the wrong alternation form perceived from the auditory prime (i.e., the mTR). The combination of both processes manifested into the LPC.

Importantly, the comparably robust LPC responses in underapplication and overapplication conditions suggest that the LPC was not solely elicited due to the ill-formed mTL+mTL tone sequence in the underapplication condition; LPC was elicited despite the fact that the mTR+mT1 overapplication violation was phonologically acceptable. This implies that the coactivation of all listed forms is not contingent upon its neighboring phonological contexts, such that the mTR can be recovered to match with the mT3 lexical representation even in a mTR+mT1 sequence which is not a phonological context triggering tone sandhi.

While the presence of N400 components was expected in all conditions, their lower latency in the underapplication and overapplication conditions was not expected. Such lower latency of N400 response as compared to the N400 in the plain violation conditions may further speak to the differential source of the N400 components elicited in the cross-modal priming paradigm and the influence of Listed UR on lexical access.

One potential source of the earlier N400 component is a phonotactic violation detection mechanism (henceforth, *phonological N400*) (Zhang et al., 2022), potentially distinct from the conventional N400 component indexing lexical access mechanisms (*lexical N400*). The presence of this putative phonological N400 is supported by a body of ERP studies which identified an N400-like component in the processing of phonotactically illegal sequences, both segmental (Domahs et al., 2009; Ulbrich et al., 2016; White & Chiu, 2017; Wiese et al., 2017) and suprasegmental (Zhang et al., 2022) in nature.

Therefore, the N400 components that differ in latency in Experiment 1 may reflect distinct components: the early N400 component elicited in the overapplication condition (which contained

a phonotactically illegal tone sequence [$*mTL + mTD$]) is phonological in nature, whereas the later N400 component elicited in the plain violation conditions with phonotactically well-formed tone sequences is lexical in nature. Interestingly, the early N400 in the overapplication condition of Experiment 1 was present despite a phonotactically well-formed tone sequence ($mTR + mT1$). Such results may reflect that this putative phonological N400 component not only reflects the detection of phonotactic violations on the surface, but also the detection of ill-derived grammatical transformation from the UR to the SR (i.e., the selection of the $[mTR]$ allomorph in a $/mT3 + mT1/$ sequence despite the lack of the sandhi-trigger phonological context).

Another possibility is that the earlier N400 may indicate more efficient detection of lexical violations. The more efficient detection of lexical violation in the underapplication and overapplication conditions is not contradictory to the LPC responses, which index a higher load of cognitive resources. In the psycholinguistics literature on the lexicon, it is known that coactivated representations exhibit both facilitative and inhibitory effects on language processing: weak coactivation leads to facilitative effects, whereas strong coactivation leads to inhibitory effects (Q. Chen & Mirman, 2012). Hence, it is likely that weak coactivation in earlier stages of processing (i.e., initial lexical access) facilitates the detection of tone alternation violation in the auditory target per se. Yet, later task-relevant stages of processing, which require an explicit and attentive mapping between the primed representation and the auditory target, may have led to stronger coactivation, which requires more cognitive resources to inhibit and restructure.

In contrast, Experiment 2 demonstrated a distinct set of the ERP patterns in the processing of Cantonese pinjam. In both underapplication and overapplication conditions, violations elicited an N400 response were comparable to the plain violation control. This N400 effect is consistent with the Single SR account, suggesting both overapplications and underapplications of pinjam were processed no differently than across-tone category plain violations.

The N400-only results suggest that the ill-formed $*cTL + cT1$ sequence in the overapplication condition was deemed the same as across-tone category plain violations; what seems surprising is that the N400 in the underapplication condition indicates lexical mismatch even though the citation cTL form appears in the underapplication violation (i.e., $*cTL + cTL$).

Traditionally, pinjam has indeed been analyzed in the theoretical literature as a morpho-phonological process that attaches a high tone target to the low tone citation form $cT4$ (Kam, 1977; M. Yip, 2002). The analysis was believed to reflect an ongoing sound change that involves the morphologization of a fossilized tone sandhi process in Cantonese (Yu, 2007). However, as mentioned in the introduction, lexical items that undergo pinjam are limited. Beyond those lexical items, Low + Low tone sequences are not phonologically ill-formed. Importantly, the productivity of pinjam is questionable, while the relationship between the pinjam form and its alleged base is not always apparent to native speakers (Kam, 1977; Yu, 2007), casting doubt on the analysis of pinjam as an active morphological process that is grammatical in nature. Therefore, one

possibility is that, at least for the specific surname + honorific compounds tested in the current study, the disyllabic pinjam forms are lexicalized as cTR + cTL disyllabic sequences and directly stored in the lexicon without further abstracting from the SR. Lexicalization refers to a process of language change that fossilizes alternation patterns originally derived by phonological processes directly into the lexical representations (Brinton & Traugott, 2005). In the process of sound change, morphologization and lexicalization reflect different stages of “dephonologization” of phonological alternations that cause morphophonological processes to cease to be part of the productive phonological system (Bostoen, 2008). Perhaps with the limited lexical distribution of pinjam, through time, certain pinjam lexical items once morphological in nature further undergo lexicalization where the whole pinjam process is degrammaticalized. Specifically, instead of a /cTL/+ /cTL/ disyllabic sequence, our lexical items that seemingly undergo pinjam from their “citation forms” are in fact lexicalized, such that the mental representation takes the form of [cTR + cTL], without undergoing any phonological processes that alter the surface tone. Therefore, any violations to this lexicalized form (e.g., the cTL + cTL sequence in the underapplication condition) are processed no differently than plain violations in terms of lexical access as indexed by N400. Meanwhile, the lexical base forms of the surnames are stored in separate phonological representations as cTL, hence a comparable N400 effect in the overapplication condition in a non-pinjam cT4 + cT4 sequence.

Whereas the current study only focuses on a very specific class of lexical items with pinjam for cross-linguistic matching of surface tone patterns, future studies can utilize a similar experimental approach to examine the neural processing of different lexical classes of pinjam. For example, a cross-sectional investigation across different age groups may provide an insightful model to assess the potential diffusion of lexicalization of pinjam in contemporary Cantonese, and how such potential degrammaticalization of pinjam may be tied to productivity of pinjam as an active grammatical process.

To further understand how other patterns of phonological alternation in the typology are represented, and how specific structures of phonological representations may reflect stages of sound change, further studies can also examine the neural processing of tone alternation in languages like Southern Min where the connection between the alternating tones is clear and recognized by speakers (unlike pinjam), but the alternation itself is not productive (unlike Mandarin tone sandhi) (e.g., Chien et al., 2017; Zhang et al., 2011).³ Another question is how alternations of sounds of different nature (e.g., segments and tones) dynamically interact to shape the phonological representation neurocognitively.

While the current study on the neural processing of phonological alternation invites various further topics of inquiry, the current cross-linguistic results provide strong support for the notion

³ We thank an anonymous reviewer for this suggestion.

that phonological representations postulated in linguistic theory are not only formal entities for linguistic analysis, but are neurocognitive in nature, active in modulating phonological processing, speech production, and lexical access. Specifically, our distinct cross-linguistic ERP patterns identified in a set of surface-similar, but distributionally distinct phonological alternation phenomena highlight that phonological representations with different structures and levels of abstractedness (listed allomorphy versus lexicalized SR) may manifest into distinct neurocognitive processes during speech processing.

3.1. Implications for the architecture and learnability of phonological representations

ERP patterns from Experiments 1 and 2 demonstrate cross-linguistic differences in the phonological representations subserving a surface-similar tone alternation pattern. These results suggest that even with a phonological alternation pattern that manifests similarly across languages on the surface, crucial differences in underlying properties such as lexical distribution can lead to a different structure of the phonological representation that subserves the alternation.

Our results hence support the notion proposed in some variants of generative theories (Mascaró, 2007; Pierrehumbert, 2016; Prince & Smolensky, 2004) that phonological representations can take different forms—from more abstract associations among allomorphs (Archangeli & Pulleyblank, 2014; Mascaró, 2007), to more concrete representations of lexically-restricted, non-productive alternation patterns as separate, surface-true representations (Bybee, 2002; Pierrehumbert, 2016). The final shapes of different phonological representations may be determined by how economical it is to transform the representations so that they accurately yield the surface phonological patterns (Prince & Smolensky, 2004; M. Yip, 1996). While the present study does not directly provide neurocognitive evidence supporting the presence of phonological representations as abstract as the Single UR account postulated in traditional generative theories (Chomsky & Halle, 1968), our cross-linguistic findings suggest that alternation patterns are at least represented in the mind at multiple levels of abstractedness as a function of lexical distributional properties.

From a learnability perspective, this multiform nature of phonological representations entails that infants and children are able to analyze distributional (e.g., a morpheme is pronounced differently when combined with another morpheme) and statistical patterns (e.g., whether a phonological pattern is pervasive across all sounds or just limited to specific words) of speech sounds while forming phonological representations through language acquisition. Infants indeed have demonstrated the ability to categorize (Perszyk & Waxman, 2018), attend to stimulus statistics (Saffran & Kirkham, 2018), and identify distributional properties of sensory signals (Seidl & Cristia, 2012), all of which are abilities demonstrated in young infancy to support learning across various aspects of cognition. Therefore, in principle, the complex and multiform

nature of phonological representations may indeed be learnable, also supported by the Emergent Phonology hypothesis that phonological learning and competence are supported by basic domain-general cognitive abilities (Archangeli & Pulleyblank, 2014). Future theoretical linguistic and experimental work may shed light on how phonological representations are learned over the course of infant language acquisition, specifically on how distinct forms of phonological representation (e.g., listed allomorphy versus lexicalization) take shape neurocognitively.

4. Limitations

Several potential limitations should be considered and addressed in future work. To maintain maximal experimental control at the phonological level, our stimuli carried surname + honorific combination which carry little semantic meaning. It is unclear how the form of phonological representations may interact with lexical factors such as meaning and lexical frequency. Future experiments utilizing a larger variety of stimuli in different languages discourses may strengthen the understanding of the multiform nature of phonological representations.

Another potential limitation lies in the specificity of the cross-modal priming paradigm in specifically tapping abstract phonological representations (e.g., the UR) versus more concrete surface representations and acoustic factors. Prior studies (Friedrich et al., 2008; Lahiri & Reetz, 2010) have suggested that an abstract level of representation corresponding to the UR is tapped in priming paradigms, especially considering the later timescale of processing (>400 ms) where higher-level lexical-access is hypothesized to take place (E. F. Lau et al., 2008), while the extraction of lower-level phonological features takes place as early as 100 ms as reflected in the N100 component (Meng, Kotzor, et al., 2021; Obleser et al., 2004). However, it remains unclear the extent to which SR-level representations are tapped in the N400 time range *in addition to* the more abstract UR level, considering recent evidence suggesting both UR and SR are encoded at different stages of tone sandhi word production (X. Chen et al., 2022). Indeed, the experimental results of both of our studies implicate that both types of tone alternation are subserved by different organizations of concrete representations (listed allomorphy and lexicalization), but not the most abstract Single UR account (although importantly, results suggest that listed URs in Mandarin T3 Sandhi are more abstract than Single SRs in Cantonese pinjam). Future studies where more stringent acoustic control can be implemented (e.g., with synthesized stimuli) and with alternation patterns that may be represented by very abstract Single URs (e.g., Turkish final devoicing (Inkelas, 1995; Kager, 2008)) may better tease these factors apart.

The behavioral responses were also puzzling, with specific patterns not predicted from any of the competing hypotheses. The behavioral results were also incompatible with the ERP results, considering there was a lack of ERP differences across conditions in Experiment 2 despite behavioral differences. One potential explanation was that compared to ERP, the abstract phonological representations were tapped to a lesser extent in behavioral manifestations

of priming, reflecting levels of downstream processing that were not tapped by the N400 and LPC components. As such, as compared to ERP components which are more specifically tied to lexical and phonological processes, behavioral responses are interlaced with factors beyond phonological representations, e.g., meaning and lexical frequency. Alternatively, this may reflect the two types of tone alternation patterns being processed differentially at the neural level to arrive at a similar behavioral pattern (i.e., priming effect) cross-linguistically.

Another possibility is that the ERP differences in Experiment 2 may be potentially minute, and therefore not detectable statistically given our sample size. Indeed, our ERP analysis was implemented with a jackknife approach to reduce potential Type I and II errors, which may have limited the sensitivity of our statistical approach. Future, larger-scale studies with a more comprehensive set of stimuli, a larger number of trials, and a larger participant sample may be able to better tease apart the relationship between behavioral and ERP responses in the cross-modal priming paradigm more conclusively.

From a methodological perspective, the inconsistent results may further suggest that different experimental methods and paradigms may tap into different levels of phonological representations. Indeed, laboratory phonology studies examining the UR structure of Mandarin tone sandhi comprise a large variety of methodologies across modalities, targeting speech perception (Chien et al., 2016; J. C. Lau et al., 2019; Li & Chen, 2015; Politzer-Ahles et al., 2016; Zeng et al., 2021) versus production (X. Chen et al., 2022; Zhang & Lai, 2010; Zhang et al., 2022), at neural (X. Chen et al., 2022; J. C. Lau et al., 2019; Li & Chen, 2015; Politzer-Ahles et al., 2016; Zeng et al., 2021; Zhang et al., 2022) versus behavioral levels (Y. Chen et al., 2011; Chien et al., 2016; Meng, Wynne, & Lahiri, 2021; Zhang & Lai, 2010), and at single-syllables (J. C. Lau et al., 2019; Li & Chen, 2015; Politzer-Ahles et al., 2016) versus larger linguistic constituents (X. Chen et al., 2022; Y. Chen et al., 2011; Chien et al., 2016; Meng, Wynne, & Lahiri, 2021; Nixon et al., 2015; Zeng et al., 2021; Zhang et al., 2022). Therefore, experimental methods, especially to the extent that they provide time-sensitive information on lexical access and production processing, may determine what level of representations are relevant. Indeed, phonological representations modulate speech and language processing tasks widely across system levels of the brain (Hickok & Poeppel, 2007). Phonological representations are tapped into by neurophysiological pathways that subserve speech and language functions across modalities (Hickok, 2012). For example, phonological representations impact speech processing as fundamentally as by actively modulating sensory encoding of auditory signals at the subcortical auditory system (J. C. Lau et al., 2019), through interactive effects between neuronal adaptation mechanisms and top-down predictive tuning through the corticofugal pathway (J. C. Lau et al., 2017). One possibility is that neural pathways across system levels and modalities may invoke different levels of the multiform phonological representations (Pierrehumbert, 2016) during speech- and language-related processing. Future experiments should consider employing a multimethod approach

performed on the same cohort of participants and the same set of stimuli, not only to unify the mixed results in lexical tone alternation, but also to provide methodological insight into how different levels of the phonological representation can be more precisely tapped in laboratory phonology inquiries.

Another limitation is this study's assumption that the phonologies of individuals speaking the same language are uniform. However, an abundance of experimental evidence on individual differences in speech sound processing (e.g., Bones & Wong, 2017; Deng et al., 2018; Maggu et al., 2021; Maggu et al., 2018) implies that phonological representations and grammars may differ even among individuals speaking the same language. Although the current study employs a within-subject design where individual differences are not a confounder, the extent to which individual differences at processing, representational, and grammatical levels may have impacted our experimental results are not clear. Individual differences in speech sound representations and processing are critical as metrics of learning (Antoniou & Wong, 2015; Ingvalson et al., 2013; Wong et al., 2017) and indices of clinical phenotypes (J. C. Lau et al., 2021, 2022; Liu et al., 2014), with implications to the genetic bases of language (Wong et al., 2012, 2017, 2020). The issue of individual differences in phonological representations and grammars in formal phonological theory should therefore be more directly addressed in future studies.

5. Conclusion

In summary, the present study identified different ERP patterns in the neural processing of a lexical tone alternation similar on the surface in two languages, but with crucial differences in lexico-phonological distribution across languages. We interpret the cross-linguistic patterns as indicative of the multiform nature of phonological representations that can be shaped with different levels of abstractedness, by taking into account factors including lexical and phonological distributions. By showing neurocognitive evidence for a multiform architecture of phonological representations, the results invite further theoretical and empirical investigations to further advance the current understandings of the role of phonological derivation in language production, and the acquisition of the multiform nature of phonological representations.

Abbreviations

The following abbreviations were used in general:

UR underlying representation

SR surface representation

F0 fundamental frequency

LMM linear mixed effect mode

The following abbreviations were used in experiments:

EEG electroencephalography

ERP event related potential

LPC late positive complex

SPL sound pressure level

RT response time

Ag/AgCl silver/silver chloride

ICA independent component analysis

The following abbreviations were used to describe tones:

mT1 Mandarin second tone

mT2 Mandarin second tone

mT3 Mandarin third tone

mTR Mandarin rising tone

mTD Mandarin dipping tone

mTL Mandarin low-falling tone

cT1 Cantonese first tone

cT4 Cantonese fourth tone

cTR Cantonese rising tone

cTL Cantonese low-falling tone

Supplementary Materials

File1: A PDF file (SupplementaryMaterials.pdf) containing supplementary materials. DOI: <https://doi.org/10.16995/labphon.10293.s1>.

Anonymized data and statistical analytic scripts are available at <https://osf.io/r43de/>.

Ethics and consent

Informed consents approved by The Joint Chinese University of Hong Kong - New Territories East Cluster Clinical Research Ethics Committee were obtained from each human subject before any experimental procedure.

Acknowledgements

The authors would like to thank René Kager, Peggy Mok, Regine Lai, and all members of the Laboratory for Language, Learning, and the Brain at The Chinese University of Hong Kong for their assistance and comments on this study. We also thank The Chinese University of Hong Kong (CUHK) – Utrecht University (UU) Joint Centre for Language, Mind and Brain for its assistance in this research. We thank Mitra Kumareswaran and Lindsay Goldman for proofreading the manuscript.

Competing interests

The authors have no competing interests to declare.

Author's contributions

J.C.Y.L. designed the study, performed research, analyzed data, and wrote the paper; P.C.M.W. funded this study via discretionary funding; P.C.M.W. edited the paper and provided general comments on the study.

References

- Alderete, J., Chan, Q., & Tanaka, S.-i. (2022). The morphology of Cantonese “changed tone”: Extensions and limitations. *GENGO KENKYU (Journal of the Linguistic Society of Japan)*, 161, 139–169.
- Anderson, J., & Jones, C. (1974). Three theses concerning phonological representations. *Journal of linguistics*, 10(1), 1–26. DOI: <https://doi.org/10.1017/S0022226700003972>
- Anderson, J. E., & Holcomb, P. J. (1995). Auditory and visual semantic priming using different stimulus onset asynchronies: An event-related brain potential study. *Psychophysiology*, 32(2), 177–190. DOI: <https://doi.org/10.1111/j.1469-8986.1995.tb03310.x>
- Antoniou, M., & Wong, P. (2015). Poor phonetic perceivers are affected by cognitive load when resolving talker variability. *The Journal of the Acoustical Society of America*, 138(2), 571–574. DOI: <https://doi.org/10.1121/1.4923362>
- Anttila, A. (2002). Morphologically conditioned phonological alternations. *Natural Language & Linguistic Theory*, 20(1), 1–42. DOI: <https://doi.org/10.1023/A:1014245408622>

- Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, 5(2), 183–207. DOI: <https://doi.org/10.1017/S0952675700002268>
- Archangeli, D., & Pulleyblank, D. (2014). Phonology as an emergent system. *The Routledge Handbook of Phonological Theory*. London: Routledge.
- Bohn, K., Knaus, J., Wiese, R., & Domahs, U. (2013). The influence of rhythmic (ir) regularities on speech processing: Evidence from an ERP study on German phrases. *Neuropsychologia*, 51(4), 760–771. DOI: <https://doi.org/10.1016/j.neuropsychologia.2013.01.006>
- Bones, O., & Wong, P. C. (2017). Congenital amusics use a secondary pitch mechanism to identify lexical tones. *Neuropsychologia*, 104, 48–53. DOI: <https://doi.org/10.1016/j.neuropsychologia.2017.08.004>
- Bostoen, K. (2008). Bantu spirantization: Morphologization, lexicalization and historical classification. *Diachronica*, 25(3), 299–356. DOI: <https://doi.org/10.1075/dia.25.3.02bos>
- Brinton, L. J., & Traugott, E. C. (2005). *Lexicalization and language change*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511615962>
- Bybee, J. (2002). Phonological evidence for exemplar storage of multiword sequences. *Studies in second language acquisition*, 24(2), 215–221. DOI: <https://doi.org/10.1017/S0272263102002061>
- Chen, P., Bobb, S. C., Hoshino, N., & Marian, V. (2017). Neural signatures of language co-activation and control in bilingual spoken word comprehension. *Brain Research*, 1665, 50–64. DOI: <https://doi.org/10.1016/j.brainres.2017.03.023>
- Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*, 119(2), 417. DOI: <https://doi.org/10.1037/a0027175>
- 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, 108367. DOI: <https://doi.org/10.1016/j.neuropsychologia.2022.108367>
- Chen, Y., Shen, R., & Schiller, N. O. (2011). Representation of allophonic tone sandhi variants. *Proceedings of Psycholinguistics Representation of Tone. Satellite Workshop to ICPhS, Hongkong*, 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. DOI: <https://doi.org/10.1080/23273798.2015.1064976>
- Chien, Y.-F., Sereno, J. A., & Zhang, J. (2017). What's in a word: Observing the contribution of underlying and surface representations. *Language and Speech*, 60(4), 643–657. DOI: <https://doi.org/10.1177/0023830917690419>
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.
- Curran, T., Tucker, D. M., Kutas, M., & Posner, M. I. (1993). Topography of the N400: Brain electrical activity reflecting semantic expectancy. *Electroencephalography and Clinical Neurophysiology/ Evoked Potentials Section*, 88(3), 188–209. DOI: [https://doi.org/10.1016/0168-5597\(93\)90004-9](https://doi.org/10.1016/0168-5597(93)90004-9)

- Daltrozzo, J., Wioland, N., & Kotchoubey, B. (2012). The N400 and Late Positive Complex (LPC) effects reflect controlled rather than automatic mechanisms of sentence processing. *Brain Sciences*, 2(3), 267–297. DOI: <https://doi.org/10.3390/brainsci2030267>
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1), 9–21. DOI: <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Deng, Z., Chandrasekaran, B., Wang, S., & Wong, P. C. (2018). Training-induced brain activation and functional connectivity differentiate multi-talker and single-talker speech training. *Neurobiology of Learning and Memory*, 151, 1–9. DOI: <https://doi.org/10.1016/j.nlm.2018.03.009>
- Domahs, U., Genc, S., Knaus, J., Wiese, R., & Kabak, B. (2013). Processing (un-) predictable word stress: ERP evidence from Turkish. *Language and Cognitive Processes*, 28(3), 335–354. DOI: <https://doi.org/10.1080/01690965.2011.634590>
- Domahs, U., Kehrein, W., Knaus, J., Wiese, R., & Schlesewsky, M. (2009). Event-related potentials reflecting the processing of phonological constraint violations. *Language and Speech*, 52(4), 415–435. DOI: <https://doi.org/10.1177/0023830909336581>
- Domahs, U., Knaus, J. A., El Shanawany, H., & Wiese, R. (2014). The role of predictability and structure in word stress processing: An ERP study on Cairene Arabic and a cross-linguistic comparison. *Frontiers in Psychology*, 5. DOI: <https://doi.org/10.3389/fpsyg.2014.01151>
- Domahs, U., Wiese, R., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). The processing of German word stress: Evidence for the prosodic hierarchy. *Phonology*, 25(1), 1–36. DOI: <https://doi.org/10.1017/S0952675708001383>
- Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of cognitive neuroscience*, 16(4), 577–583. DOI: <https://doi.org/10.1162/089892904323057308>
- Finnigan, S., Humphreys, M. S., Dennis, S., & Geffen, G. (2002). ERP ‘old/new’ effects: Memory strength and decisional factor (s). *Neuropsychologia*, 40(13), 2288–2304. DOI: [https://doi.org/10.1016/S0028-3932\(02\)00113-6](https://doi.org/10.1016/S0028-3932(02)00113-6)
- Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1–2), 3–71. DOI: [https://doi.org/10.1016/0010-0277\(88\)90031-5](https://doi.org/10.1016/0010-0277(88)90031-5)
- Friedman, D., & Johnson, R. (2000). Event-related potential (ERP) studies of memory encoding and retrieval: A selective review. *Microscopy research and technique*, 51(1), 6–28. DOI: [https://doi.org/10.1002/1097-0029\(20001001\)51:1 <::AID-JEMT2>3.0.CO;2-R](https://doi.org/10.1002/1097-0029(20001001)51:1 <::AID-JEMT2>3.0.CO;2-R)
- Friedrich, C. K., Lahiri, A., & Eulitz, C. (2008). Neurophysiological evidence for underspecified lexical representations: Asymmetries with word initial variations. *Journal of Experimental Psychology: Human Perception and Performance*, 34(6), 1545. DOI: <https://doi.org/10.1037/a0012481>
- Henrich, K., Alter, K., Wiese, R., & Domahs, U. (2014). The relevance of rhythmical alternation in language processing: An ERP study on English compounds. *Brain and Language*, 136, 19–30. DOI: <https://doi.org/10.1016/j.bandl.2014.07.003>

- Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature reviews neuroscience*, 13(2), 135–145. DOI: <https://doi.org/10.1038/nrn3158>
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature reviews neuroscience*, 8(5), 393–402. DOI: <https://doi.org/10.1038/nrn2113>
- Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, 30(1), 47–61. DOI: <https://doi.org/10.1111/j.1469-8986.1993.tb03204.x>
- Holcomb, P. J., Anderson, J., & Grainger, J. (2005). An electrophysiological study of cross-modal repetition priming. *Psychophysiology*, 42(5), 493–507. DOI: <https://doi.org/10.1111/j.1469-8986.2005.00348.x>
- Honbolygó, F., & Csépe, V. (2013). Saliency or template? ERP evidence for long-term representation of word stress. *International Journal of Psychophysiology*, 87(2), 165–172. DOI: <https://doi.org/10.1016/j.ijpsycho.2012.12.005>
- Ingvalson, E. M., Barr, A. M., & Wong, P. C. (2013). Poorer phonetic perceivers show greater benefit in phonetic-phonological speech learning. *Journal of Speech, Language, and Hearing Research*, 56, 1045–50. DOI: [https://doi.org/10.1044/1092-4388\(2012/12-0024](https://doi.org/10.1044/1092-4388(2012/12-0024)
- Inkelas, S. (1995). The consequences of optimization for underspecification. In J. Beckman (Ed.), *Proceedings of the 26th annual meeting of the north east linguistic society* (pp. 287–302). Amherst, Mass: GLSA.
- Jackendoff, R. S. (1995). *Languages of the mind: Essays on mental representation*. MIT Press.
- Jakimik, J., Cole, R. A., & Rudnicky, A. I. (1985). Sound and spelling in spoken word recognition. *Journal of Memory and Language*, 24(2), 165–178. DOI: [https://doi.org/10.1016/0749-596X\(85\)90022-1](https://doi.org/10.1016/0749-596X(85)90022-1)
- Johnson, K. (2007). Decisions and mechanisms in exemplar-based phonology. *Experimental approaches to phonology*, 25–40. DOI: <https://doi.org/10.1093/oso/9780199296675.003.0003>
- Jones, D. (1957). The history and meaning of the term “phoneme”. *Le maître phonétique*, 35, 1–20.
- Juottonen, K., Revonsuo, A., & Lang, H. (1996). Dissimilar age influences on two ERP waveforms (LPC and N400) reflecting semantic context effect. *Cognitive Brain Research*, 4(2), 99–107. DOI: [https://doi.org/10.1016/0926-6410\(96\)00022-5](https://doi.org/10.1016/0926-6410(96)00022-5)
- Kager, R. (2008). Lexical irregularity and the typology of contrast. In K. Hanson & S. Inkelas (Eds.), *The nature of the word: Studies in honor of Paul Kiparsky*. Cambridge, Massachusetts: MIT Press. DOI: <https://doi.org/10.7551/mitpress/9780262083799.003.0017>
- Kam, T. H. (1977). Derivation by tone change in cantonese: A preliminary survey. *Journal of Chinese Linguistics*, 186–210.
- Karayanidis, F., Andrews, S., Ward, P. B., & McConaghy, N. (1991). Effects of inter-item lag on word repetition: An event-related potential study. *Psychophysiology*, 28(3), 307–318. DOI: <https://doi.org/10.1111/j.1469-8986.1991.tb02200.x>

- Kiyonaga, K., Grainger, J., Midgley, K., & Holcomb, P. J. (2007). Masked cross-modal repetition priming: An event-related potential investigation. *Language and Cognitive Processes*, 22(3), 337–376. DOI: <https://doi.org/10.1080/01690960600652471>
- Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & psychophysics*, 50(2), 93–107. DOI: <https://doi.org/10.3758/BF03212211>
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 979–1000. DOI: <https://doi.org/10.1098/rstb.2007.2154>
- Kuo, W.-J., Yeh, T.-C., Lee, J.-R., Chen, L.-F., Lee, P.-L., Chen, S.-S., Ho, L.-T., Hung, D. L., Tzeng, O. J.-L., & Hsieh, J.-C. (2004). Orthographic and phonological processing of Chinese characters: An fMRI study. *Neuroimage*, 21(4), 1721–1731. DOI: <https://doi.org/10.1016/j.neuroimage.2003.12.007>
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4(12), 463–470. DOI: [https://doi.org/10.1016/S1364-6613\(00\)01560-6](https://doi.org/10.1016/S1364-6613(00)01560-6)
- Lahiri, A., & Reetz, H. (2002). Underspecified recognition. *Laboratory phonology*, 7. DOI: <https://doi.org/10.1515/9783110197105.2.637>
- Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, 38. DOI: <https://doi.org/10.1016/j.wocn.2010.01.002>
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, 9(12), 920–933. DOI: <https://doi.org/10.1038/nrn2532>
- Lau, J. C., Patel, S., Kang, X., Nayar, K., Martin, G. E., Choy, J., Wong, P. C., & Losh, M. (2022). Cross-linguistic patterns of speech prosodic differences in autism: A machine learning study. *PLoS one*, 17(6), e0269637. DOI: <https://doi.org/10.1371/journal.pone.0269637>
- Lau, J. C., To, C. K., Kwan, J. S., Kang, X., Losh, M., & Wong, P. C. (2021). Lifelong tone language experience does not eliminate deficits in neural encoding of pitch in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 51, 3291–3310. DOI: <https://doi.org/10.1007/s10803-020-04796-7>
- Lau, J. C., Wong, P., & Chandrasekaran, B. (2019). Interactive effects of linguistic abstraction and stimulus statistics in the online modulation of neural speech encoding. *Attention, Perception, & Psychophysics*, 81(4), 1020–1033. DOI: <https://doi.org/10.3758/s13414-018-1621-9>
- Lau, J. C., Wong, P. C., & Chandrasekaran, B. (2017). Context-dependent plasticity in the subcortical encoding of linguistic pitch patterns. *Journal of neurophysiology*, 117(2), 594–603. DOI: <https://doi.org/10.1152/jn.00656.2016>
- Li, X., & Chen, Y. (2015). Representation and processing of lexical tone and tonal variants: Evidence from the mismatch negativity. *PLOS One*, 10(12), e0143097. DOI: <https://doi.org/10.1371/journal.pone.0143097>

- Liang, B., & Du, Y. (2018). The functional neuroanatomy of lexical tone perception: An activation likelihood estimation meta-analysis. *Frontiers in neuroscience*, *12*, 495. DOI: <https://doi.org/10.3389/fnins.2018.00495>
- Liu, F., Maggu, A. R., Lau, J. C., & Wong, P. C. (2014). Brainstem encoding of speech and musical stimuli in congenital amusia: Evidence from Cantonese speakers. *Front Hum Neurosci*, *8*. DOI: <https://doi.org/10.3389/fnhum.2014.01029>
- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, 213. DOI: <https://doi.org/10.3389/fnhum.2014.00213>
- Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.
- Maggu, A. R., Lau, J. C., Wayne, M. M., & Wong, P. C. (2021). Combination of absolute pitch and tone language experience enhances lexical tone perception. *Scientific reports*, *11*(1), 1485. DOI: <https://doi.org/10.1038/s41598-020-80260-x>
- Maggu, A. R., Wong, P. C., Antoniou, M., Bones, O., Liu, H., & Wong, F. C. (2018). Effects of combination of linguistic and musical pitch experience on subcortical pitch encoding. *Journal of Neurolinguistics*, *47*, 145–155. DOI: <https://doi.org/10.1016/j.jneuroling.2018.05.003>
- Mascaró, J. (2007). External allomorphy and lexical representation. *Linguistic Inquiry*, *38*(4), 715–735. DOI: <https://doi.org/10.1162/ling.2007.38.4.715>
- McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive science*, *30*(6), 1113–1126. DOI: https://doi.org/10.1207/s15516709cog0000_79
- Meng, Y., Kotzor, S., Xu, C., Wynne, H. S. Z., & Lahiri, A. (2021). Asymmetric influence of vocalic context on Mandarin sibilants: Evidence from ERP studies. *Frontiers in Human Neuroscience*, *15*, 617318. DOI: <https://doi.org/10.3389/fnhum.2021.617318>
- Meng, Y., Wynne, H., & Lahiri, A. (2021). Representation of “T3 sandhi” in mandarin: Significance of context. *Language, Cognition and Neuroscience*, *36*(6), 791–808. DOI: <https://doi.org/10.1080/23273798.2021.1893769>
- Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*(6174), 1006–1010. DOI: <https://doi.org/10.1126/science.1245994>
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227. DOI: <https://doi.org/10.1037/h0031564>
- Miller, G. A. (1990). The place of language in a scientific psychology. *Psychological Science*, *1*(1), 7–14. DOI: <https://doi.org/10.1111/j.1467-9280.1990.tb00059.x>
- Miller, J., Patterson, T., & Ulrich, R. (1998). Jackknife-based method for measuring LRP onset latency differences. *Psychophysiology*, *35*(1), 99–115. DOI: <https://doi.org/10.1111/1469-8986.3510099>
- Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, *48*(2), 229–240. DOI: <https://doi.org/10.1111/j.1469-8986.2010.01061.x>

- Molczanow, J., Domahs, U., Knaus, J., & Wiese, R. (2013). The lexical representation of word stress in Russian: Evidence from event-related potentials. *The Mental Lexicon*, 8(2), 164–194. DOI: <https://doi.org/10.1075/ml.8.2.03mol>
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385, 432–434. DOI: <https://doi.org/10.1038/385432a0>
- Nixon, J. S., Chen, Y., & Schiller, N. O. (2015). Multi-level processing of phonetic variants in speech production and visual word processing: Evidence from Mandarin lexical tones. *Language, Cognition and Neuroscience*, 30(5), 491–505. DOI: <https://doi.org/10.1080/23273798.2014.942326>
- Obleser, J., Lahiri, A., & Eulitz, C. (2004). Magnetic brain response mirrors extraction of phonological features from spoken vowels. *Journal of Cognitive Neuroscience*, 16(1), 31–39. DOI: <https://doi.org/10.1162/089892904322755539>
- Perszyk, D. R., & Waxman, S. R. (2018). Linking language and cognition in infancy. *Annual review of psychology*, 69, 231. DOI: <https://doi.org/10.1146/annurev-psych-122216-011701>
- Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. *Typological studies in language*, 45, 137–158. DOI: <https://doi.org/10.1075/tsl.45.08pie>
- Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, 2(1). DOI: <https://doi.org/10.1146/annurev-linguistics-030514-125050>
- Politzer-Ahles, S., Schluter, K., Wu, K., & Almeida, D. (2016). Asymmetries in the perception of Mandarin tones: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance*, 42(10), 1547. DOI: <https://doi.org/10.1037/xhp0000242>
- Prince, A., & Smolensky, P. (2004). *Optimality theory: Constraint interaction in generative grammar*. John Wiley & Sons. DOI: <https://doi.org/10.1002/9780470759400>
- Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, 11(6), 251–257. DOI: <https://doi.org/10.1016/j.tics.2007.04.004>
- Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual review of psychology*, 69, 181. DOI: <https://doi.org/10.1146/annurev-psych-122216-011805>
- Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory*, 5(6), 546. DOI: <https://doi.org/10.1037//0278-7393.5.6.546>
- Seidl, A., & Cristia, A. (2012). Infants' learning of phonological status. *Frontiers in psychology*, 3, 448. DOI: <https://doi.org/10.3389/fpsyg.2012.00448>
- Tan, L. H., Laird, A. R., Li, K., & Fox, P. T. (2005). Neuroanatomical correlates of phonological processing of Chinese characters and alphabetic words: A meta-analysis. *Human brain mapping*, 25(1), 83–91. DOI: <https://doi.org/10.1002/hbm.20134>
- Thagard, P. (2005). *Mind: Introduction to cognitive science*. MIT press.
- Trubetzkoy, N. (1939). *Grundzüge der phonologie* [Christiane Baltaxe, trans. *Principles of Phonology*. Berkeley: University of California Press. 1969.]. Travaux du cercle linguistique de Prague 7.

- Ulbrich, C., Alday, P. M., Knaus, J., Orzechowska, P., & Wiese, R. (2016). The role of phonotactic principles in language processing. *Language, Cognition and Neuroscience*, 31(5), 662–682. DOI: <https://doi.org/10.1080/23273798.2015.1136427>
- Ulrich, R., & Miller, J. (2001). Using the jackknife-based scoring method for measuring LRP onset effects in factorial designs. *Psychophysiology*, 38(5), 816–827. DOI: <https://doi.org/10.1111/1469-8986.3850816>
- Ventura, P., Morais, J., Pattamadilok, C., & Kolinsky, R. (2004). The locus of the orthographic consistency effect in auditory word recognition. *Language and Cognitive processes*, 19(1), 57–95. DOI: <https://doi.org/10.1080/01690960344000134>
- White, J., & Chiu, F. (2017). Disentangling phonological well-formedness and attestedness: An ERP study of onset clusters in English. *Acta Linguistica Academica*, 64(4), 513–537. DOI: <https://doi.org/10.1556/2062.2017.64.4.2>
- Wiese, R., Orzechowska, P., Alday, P. M., & Ulbrich, C. (2017). Structural principles or frequency of use? an ERP experiment on the learnability of consonant clusters. *Frontiers in Psychology*, 7, 2005. DOI: <https://doi.org/10.3389/fpsyg.2016.02005>
- Wong, P. C., Chandrasekaran, B., & Zheng, J. (2012). The derived allele of *ASPM* is associated with lexical tone perception. *PloS one*, 7(4), e34243. DOI: <https://doi.org/10.1371/journal.pone.0034243>
- Wong, P. C., Kang, X., Wong, K. H., So, H.-C., Choy, K. W., & Geng, X. (2020). *ASPM*-lexical tone association in speakers of a tone language: Direct evidence for the genetic-biasing hypothesis of language evolution. *Science Advances*, 6(22), eaba5090. DOI: <https://doi.org/10.1126/sciadv.aba5090>
- Wong, P. C., Vuong, L. C., & Liu, K. (2017). Personalized learning: From neurogenetics of behaviors to designing optimal language training. *Neuropsychologia*, 98, 192–200. DOI: <https://doi.org/10.1016/j.neuropsychologia.2016.10.002>
- Woodward, S. H., Ford, J. M., & Hammett, S. C. (1993). N4 to spoken sentences in young and older subjects. *Electroencephalography and Clinical Neurophysiology*, 87(5), 306–320. DOI: [https://doi.org/10.1016/0013-4694\(93\)90184-W](https://doi.org/10.1016/0013-4694(93)90184-W)
- Wu, C.-Y., Ho, M.-H. R., & Chen, S.-H. A. (2012). A meta-analysis of fMRI studies on Chinese orthographic, phonological, and semantic processing. *Neuroimage*, 63(1), 381–391. DOI: <https://doi.org/10.1016/j.neuroimage.2012.06.047>
- Yip, M. (1996). Lexicon optimization in languages without alternations. *Rutgers Optimality Archive*, 35.
- Yip, M. (2002). *Tone*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9781139164559>
- Yip, M. J. (1980). *The tonal phonology of Chinese* [Doctoral dissertation, Massachusetts Institute of Technology]. DOI: <https://doi.org/10.3406/clao.1980.1072>
- You, W., Zhang, Q., & Verdonschot, R. G. (2012). Masked syllable priming effects in word and picture naming in Chinese. *PloS One*, 7(10), e46595. DOI: <https://doi.org/10.1371/journal.pone.0046595>

- Yu, A. C. L. (2007). Understanding near mergers: The case of morphological tone in Cantonese. *Phonology*, 24, 187–214. DOI: <https://doi.org/10.1017/S0952675707001157>
- Zeng, Y., Fiorentino, R., & Zhang, J. (2021). Electrophysiological signatures of perceiving alternated tone in Mandarin Chinese: Mismatch negativity to underlying tone conflict. *Frontiers in Psychology*, 12, 735593. DOI: <https://doi.org/10.3389/fpsyg.2021.735593>
- Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, 27(1), 153–201. DOI: <https://doi.org/10.1017/S0952675710000060>
- Zhang, J., Lai, Y., & Sailor, C. (2011). Modeling Taiwanese speakers' knowledge of tone sandhi in reduplication. *Lingua*, 121(2), 181–206. DOI: <https://doi.org/10.1016/j.lingua.2010.06.010>
- Zhang, J., Zhang, C., Politzer-Ahles, S., Pan, Z., Huang, X., Wang, C., Peng, G., & Zeng, Y. (2022). The neural encoding of productive phonological alternation in speech production: Evidence from Mandarin tone 3 sandhi. *Journal of Neurolinguistics*, 62, 101060. DOI: <https://doi.org/10.1016/j.jneuroling.2022.101060>
- Zhou, W., Shu, H., Miller, K., & Yan, M. (2018). Reliance on orthography and phonology in reading of Chinese: A developmental study. *Journal of Research in Reading*, 41(2), 370–391. DOI: <https://doi.org/10.1111/1467-9817.12111>
- Zhou, X., & Marslen-Wilson, W. (1997). The abstractness of phonological representation in the Chinese mental lexicon. *Cognitive processing of Chinese and related Asian languages*, 3–26.

