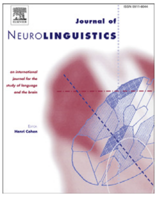
Journal of Neurolinguistics 33 (2015) 149–162

Contents lists available at ScienceDirect

Journal of Neurolinguistics

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



Mandarin third tone sandhi requires more effortful phonological encoding in speech production: Evidence from an ERP study

Caicai Zhang [a, b], Quansheng Xia [a, b], Gang Peng [a, b, c, *]

[a] CUHK-PKU-UST Joint Research Centre for Language and Human Complexity, The Chinese University of Hong Kong, Hong Kong, China
[b] Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong, China
[c] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

ARTICLE INFO

Article history:
Available online 20 August 2014

Keywords:
Mandarin
Tone sandhi
Computation mechanism
Lexical mechanism
Event-related potentials

ABSTRACT

In Mandarin Chinese, the third tone (T3) is changed to the second tone (T2) or T2-like when followed by another T3 syllable in speech production. It has long been debated whether the production of a linguistic pattern like Mandarin T3 sandhi is operated via a computation mechanism or a lexical mechanism. The computation mechanism is that the sandhi/non-sandhi form of a tone is computed according to the phonological context, irrespective of real words or novel words. The lexical mechanism is that the lexical representation of T3 + T3 words and the associated phonological forms are accessed in production, suggesting that T3 sandhi only applies to real words. To investigate whether T3 sandhi is mediated by a computation mechanism or a lexical mechanism, we examined the event-related potentials (ERPs) during the covert production of T2 + T3 and T3 + T3 sequences in real words and pseudowords in this study. We found that the second syllable elicited greater P2 amplitude in T3 + T3 sequences than in T2 + T3 sequences, indicating that the phonological encoding of sequences with T3 sandhi may be more effortful. Moreover, the phonological processing may not differ qualitatively between real words and pseudowords in the P2 time-window. It suggests that the phonological encoding of T3 sandhi may be mediated by a common computation mechanism in both real words and pseudowords. Alternative interpretations were also

---

[*] Corresponding author. Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Hong Kong, China. Tel.: +852 39434711.

E-mail address: gpengjack@gmail.com (G. Peng).

http://dx.doi.org/10.1016/j.jneuroling.2014.07.002

discussed. These findings, which are in line with previous behavioral findings that T3 sandhi occurs in phonological/phonetic encoding before the initiation of articulation, shed some light on the online encoding of linguistic patterns in production.

1. Introduction

It has long been debated whether the neural processing of speech sounds relies on domain-specific or domain-general mechanisms (see Zatorre & Gandour, 2008 for a review and many references therein). According to the domain-specific mechanism, language is predominantly processed in the left hemisphere (LH) (Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; Mäkelä, Alku, & Tiitinen, 2003; Shestakova et al., 2002; Tervaniemi et al., 2000) and music is predominantly processed in the right hemisphere (RH) (Peretz, 1990; Zatorre, Belin, & Penhune, 2002). According to the domain-general mechanism, the auditory processing is shaped by the acoustic properties of sounds, no matter whether they are speech or non-speech. One domain-general hypothesis is that the auditory processing of temporal cues is lateralized to the LH whereas that of spectral cues is lateralized to the RH (Jamison, Watkins, Bishop, & Matthews, 2006; Okamoto, Stracke, Draganova, & Pantev, 2009; Schonwiesner, Rubsamen, & von Cramon, 2005; Theunissen & Elie, 2014; Zatorre & Belin, 2001). Another hypothesis is that the left auditory cortex is sensitive to the analysis of acoustic parameters via a smaller temporal integration window, whereas the right auditory cortex is sensitive to the analysis of long, steady acoustic parameters (Bedoin, Ferragne, & Marsico, 2010; Poeppel, 2003).

As far as lexical tones are concerned, the domain-specific mechanism predicts that lexical tone processing would be lateralized to the LH, because lexical tones contrast lexical meanings in tone languages; however, the domain-general mechanism predicts that the auditory processing of lexical tones would be lateralized to the RH, because lexical tones are distinguished by patterns of spectral/pitch cues that occur on a long temporal window (hundreds of milliseconds). Previous studies have found evidence for both mechanisms, showing that lexical tone perception activates neural systems in both hemispheres (Gandour et al., 2002; Gandour et al., 2003; Gandour et al., 2004; Zhang, Xi, Xu, Shu, Wang, & Li, 2011). Auditory processing of lexical tones without attention elicited a mismatch negativity (MMN) with stronger amplitude over RH electrode sites than over LH electrode sites, in contrast to consonants (Luo et al., 2006). But lexical tone processing is lateralized to the LH in native tone language speakers when phonological or semantic processing is required beyond the auditory analysis of pitch cues in the stimuli (Brown-Schmidt & Canseco-Gonzalez, 2004; Gandour et al., 2002, 2003, 2004; Gu, Zhang, Hu, & Zhao, 2013; Malins & Joanisse, 2012; Schirmer, Tang, Penney, Gunter, & Chen, 2005; Xi, Zhang, Shu, Zhang, & Li, 2010; Zhang, Lai, & Sailor, 2011; Zhang et al., 2011; Zhang, Xi, Wu, Shu, & Li, 2012).

Relatively less investigated, however, are the neural bases of tone production. An interesting phenomenon in tone production is tone sandhi, namely, the change of a tone to other tones according to the linguistic context where it occurs. In some languages, tone sandhi involves tone alternation, which is not a product of tone co-articulation because co-articulation mostly influences the beginning or ending of a tone (e.g., Xu, 1997). For example, in Mandarin Chinese, when the third tone (T3) – a falling-rising tone – is followed by another T3 syllable, the first T3 is changed to the second tone (T2) – a high rising tone – or T2-like in speech production[1] (e.g., Chao, 1948; Cheng, 1968; Myers & Tsay, 2003; Peng, 2000; Shen, 1990; Wang & Li, 1967; Xu, 1997; Yuan & Chen, 2014; Zee, 1980; Zhang &

---

[1] There is another tone sandhi concerning T3 called half-T3 sandhi, i.e., T3 is replaced by a low falling tone (i.e., the first half of T3) when followed by non-T3 syllables. However, this half-T3 sandhi might be driven by simplification in articulation, and it is debatable whether it is stored as linguistic knowledge or not. For this reason, we investigated the T3 sandhi rather than the half-T3 sandhi in this study.

Peng, 2013). Previous behavioral studies have found that T3 sandhi is operated in phonological/phonetic encoding before the initiation of articulation (Politzer-Ahles & Zhang, in press; Xu, 1991). However, it is yet unknown whether the online encoding of sandhi patterns is mediated by a domain-specific or a domain-general mechanism. On the one hand, the encoding of sandhi patterns may be mediated by a domain-specific mechanism, as the sandhi patterns are likely stored specifically as linguistic knowledge. On the other hand, the encoding of sandhi patterns may rely on a domain-general mechanism, as tone production may be analogous to singing.

A recent functional magnetic resonance imaging (fMRI) study has found that the production of sequences with T3 sandhi activated the right posterior inferior frontal gyrus (IFG) (Chang, Lee, Tzeng, & Kuo, 2014). The left posterior IFG has been found to be involved in syllabification/prosodification in speech production (Indefrey & Levelt, 2004). The activation of posterior IFG in the RH therefore may suggest that the production of T3 sandhi is mediated by a domain-general mechanism, similar to singing. However, lack of evidence for the involvement of the domain-specific mechanism is not equal to negative evidence. More studies are required to obtain a fuller understanding of the nature of the production of T3 sandhi.

Importantly, in those studies that examined the productivity of linguistic patterns in novel words, it has long been debated whether the encoding of a linguistic pattern like tone sandhi is operated via a computation mechanism or a lexical mechanism (e.g., Berko, 1958; Hsieh, 1970, 1975, 1976; Pinker, 1998; Pinker, Lebeaux, & Frost, 1987; Zhang & Lai, 2010; Zhang, Lai, et al., 2011; Zhang & Peng, 2013). The computation mechanism suggests that the sandhi and non-sandhi forms of a tone are computed according to the phonological context, no matter whether it is a real word or a novel word (e.g., Zhang & Lai, 2010; Zhang & Peng, 2013). The lexical mechanism is that the encoding of tone sandhi is operated via access to the lexical representation and the phonological forms attached to it, and therefore only applies to real words (e.g., Hsieh, 1970, 1975, 1976). Both the computation and lexical mechanisms are possibly domain-specific mechanisms, but the computation mechanism is likely *less* domain-specific, because it predicts the productivity of linguistic patterns in novel words, which could be relatively free of semantic and syntactic constraints. Therefore, an examination of the computation and lexical mechanisms could also shed some light on the nature of the encoding of T3 sandhi, which is the aim of the present study. Moreover, it is worth examining whether the neural activities are lateralized to the LH or the RH, to further determine the domain-specificity/generality of the computation mechanism and the lexical mechanism.

Previous studies have consistently shown that Mandarin T3 sandhi is productive in novel disyllabic sequences (Xu, 1997; Zhang & Lai, 2010; Zhang & Peng, 2013). Xu (1997) found that T3 sandhi applied to a nonsense disyllabic sequence, 马马 (*ma3* – 'horse', *ma3* – 'horse'). Zhang and Lai (2010) provided by far the most extensive evidence for the productivity of T3 sandhi, showing that T3 sandhi applied to three types of novel disyllabic sequences: (1) a non-occurring sequence of two occurring morphemes, such as 尺洒 (*chi3* – 'a ruler', *sa3* – 'to spray'); (2) a combination of one occurring morpheme and a pseudo-syllable, the base syllable of which occurs but its combination with T3 does not exist,[2] such as 闯*zeng3* (*chuang3* – 'to break in') or *ping3*马 (*ma3* – 'horse'); and (3) a sequence of two pseudo-syllables, such as *ping3 zeng3*. Moreover, our earlier study extended the finding of Zhang and Lai (2010), showing that T3 sandhi also applies productively to non-occurring syllables in Mandarin, such as *tiang3 lua3* (Zhang & Peng, 2013).

The above studies provide evidence for the computation mechanism mediating the encoding of T3 sandhi in novel disyllabic sequences. However, it is yet unclear whether real words and novel words are operated by a common mechanism (presumably the computation mechanism), or by the lexical and computation mechanism respectively. Moreover, it is unknown *when* the lexical and computation mechanisms are operated in the online encoding of spoken words, and whether the neural activities of these mechanisms are lateralized or not.

To this end, this study used event-related potentials (ERPs) to examine the time-course of the encoding of T3 sandhi in real words and novel words. We adopted a within-subjects *lexicality* (real

---

[2] For example, the base syllable *zeng* occurs, because *zeng1* (e.g., 曾 'a family name') and *zeng4* (e.g., 憎, 'hate') are morphemes in Mandarin Chinese. But the combination of *zeng* with T3, i.e., *zeng3*, does not occur.

words vs. pseudowords) × *tone sandhi* (T2 + T3 vs. T3 + T3) design. ERPs were recorded while subjects *covertly* produced two auditorily presented syllables as a disyllabic sequence. We compared the covert production of T2 + T3 vs. T3 + T3 sequences, both of which would be encoded as T2 + T3 in the output of speech production, but only T3 + T3 sequences involving tone sandhi. Moreover, real words and pseudowords were compared to examine whether the encoding of T3 sandhi recruits a similar computation mechanism, or relies on the lexical mechanism and computation mechanism respectively. If mediated by a common computation mechanism, tone sequences with T3 sandhi, irrespective of real words or pseudowords, may modulate the P2 (or related components), which is sensitive to phonological processing (Crowley & Colrain, 2004; Landi, Crowley, Wu, Bailey, & Mayes, 2012; Tremblay, Kraus, McGee, Ponton, & Otis, 2001) and visual word encoding (Dunn, Dunn, Languis, & Andrews, 1998). If the encoding of real words is mediated by the lexical mechanism, different form pseudowords, the stored lexical representation and phonological forms may be accessed in real words but not in pseudowords, modulating the N400 (or related components), which indexes the access to lexical representation and semantic memory (Kutas & Federmeier, 2011). Moreover, where the ERPs differ between T2 + T3 and T3 + T3 sequences, we compared the ERPs at LH and RH electrode sites to determine their lateralization pattern.

2. Methods

2.1. Participants

Seventeen participants (11 female, 6 male; mean age = 23.5 years, SD = 2.3, aged 19–28.3 years) were paid to participate in this experiment. All subjects were native Mandarin speakers from Northern China. They were all university students, right-handed, with normal hearing, no musical training and no reported history of neurological illness. Four additional participants (2 female, 2 male) were rejected from the analysis due to over 20% of trials with mispronunciations in the pseudoword condition. The experimental procedures were approved and informed written consent was obtained from each participant in compliance with a protocol approved by the Survey and Behavioral Research Ethics Committee of The Chinese University of Hong Kong.

2.2. Stimuli and experimental design

The real word condition included ten minimal pairs of T2 + T3 and T3 + T3 words (see Table 1) and the pseudoword condition included ten minimal pairs of T2 + T3 and T3 + T3 disyllabic sequences (see Table 2). Ten native Mandarin speakers who did not participate in this ERP experiment were asked to rate the subjective familiarity of each word on a 1–9 scale (1 = very unfamiliar, 9 = very familiar), and T2 + T3 and T3 + T3 stimuli were matched in subjective familiarity (T2 + T3: 7.13, T3 + T3: 5.63; $t = 1.882$, $p = 0.063$). Syntactic structure was also largely matched between the T2 + T3 and T3 + T3 real words. Moreover, all of the second T3 syllables have a sonorant or glide initial (/m/, /l/ and /j/), in

Table 1
List of real words.

No.	T2 + T3 words			T3 + T3 words		
	Pinyin	Sinogram	Gloss	Pinyin	Sinogram	Gloss
1	*er2 yu3*	儿语	Baby talk	*er3 yu3*	耳语	Whisper
2	*tao2 mi3*	淘米	To wash rice	*tao3 mi3*	讨米	To ask for rice
3	*bai2 ma3*	白马	White horse	*bai3 ma3*	百马	One hundred horses
4	*bai2 mi3*	白米	White rice	*bai3 mi3*	百米	One hundred meters
5	*du2 yin3*	毒瘾	Drug addiction	*du3 yin3*	赌瘾	Gambling addiction
6	*cai2 li3*	财礼	Dowry	*cai3 li3*	彩礼	Bride price
7	*qi2 ma3*	骑马	To ride a horse	*qi3 ma3*	起码	At least
8	*nian2 mi3*	黏米	Sticky rice	*nian3 mi3*	碾米	To grind the rice
9	*wu2 li3*	无礼	Rude	*wu3 li3*	五里	Five miles
10	*yi2 lao3*	遗老	Remnant	*yi3 lao3*	倚老	Self-important

Table 2
List of pseudowords in *pinyin* form.

No.	T2 + T3 sequences	T3 + T3 sequences
1	*shong2 muai3*	*shong3 muai3*
2	*cua2 luang3*	*cua3 luang3*
3	*diang2 mua3*	*diang3 mua3*
4	*dv2 mia3*	*dv3 mia3*
5	*tiang2 lua3*	*tiang3 lua3*
6	*fiu2 mv3*	*fiu3 mv3*
7	*pe2 luai3*	*pe3 luai3*
8	*tei2 muang3*	*tei3 muang3*
9	*fiang2 nua3*	*fiang3 nua3*
10	*tia2 nia3*	*tia3 nia3*

order to ensure F0 starts at the syllable onset. For the pseudowords, we used non-occurring syllables, which were combinations of occurring initials and finals in Mandarin (also see Zhang & Peng, 2013). For example, initial 't' and final 'iang' occur in Mandarin individually but their combination 'tiang' does not occur. Similar to real words, all the second T3 syllables in pseudowords have a sonorant initial (/m/, /l/ and /n/) in order to ensure F0 starts at the syllable onset.

In addition to the test items, fillers were included to conceal the experimental purpose of examining T3 sandhi. Fillers were 20 disyllabic real words and 20 disyllabic pseudowords, with half of the items being a T2 syllable followed by a syllable with any tone but T3 (i.e., T2 + T1/T2/T4), and the other half being a T3 syllable followed by a syllable with any tone but T3 (i.e., T3 + T1/T2/T4) in real words and pseudowords respectively. The ratio of test items and fillers was 1:1.

Each syllable in the real words and pseudoword sequences was recorded in isolation from a female native speaker of Beijing Mandarin at a sampling frequency of 22,050 Hz in a quiet room. In particular, this speaker was instructed to pronounce T3 in its full form, i.e., low falling-rising tone. All the syllables were normalized to 400 ms in duration and to 70 dB in average intensity using Praat (Boersma & Weenink, 2012). Moreover, the F0 trajectory of all of the second T3 syllables was manipulated to be identical. Specifically, we selected one T3 syllable, the F0 trajectory of which is closest to the mean of all of the second T3 syllables and replaced the F0 trajectory of other T3 syllables with that of the selected syllable. This manipulation is important for maintaining a similar recognition point of T3, in order to minimize differences between trials in the onset of phonological/phonetic encoding following the recognition of the second T3. The manipulation of F0 trajectory was done for real words and pseudowords separately. The F0 trajectory of the first T3 syllables was not manipulated.

2.3. Procedure

We used a speech production task in this study. A fixation, which indicates the beginning of a trial, appeared and stayed on the screen for 500 ms, following which, two syllables were presented consecutively to the subjects via the earphone binaurally. These two syllables were separated by a long silent interval jittered between 600 and 1000 ms, in order to minimize transient effects of the neural processing of the first syllable persisting into the processing of the second syllable. Subjects were instructed to *covertly* produce these two syllables as a disyllabic word as soon as they heard the second syllable, and *overtly* produce the disyllabic word only when a microphone sign appeared on the screen. The overt production (cued by the microphone sign) was delayed by a long interval jittered between 1000 and 1600 ms after the offset of the second syllable, in order to avoid any influence of articulation or preparation for articulation on the neural processing of the second syllable. The experimental procedure was illustrated in Fig. 1.

The real word and pseudoword stimuli were presented in two separate blocks. In each block, 20 test trials (T2 + T3 and T3 + T3) and 20 filler trials (T2 + T1/T2/T4 and T3 + T1/T2/T4) were randomized and repeated twice (i.e., 80 trials per block). The second T3 syllable in T2 + T3 and T3 + T3 trials was acoustically identical in a block. Each block was repeated three times. Half of the subjects did three blocks of real words first, and the other half did three blocks of pseudowords first. Subjects were given

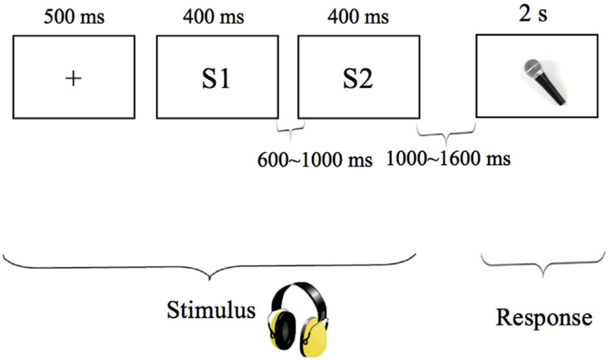


Fig. 1. Illustration of the experimental procedure. 'S1' and 'S2' refer to the first and second syllable respectively.

a few minutes to rest between two blocks. In total, there were 60 trials for each of the four conditions – T2 + T3 real words, T3 + T3 real words, T2 + T3 pseudowords and T3 + T3 pseudowords.

Before the experiment, the subjects had a familiarization test of the pseudowords, in which each isolated syllable were randomized and presented to the subjects auditorily with its pinyin form visually displayed on the screen to facilitate recognition. The subjects were instructed to repeat each syllable three times overtly and were allowed to listen to a syllable multiple times. Mispronunciations were pointed out by the experimenter and a correct pronunciation was demonstrated. The presentation pace was controlled by the subjects. Before the real word and pseudoword blocks, a practice block with two items from the filler real words and filler pseudowords was presented respectively to familiarize the subjects with the procedure.

2.4. EEG recording and data analysis

Electroencephalographic (EEG) data were recorded using SynAmps 2 amplifier (NeuroScan, Charlotte, NC, U.S.) with a cap carrying 64 Ag/AgCl electrodes placed on the scalp at the standard locations according to the extended international 10–20 system. Vertical electrooculography (EOG) was recorded using bipolar channel placed above and below the left eye, and horizontal EOG was recorded using bipolar channel placed lateral to the outer canthi of both eyes. The online reference electrode was located between Cz and CPz, and two more electrodes attached to each mastoid were used as offline references. Impedance between the online reference electrode and any recording electrode was kept below 5 kΩ. Alternating current signals (0.15–400 Hz) were continuously recorded and digitized at the sampling rate of 1000 Hz.

We time-locked the ERP analysis to the onset of the second T3 syllable, comparing the neural processing of the second T3 syllable when it was preceded by a T3 syllable (i.e., T3 + T3, where T3 sandhi would apply) versus when it was preceded by a T2 syllable (i.e., T2 + T3, where no T3 sandhi would apply). As soon as the second auditory syllable is recognized, these two syllables would be concatenated and encoded as a disyllabic sequence in covert speech. In other words, the onset of the encoding of T3 sandhi depends on the recognition time of the second T3 syllable. We compared the ERPs of the T3 + T3 and T2 + T3 stimuli to examine the encoding of T3 sandhi in real words and pseudowords.

The BESA EEG V.5.1 was used to analyze the data. Epochs ranged from −100 to 1200 ms after the onset of the second T3 syllable. The EEG recordings were re-referenced offline against average-mastoid, and re-filtered with a 0.5–30 Hz band-pass zero-phase shift digital filter (slope 12 dB/Oct). Baseline correction was performed according to prestimulus activity.

Trials with mispronunciations (mispronunciation of rhymes or tones) and long pause between two syllables were excluded form the ERP analysis. Trials where there was uncertainty in determining whether T3 sandhi has applied or not were also excluded from analysis. Two experimenters evaluated whether a T3 syllable was produced in its sandhi form or non-sandhi form, and only trials where both experimenters agreed on the application of T3 sandhi were included in the analysis. For quality control purpose, four subject's data were excluded altogether due to more than 20% of trials rejected in the pseudoword condition. For the remaining seventeen subjects, the average rejection rate was 2.25% (SD = 2.63%) for T2 + T3 real words, 6.96% (SD = 6.54%) for T3 + T3 real words, 0.39% (SD = 0.73%) for T2 + T3 pseudowords, and 1.96% (SD = 2.52%) for T3 + T3 pseudowords. Furthermore, epochs with potentials exceeding ±120 μV at any electrode were rejected from analysis. The mean rejection rate (including trials with mispronunciations or potentials exceeding ±120 μV) across the four conditions is 8.83% (SD = 5.07%). For each subject, ERPs were then averaged across the remaining trials for each condition.

Five ERP components – the N1 (140–230 ms), the P2 (230–320 ms), the N2 (320–520 ms), the N400 (520–820 ms), and the Late Positive Component (LPC, 520–820 ms) – were determined from the global field power[3] and topographical plots (Fig. 2). The N400 and the LPC occurred in a similar time-window but had different topographical distributions. The N400 showed a fronto-central distribution whereas the LPC showed a left posterior distribution. Different sets of electrodes were selected for the N1, P2, N2, N400 and LPC according to the topographic distributions. Eighteen frontal, central and posterior electrodes (FC3, FC1, FCz, FC2, FC4, C3, C1, Cz, C2, C4, CP3, CP1, CPz, CP2, CP4, P1, Pz, P2) were selected for the N1, thirteen fronto-central electrodes (FP1, FPz, FP2, AF3, AF4, F3, F1, Fz, F2, F4, FC1, FCz, FC2) were selected for the P2, fifteen frontal and central electrodes (F3, F1, Fz, F2, F4, FC3, FC1, FCz, FC2, FC4, C3, C1, Cz, C2, C4) were selected for the N2, thirteen fronto-central electrodes (FP1, FPz, FP2, AF3, AF4, F3, F1, Fz, F2, F4, FC1, FCz, FC2) were selected for the N400, and eight posterior electrodes (P7, P5, P3, P1, PO7, PO5, PO3, POz) were selected for the LPC.

For each subject, the ERPs were averaged across the selected electrodes for each ERP component respectively. The peak latency of a component was determined from the minimal (for N1, N2 and N400) or maximal point (for P2 and LPC) of the ERP waves within the defined time-windows for each condition. Mean amplitude of the N1, P2, N2, N400 and LPC was also obtained from the defined time-windows for each condition.

3. Results

Fig. 3 shows the ERPs of the four conditions at nine electrodes (F3, Fz, F4, C3, Cz, C4, P3, Pz and P4). To contrast the ERPs between T2 + T3 and T3 + T3 stimuli, Fig. 4 plots the ERPs of read words and pseudowords separately at two electrodes, F3 and F4. Two-way repeated measures ANOVAs were conducted on the peak latency and mean amplitude of each component by indicating *lexicality* (real words vs. pseudowords) and *tone sandhi* (T2 + T3 vs. T3 + T3) as two within-subjects factors.

No effects were significant for the peak latency or mean amplitude of the N1.

For P2, there was a significant main effect of *tone sandhi* for its amplitude ($F(1, 16) = 8.375$, $p < 0.05$), with greater P2 amplitude elicited by T3 syllables preceded by a T3 syllable than those preceded by a T2 syllable (2.958 μV vs. 2.449 μV). No other effects were significant. Moreover, to examine the lateralization pattern of the P2 amplitude, mean P2 amplitude was obtained for each condition from two LH electrodes (F3 and FC3), two mid-line electrodes (Fz and FCz) and two RH electrodes (F4 and FC4). Three-way repeated measures ANOVA was conducted on the P2 amplitude by indicating *lexicality* (real words vs. pseudowords), *tone sandhi* (T2 + T3 vs. T3 + T3) and *lateralization* (LH vs. mid-line vs. RH) as within-subjects factors. There were a significant main effect of *lateralization* ($F(2, 32) = 12.140$, $p < 0.001$), but no interaction effects of *lateralization* with other factors. Post-hoc tests with Bonferroni correction for multiple comparisons reveal that the P2 amplitude at mid-line electrodes was larger than that at LH electrodes (2.896 μV vs. 2.323 μV, $p < 0.001$) and RH electrodes (2.896 μV vs. 2.506 μV, $p < 0.01$). There was no significant difference between LH and RH electrodes.

For N2, there was only a significant main effect of *lexicality* for its peak latency ($F(1, 16) = 6.243$, $p < 0.05$), with N2 peaking earlier in the real word condition than in the pseudoword condition (408 ms vs. 432 ms).

---

[3] Global field power was calculated as the square root mean of the amplitudes of the ERP wave averaged across all conditions and all subjects at all electrodes, at each time point between −100 and 1200 ms of the onset of the second T3 syllable.

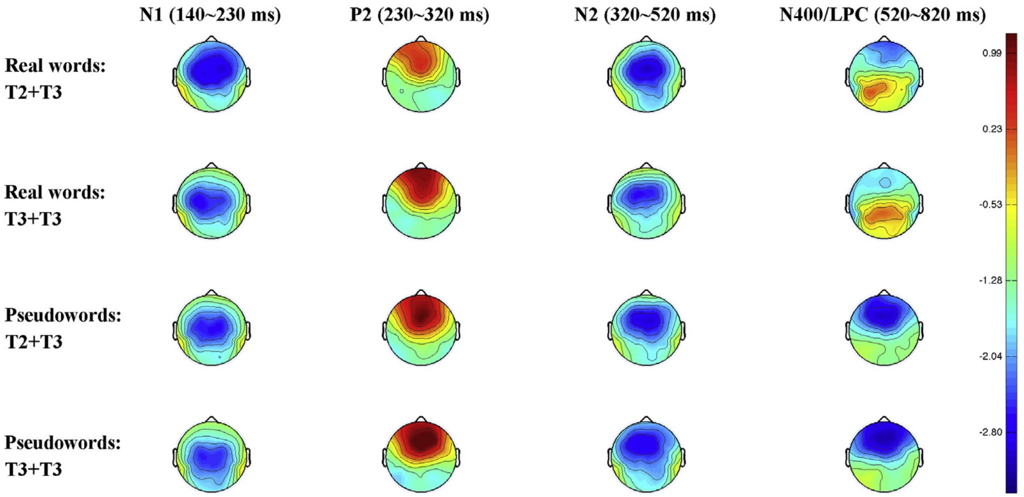


Fig. 2. Topographical maps for the four conditions, T2 + T3 real words, T3 + T3 real words, T2 + T3 pseudowords and T3 + T3 pseudowords, at the time-windows of the N1 (140–230 ms), P2 (230–320 ms), N2 (320–520 ms), N400 (520–820 ms), and LPC (520–820 ms).

For N400, there was only a significant main effect of *lexicality* for its amplitude ($F(1, 16) = 32.940$, $p < 0.001$), with stronger N400 amplitude elicited by pseudowords than real words (−2.173 μV vs. −0.737 μV).

For LPC, there was only a significant main effect of *lexicality* for its amplitude ($F(1, 16) = 29.426$, $p < 0.001$), with stronger LPC amplitude elicited by real words than by pseudowords (1.594 μV vs. 0.093 μV). Moreover, the topographical maps (Fig. 2) show that LPC elicited by real words appeared to

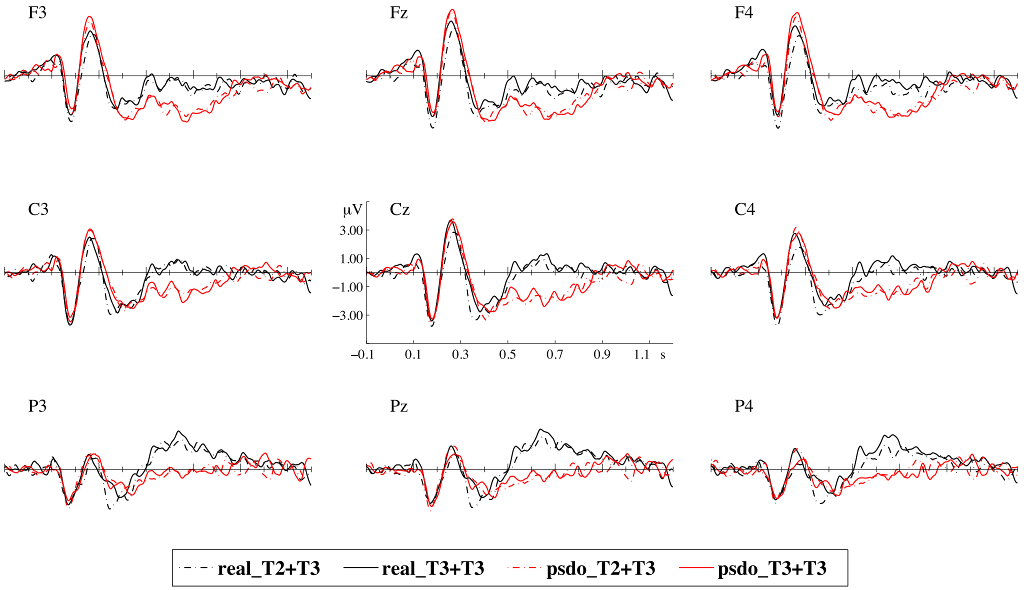


Fig. 3. ERP waves of the four conditions, T2 + T3 real words (‘real_T2 + T3’), T3 + T3 real words (‘real_T3 + T3’), T2 + T3 pseudowords (‘psdo_T2 + T3’), and T3 + T3 pseudowords (‘psdo_T3 + T3’) at nine electrode sites, F3, Fz, F4, C3, Cz, C4, P3, Pz and P4.

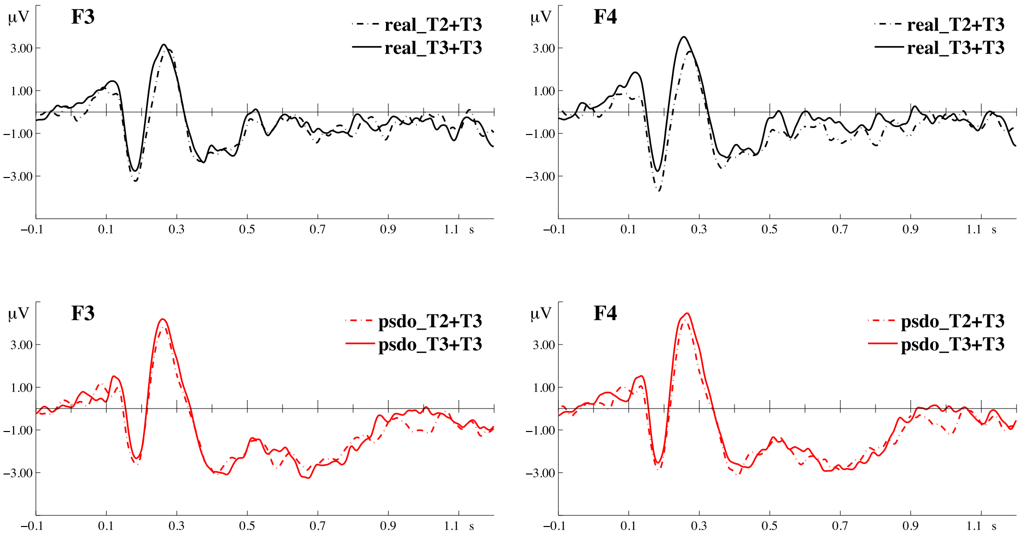


Fig. 4. ERP waves plotted for real words and pseudowords separately at two representative electrodes, F3 and F4. Top panel: T2 + T3 ('real_T2 + T3') and T3 + T3 ('real_T3 + T3') real words; bottom panel: T2 + T3 ('psdo_T2 + T3') and T3 + T3 ('psdo_T3 + T3') pseudowords.

be lateralized to LH electrodes whereas no lateralization tendency was observed for pseudowords. This observation was confirmed by an analysis of the lateralization pattern of the LPC amplitude. Mean amplitude of LPC for each condition was obtained from two LH electrodes (P3 and PO3), two mid-line electrodes (Pz and POz) and two RH electrodes (P4 and PO4). Three-way repeated measures ANOVA was then conducted on the LPC amplitude by indicating *lexicality* (real words vs. pseudowords), *tone sandhi* (T2 + T3 vs. T3 + T3) and *lateralization* (LH vs. mid-line vs. RH) as within-subjects factors. There were a significant main effect of *lateralization* ($F(1, 16) = 5.530$, $p < 0.05$) and an interaction effect of *lexicality* by *lateralization* ($F(1, 16) = 3.780$, $p < 0.05$). Post-hoc tests with Bonferroni correction suggest that for real words, the LPC amplitude at LH (1.612 μV vs. 1.277 μV, $p < 0.05$) and mid-line electrodes (1.595 vs. 1.277, $p < 0.01$) was significantly stronger than that at RH electrodes, whereas there was no significant difference between LH, mid-line and RH electrodes (0.022 μV vs. −0.244 μV vs. −0.277 μV) for pseudowords.

4. Discussion

4.1. Encoding of T3 sandhi in the P2 time-window: computation mechanism vs. lexical mechanism

An important and unresolved question is whether the production of linguistic patterns like Mandarin T3 sandhi is operated by a common mechanism (presumably the computation mechanism), or by the lexical and computation mechanism in real words and pseudowords respectively. To examine this question, we compared the ERPs during the covert production of T2 + T3 and T3 + T3 sequences of real words and pseudowords.

We found a significant effect of tone sandhi at the P2 time-window, with the second T3 syllables eliciting greater P2 amplitude in T3 + T3 sequences than in T2 + T3 sequences. Moreover, the encoding of T3 + T3 stimuli may not be qualitatively different between real words and pseudowords at the P2 time-window (given the lack of interaction effect of *tone sandhi* by *lexicality*). No effect of tone sandhi was found in the N400 time-window. Taken together, our findings are most compatible with a common computation mechanism mediating the phonological encoding of T3 sandhi in both real words and pseudowords. There is no evidence to suggest that the encoding of T3 sandhi in real words is mediated by the lexical mechanism, different from pseudowords.

Earlier ERP studies found that previously encountered auditory non-words elicited reduced N1 (i.e., less negative) and P2 (i.e., less positive) amplitude compared to novel non-words in normal children, suggesting that familiarity facilitates auditory and phonological processing of non-words (Landi et al., 2012). Moreover, the P2 was elicited during the encoding of visual words in a memory task, where its amplitude differed between poor recallers and good recallers (Dunn et al., 1998). Greater frontal P2 amplitude was elicited in poor recallers, which may indicate more effortful word encoding. In a similar line, the P2 was found to be modulated by task difficulty, with greater P2 amplitude elicited by harder tasks (Kim, Kim, Yoon, & Jung, 2008). Given these findings, the greater P2 amplitude elicited in the T3 + T3 sequences likely suggests that the encoding of sequences with T3 sandhi requires additional or more effortful phonological processing. In particular, the perceived T3 + T3 sequence has to be encoded as T2 + T3, whereas the encoding of T2 + T3 stimuli would be more transparent, because the sequence would be encoded as perceived.

Our finding is consistent with previous behavioral findings that that T3 sandhi occurs in phonological/phonetic encoding before the initiation of articulation (Politzer-Ahles & Zhang, in press; Xu, 1991), and provides further evidence for its ERP correlates (likely encoded in the P2 time-window). Xu (1991) found that subjects made significantly more errors recalling items from those lists containing T3 + T3 sequences, presumably because T3 became T2-like in covert production when encoding those sequences phonologically/phonetically into the short-term memory. Politzer-Ahles and Zhang (in press) reached a similar conclusion using a priming paradigm. All these studies provide evidence for the encoding of T3 sandhi online, suggesting that it is not a product of lower-level articulatory interaction of neighboring tones.

As mentioned earlier, the computation mechanism is likely less domain-specific than the lexical mechanism, because it predicts the productivity of T3 sandhi in novel words that are relatively free of semantic and syntactic constraints. Our finding that the encoding of T3 sandhi in real words and pseudowords is mediated by a similar computation mechanism therefore may suggest that the nature of T3 sandhi is less domain-specific. However, the P2 amplitude seems to be the strongest along midline electrodes, showing no obvious lateralization over LH or RH electrodes. There is no clear evidence as to whether the computation mechanism is operated by neural processes in the LH or the RH. This finding is less consistent with the previous fMRI study, which found activation in the right posterior IFG mediating the production of sequences with T3 sandhi (Chang et al., 2014). However, the underlying source activity of the P2 may be different from the scalp voltage topography (see Van Petten & Luka, 2006 for a similar claim). The differences may arise from a slight tilt of the summed electrical dipole in one hemisphere which points slightly toward the other hemisphere. To further examine the lateralization of the encoding of T3 sandhi, MEG or fMRI studies are required.

Our findings are roughly consistent with previous speech production models. According to Indefrey and Levelt (2004), picture naming involves three successive operations: (1) the activation and selection of concept associated with a picture (175 ms), (2) lexical retrieval (75 ms), and (3) form encoding, which lasts for about 350 ms for a five-segment word till the initiation of articulation (phonological code retrieval: 80 ms; syllabification: 25 ms/segment; phonetic encoding: 145 ms). Consistent with this model, previous ERP studies show that visually presented words with non-anticipated lexical stress patterns (e.g., words with initial stress when words with final stress were anticipated) elicited more negative N2 (400–500 ms) than words with anticipated lexical stress patterns (Schiller, 2006; Schiller, Bles, & Jansma, 2003). It suggests that lexical stress patterns could have been encoded during or before the 400–500 ms time-window, therefore allowing non-anticipated lexical stress patterns to be detected neurally. Importantly, the encoding of lexical stress patterns is similar to the encoding of tones, both involving the assigning of pitch patterns to neighboring syllables (though lexical stress patterns also involve duration and intensity differences between stressed and unstressed syllables) (Zhang, Nissen, & Francis, 2008). If this is the case, the encoding of tone patterns might occur at a similar time-window as that of lexical stress patterns. However, our finding of T3 sandhi effect at the P2 time-window (230–320 ms) seems to be earlier than the N2 time-window (400–500 ms). This discrepancy may be because the syllables were presented auditorily in this study, which may allow access of phonological code directly from the auditory signals, without necessarily retrieving the lexical representation of a word. Moreover, in the case of pseudowords, no lexical representation could be accessed. Therefore, it could have minimized the time required for concept selection and lexical retrieval, allowing the form encoding to start earlier. If this is the case, the encoding of tone patterns might indeed occur at a roughly similar time-window as that of lexical stress patterns. This possibility needs to be examined in future studies with a design similar to Schiller (2006).

Finally, our finding supports a common computation mechanism mediating the operation of tone sandhi in Mandarin Chinese, but tone sandhi in other Chinese dialects may be mediated by a different mechanism. For example, it has been found that tone sandhi patterns in Taiwanese are much less productive in novel words (Hsieh, 1970, 1976; Zhang, Lai, et al., 2011). It suggests that Taiwanese tone sandhi is mostly mediated by a lexical mechanism, which retrieves the stored sandhi forms of real words from memory. Novel words, which have no stored sandhi forms, therefore do not undergo tone sandhi as much. Future studies are needed to examine the neural correlates of the computation and lexical mechanism in different Chinese dialects.

4.2. Alternative explanations for the P2 effect

Some alternative interpretations are available, which may attribute the P2 effect to the perceptual processing differences of T3 + T3 vs. T2 + T3 sequences, rather than the more effortful phonological encoding. One such explanation is that the first T3 syllable might have a priming effect on the perceptual processing of the second T3 syllable, which shares the same tone. Therefore it eases the phonological processing of the second T3 syllable in T3 + T3 stimuli compared to T2 + T3 stimuli. However, priming and repetition effects are often associated with reduced amplitude of ERP components (Besson, Kutas, & van Petten, 1992; Guillaume et al., 2009; Penney, Mecklinger, & Nessler, 2001; Rodriguez-Fornells, Münte, & Clahsen, 2002), which is less consistent with our finding of greater P2 amplitude elicited by the T3 + T3 sequences. Another explanation is that the subjects might perceive the two auditorily presented T3 syllables as one disyllabic word. In connected speech, a T3 syllable in its non-sandhi form is rarely followed by another T3 syllable, due to the application of T3 sandhi. The low probability of a non-sandhi T3 syllable followed by another T3 syllable therefore might increase the effort of phonological processing of the second T3 syllable. This explanation might account for the greater P2 amplitude elicited by the T3 + T3 stimuli. However, the two auditorily presented syllables were separated by a long silent interval of 600–1000 ms in this study, which makes it less likely for the subjects to perceive these two separated syllables as one disyllabic word. This interpretation does not fit our study very well either.

Moreover, previous tone perception studies also support that the encoding of T3 sandhi may start in the P2 time-window. As mentioned earlier, the onset of encoding T3 sandhi depends on the recognition time of the second T3 syllable. Previous eye-tracking and tone perception studies suggest that a T3 syllable is likely to be recognized with very brief auditory input at the beginning of a syllable (Gottfried & Suiter, 1997; Lee, 2009; Lee, Tao, & Bond, 2008; Shen, Deutsch, & Rayner, 2013). Shen et al. (2013) found that the T3 began to be recognized unambiguously at about 250 ms after the onset of an auditory T3 syllable, as indicated by the divergence of fixation at the picture associated with the target T3 syllable and at pictures associated with other tones from 250 ms onwards. Moreover, Gottfried and Suiter (1997) found that isolated Mandarin tone stimuli excised from the first six glottal pulses of a syllable (e.g., about 60 ms for F0 of 100 Hz) could be identified quite accurately (T3: 65%), a finding which was replicated by Lee et al. (2008) with a different set of Mandarin syllables (T3: 87%). The above studies suggest that although speech recognition is an incremental process, the cues necessary to recognize T3 might become available at a rather early time, possibly during the N1 and P2 time-windows.

In summary, we conclude that the greater P2 amplitude elicited by T3 + T3 sequences than T2 + T3 sequences is best explained by the more effortful phonological encoding of T3 sandhi, rather than by perceptual processing differences.

4.3. N2, N400 and LPC effects: real words vs. pseudowords

The effect of lexicality we found in the N2, N400 and LPC time-windows also requires an explanation. Specifically, real words elicited an earlier-peaking N2, smaller N400 amplitude and greater LPC amplitude than pseudowords. Moreover, the LPC amplitude of real words was stronger at LH and midline electrodes than RH electrodes.

The N2 is usually elicited in the go/no-go tasks and is thought to index response inhibition or conflict monitoring (Donkers & van Boxtel, 2004; Folstein & van Petten, 2008; Heil, Osman, Wiegelmann, Rolke, & Hennighausen, 2000). The N2 differs between trials that require a response (go-trials) and trials that do not require a response (no-go trials), suggesting that the relevant information necessary to withhold a behavioral response must have been available during the N2 time-window. Although we did not use a go/no-go task in this study, we found that the N2 peaked earlier for real words than pseudowords. It might suggest that the information needed to initiate spoken word encoding becomes available earlier in real words than in pseudowords.

The N400 has been hypothesized to index the binding of information obtained from stimulus input with representations from long-term (i.e., activation of mental lexicon) and short-term memory (i.e., violation of semantic expectancy from a recent context) (Hagoort, Baggio, & Willems, 2009; Kutas & Federmeier, 2011). Given this account, our results may suggest that it is easier to access the semantic memory of T3 syllables in real words than in pseudowords, which were non-occurring syllables with no lexical representation in Mandarin Chinese.

The LPC has been found to index domain-general decisional process with regard to stimulus categorization (e.g. Bornkessel-Schlesewsky et al., 2011). For example, previous studies found that the LPC amplitude is sensitive to decision accuracy, with larger LPC amplitude elicited in response to trials that were accurately categorized (Finnigan, Humphreys, Dennis, & Geffen, 2002; Zhang, Peng, & Wang, 2013). Our results may be interpreted as indicating that real words were easier to categorize than pseudowords, and that the categorization of real words was mostly mediated by language knowledge in the LH.

5. Conclusion

In this ERP study, we examined the time-course of covert production of T2 + T3 and T3 + T3 sequences of real words and pseudowords. We found that the second T3 syllables elicited greater P2 amplitude in T3 + T3 sequences than in T2 + T3 sequences, indicating that the phonological encoding of T3 sandhi may be more effortful. Moreover, the phonological processing may not differ qualitatively between real words and pseudowords in the P2 time-window. Our findings suggest that the phonological encoding of T3 sandhi in both real words and pseudowords may be mediated by a common computation mechanism. It also provides evidence for the online encoding of T3 sandhi before the initiation of articulation (see Politzer-Ahles & Zhang, in press; Xu, 1991 for similar findings), suggesting that T3 sandhi is not a product of lower-level articulatory interaction of neighboring tones. Our findings shed some light on the online encoding of linguistic patterns.

Acknowledgements

This work is supported in part by grants from National Natural Science Foundation of China (NSFC: 61135003) and the Research Grant Council of Hong Kong (GRF: 448413). We thank Dr. Jie Zhang for his constructive comments on a broad spectrum of tone sandhi patterns in Chinese dialects. We thank Ms. Xiao Wang for her help with the data collection and analysis. We also thank members of the Joint Research Centre for Language and Human Complexity for their useful discussions.

References

Bedoin, N., Ferragne, E., & Marsico, E. (2010). Hemispheric asymmetries depend on the phonetic feature: a dichotic study of place of articulation and voicing in French stops. *Brain and Language, 115*(2), 133–140.

Berko, J. (1958). The child's learning of English morphology. *Word, 14*, 150–177.

Besson, M., Kutas, M., & van Petten, C. (1992). An event-related potential (ERP) analysis of semantic congruity and repetition effects in sentences. *Journal of Cognitive Neuroscience, 4*(2), 132–149.

Boersma, P., & Weenink, D. (2012). *Praat: Doing phonetics by computer (Version 5.3.23)* [Computer program] http://www.praat.org. Last viewed on August 7, 2012.

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., et al. (2011). Think globally: cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language, 117*, 133–152.

Brown-Schmidt, S., & Canseco-Gonzalez, E. (2004). Who do you love, your mother or your horse? An event-related brain potential analysis of tone processing in Mandarin Chinese. *Journal of Psycholinguistic Research, 33*(2), 103–135.

Chang, H.-C., Lee, H.-J., Tzeng, O. J. L., & Kuo, W.-J. (2014). Implicit target substitution and sequencing for lexical tone production in Chinese: an fMRI study. *PLoS ONE, 9*(1), e83126.

Chao, Y. R. (1948). *Mandarin Primer*. Cambridge: Harvard University Press.

Cheng, C.-C. (1968). English stresses and Chinese tones in Chinese sentences. *Phonetica, 18*, 77–88.

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clinical Neurophysiology, 115*(4), 732–744.

Donkers, F. C. L., & van Boxtel, G. J. M. (2004). The N2 in go/no-go tasks reflects conflict monitoring not response inhibition. *Brain and Cognition, 56*(2), 165–176.

Dunn, B. R., Dunn, D. A., Languis, M., & Andrews, D. (1998). The relation of ERP components to complex memory processing. *Brain and Cognition, 36*(3), 355–376.

Finnigan, S., Humphreys, M. S., Dennis, S., & Geffen, G. (2002). ERP 'old/new' effects: memory strength and decisional factor(s). *Neuropsychologia, 40*, 2288–2304.

Folstein, J. R., & van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology, 45*(1), 152–170.

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., et al. (2003). Temporal integration of speech prosody is shaped by language experience: an fMRI study. *Brain and Language, 84*(3), 318–336.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., et al. (2004). Hemispheric roles in the perception of speech prosody. *NeuroImage, 23*, 344–357.

Gandour, J., Wong, D., Lowe, M., Dzemidzic, M., Satthamnuwong, N., Tong, Y., et al. (2002). A cross-linguistic fMRI study of spectral and temporal cues underlying phonological processing. *Journal of Cognitive Neuroscience, 14*(7), 1076–1087.

Gottfried, T. L., & Suiter, T. L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics, 25*(2), 207–231.

Gu, F., Zhang, C., Hu, A., & Zhao, G. (2013). Left hemisphere lateralization for lexical and acoustic pitch processing in Cantonese speakers as revealed by mismatch negativity. *NeuroImage, 83*, 637–645.

Guillaume, C., Guillery-Girard, B., Chaby, L., Lebreton, K., Hugueville, L., Eustache, F., et al. (2009). The time course of repetition effects for familiar faces and objects: an ERP study. *Brain Research, 1248*, 149–161.

Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 819–836). Cambridge: MIT Press.

Heil, M., Osman, A., Wiegelmann, J., Rolke, B., & Hennighausen, E. (2000). N200 in the Eriksen-task: inhibitory executive process? *Journal of Psychophysiology, 14*(4), 218–225.

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

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

Hsieh, H.-I. (1976). On the unreality of some phonological rules. *Lingua, 38*, 1–19.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1–2), 101–144.

Jamison, H. L., Watkins, K. E., Bishop, D. V. M., & Matthews, P. M. (2006). Hemispheric specialization for processing auditory nonspeech stimuli. *Cerebral Cortex, 16*, 1266–1275.

Kim, K. H., Kim, J. H., Yoon, J., & Jung, K.-Y. (2008). Influence of task difficulty on the features of event-related potential during visual oddball task. *Neuroscience Letters, 445*(2), 179–183.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 11–27.

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

Lee, C.-Y. (2009). Identifying isolated, multispeaker Mandarin tones from brief acoustic input: a perceptual and acoustic study. *Journal of the Acoustical Society of America, 125*(2), 1125–1137.

Lee, C. Y., Tao, L., & Bond, Z. S. (2008). Identification of acoustically modified Mandarin tones by native listeners. *Journal of Phonetics, 36*(4), 537–563.

Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex, 15*(10), 1621–1631.

Luo, H., Ni, J.-T., Li, Z.-H., Li, X.-O., Zhang, D.-R., Zeng, F.-G., et al. (2006). Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proceedings of the National Academy of Sciences, 103*(51), 19558–19563.

Mäkelä, A. M., Alku, P., & Tiitinen, H. (2003). The auditory N1m reveals the left-hemispheric representation of vowel identity in humans. *Neuroscience Letters, 353*(2), 111–114.

Malins, J. G., & Joanisse, M. F. (2012). Setting the tone: an ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia, 50*(8), 2032–2043.

Myers, J., & Tsay, J. (2003). Investigating the phonetics of Mandarin tone sandhi. *Taiwan Journal of Linguistics, 1*, 29–68.

Okamoto, H., Stracke, H., Draganova, R., & Pantev, C. (2009). Hemispheric asymmetry of auditory evoked fields elicited by spectral versus temporal stimulus change. *Cerebral Cortex, 19*, 2290–2297.

Peng, S.-H. (2000). Lexical versus 'phonological' representations of Mandarin sandhi tones. In M. B. Broe, & J. Pierrehumbert (Eds.), *Papers in laboratory phonology V: Acquisition and the lexicon* (pp. 152–167). Cambridge: Cambridge University Press.

Penney, T. B., Mecklinger, A., & Nessler, D. (2001). Repetition related ERP effects in a visual object target detection task. *Cognitive Brain Research, 10*(3), 239–250.

Peretz, I. (1990). Processing of local and global musical information by unilateral brain-damaged patients. *Brain, 113*, 1185–1205.

Pinker, S. (1998). Words and rules. *Lingua, 106*, 219–242.

Pinker, S., Lebeaux, D. S., & Frost, L. A. (1987). Productivity and constraints in the acquisition of the passive. *Cognition, 26*, 195–267.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as “asymmetric sampling in time”. *Speech Communication, 41*(1), 245–255.

Politzer-Ahles, S., & Zhang, J. (in press). Evidence for the role of tone sandhi in Mandarin speech production. Journal of Chinese Linguistics, Monograph.

Rodriguez-Fornells, A., Münte, T. F., & Clahsen, H. (2002). Morphological priming in Spanish verb forms: an ERP repetition priming study. *Journal of Cognitive Neuroscience, 14*(3), 443–454.

Schiller, N. O. (2006). Lexical stress encoding in single word production estimated by event-related brain potentials. *Brain Research, 1112*(1), 201–212.

Schiller, N. O., Bles, M., & Jansma, B. M. (2003). Tracking the time course of phonological encoding in speech production: an event-related brain potential study. *Cognitive Brain Research, 17*, 819–831.

Schirmer, A., Tang, S.-L., Penney, T. B., Gunter, T. C., & Chen, H.-C. (2005). Brain responses to segmentally and tonally induced semantic violations in Cantonese. *Journal of Cognitive Neuroscience, 17*(1), 1–12.

Schonwiesner, M., Rubsamen, R., & von Cramon, D. Y. (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *European Journal of Neuroscience, 22*, 1521–1528.

Shen, X.-N. S. (1990). Tonal coarticulation in Mandarin. *Journal of Phonetics, 18*, 281–295.

Shen, J., Deutsch, D., & Rayner, K. (2013). On-line perception of Mandarin Tones 2 and 3: evidence from eye movements. *Journal of the Acoustical Society of America, 5*, 3016–3029.

Shestakova, A., Brattico, E., Huotilainen, M., Galunov, V., Soloviev, A., Sams, M., et al. (2002). Abstract phoneme representations in the left temporal cortex: magnetic mismatch negativity study. *NeuroReport, 13*(14), 1813–1816.

Tervaniemi, M., Medvedev, S. V., Alho, K., Pakhomov, S. V., Roudas, M. S., van Zuijen, et al.. (2000). Lateralized automatic auditory processing of phonetic versus musical information: a PET study. *Human Brain Mapping, 10*(2), 74–79.

Theunissen, F. E., & Elie, J. E. (2014). Neural processing of natural sounds. *Nature Reviews Neuroscience, 15*(6), 355–366.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear and Hearing, 22*(2), 79–90.

Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language, 97*(3), 279–293.

Wang, W. S.-Y., & Li, K.-P. (1967). Tone 3 in Pekinese. *Journal of Speech and Hearing Research, 10*, 629–636.

Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience, 170*(1), 223–231.

Xu, Y. (1991). Depth of phonological recoding in short-term memory. *Memory and Cognition, 19*, 263–273.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics, 25*, 61–83.

Yuan, J., & Chen, Y. (2014). Third tone sandhi in Standard Chinese: a corpus approach. *Journal of Chinese Linguistics, 42*(1), 218–236.

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11*, 946–953.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences, 6*, 37–46.

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1493), 1087–1104.

Zee, E. (1980). A spectrographic investigation of Mandarin tone sandhi. *UCLA Working Papers in Phonetics, 49*, 98–116.

Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology, 27*, 153–201.

Zhang, J., Lai, Y., & Sailor, C. (2011). Modeling Taiwanese speakers' knowledge of tone sandhi in reduplication. *Lingua, 121*, 181–206.

Zhang, Y., Nissen, S. L., & Francis, A. L. (2008). Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *Journal of the Acoustical Society of America, 123*(6), 4498–4513.

Zhang, C., & Peng, G. (2013). Productivity of Mandarin third tone sandhi: a wug test. In G. Peng, & F. Shi (Eds.), *Eastward flows the great river: Festschrift in honor of Prof. William S-Y. Wang on his 80th birthday* (pp. 256–282). Hong Kong: City University of Hong Kong Press.

Zhang, C., Peng, G., & Wang, W. S.-Y. (2013). Achieving constancy in spoken word identification: time course of talker normalization. *Brain and Language, 126*(2), 193–202.

Zhang, L., Xi, J., Wu, H., Shu, H., & Li, P. (2012). Electrophysiological evidence of categorical perception of Chinese lexical tones in attentive condition. *NeuroReport, 23*(1), 35–39.

Zhang, L., Xi, J., Xu, G., Shu, H., Wang, X., & Li, P. (2011). Cortical dynamics of acoustic and phonological processing in speech perception. *PLoS ONE, 6*(6), e20963.