

Chapter 7

Neural Processing of Tone Sandhi in Production and Perception: The Case of Mandarin Tone 3 Sandhi



Claire H. C. Chang and Wen-Jui Kuo

Abstract Language-specific and context-dependent phonological rules of lexical tone are prevalent in tone languages. Such rules are commonly referred to as tone sandhi. One of the most studied sandhi rules is Mandarin Tone 3 sandhi. In Mandarin, Tone 3 followed by another Tone 3 is pronounced as Tone 2 (33 → 23). In this chapter, we reviewed our current understanding of the processing of Tone 3 sandhi. Two important and relatively well-investigated questions are whether Tone 3 sandhi involves on-line tone substitution in speech production and whether the auditory representations of Tone 2 and Tone 3 are less distinct from each other due to the acquisition of Tone 3 sandhi. Recent behavioral studies demonstrated that in the lexical decision task, only Tone 3 had a facilitation effect on targets carrying tone sequence 33, while in the picture-naming task, a facilitation effect was found with both Tone 2 and Tone 3. These results supported that Tone 3 sandhi involves on-line tone substitution, in line with fMRI studies showing that Tone 3 sandhi resulted in higher activation in the right pIFG, which is known to engage in articulatory representations and their sequencing. Regarding tone perception, previous behavioral studies showed that the acquisition of Tone 3 sandhi led to worse performance at discriminating Tone 2 and Tone 3. Further, the contrast between Tone 2 and Tone 3 is consistently reported to elicit reduced MMN compared to other tone pairs only in native speakers. One explanation of these findings is that the auditory representations of Tone 2 and Tone 3 activated each other due to Tone 3 sandhi. Namely, high-level phonological rule could modulate pre-attentive auditory processing. In the future, the role of linguistic context in the processing of tone sandhi needs more investigation, especially regarding how listeners retrieve the correct word/morpheme based on the contextual information.

C. H. C. Chang

Princeton Neuroscience Institute, Princeton University, Princeton, USA

W.-J. Kuo (✉)

Institute of Neuroscience, National Yang-Ming University, Taipei, Taiwan

e-mail: wjkuo@ym.edu.tw

Brain Research Center, National Yang-Ming University, Taipei, Taiwan

© Springer Nature Singapore Pte Ltd. 2020

H.-M. Liu et al. (eds.), *Speech Perception, Production and Acquisition*,

Chinese Language Learning Sciences,

https://doi.org/10.1007/978-981-15-7606-5_7

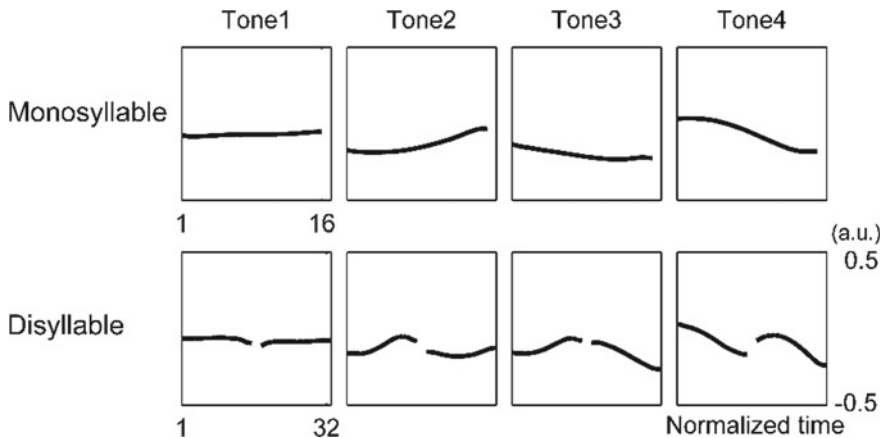


Fig. 7.1 Pitch contours of the four Mandarin lexical tones and their disyllable sequences (Chang & Kuo, 2016) (The low-falling contour of monosyllable Tone 3 in this figure is different from the falling–rising pattern in standard Mandarin, but consistent with other studies in Taiwanese Mandarin (Chang, 2010; Li, Xiong, & Wang, 2006), which might reflect the influence from Taiwanese dialect). Tone 3 sandhi is applied to disyllable Tone 3 (33 → 23)

7.1 Introduction

Phonological context-dependent tone substitution is widely found in East Asian languages (Chen, 2000) and often referred to as tone sandhi. A well-known example is Mandarin Tone 3 sandhi. Mandarin has four tones. Each syllable carries one tone. Tone 3 (T3) is pronounced as Tone 2 (T2) when it is followed by another Tone 3 (33 → 23) (Fig. 7.1) (Chao, 1948) (see also Chap. 2 in this volume). Tone 3 sandhi is a language-specific phonological rule similar to the *a/an* alternation in English (an apple vs. a dog) but much more frequent. The accumulated frequency of words inducing Tone 3 sandhi is around 1.6% in Mandarin (Academia Sinica Balanced Corpus of Modern Chinese: https://asbc.iis.sinica.edu.tw/index_readme.htm), approximating the frequency of the word “in” in English (1.5% according to Corpus of Contemporary American English: https://www.wordfrequency.info/fre_e.asp). Sandhi Tone 3 and Tone 2 are perceptually indistinguishable (Peng, 2000; Wang & Li, 1967). Because tone is used to distinguish words in tone languages, tone sandhi can result in word/morpheme ambiguity, e.g., 馬臉/ma3 ljen3/ → 麻臉/ma2 ljen3/¹ (*horse face* → *Hemp face*). In other words, for the speakers, the word they have in mind is different from the word they pronounce. For the listeners, the pronounced word is different from what they subjectively perceive.

Why does a phonological rule that induces word/morpheme ambiguity come to exist in the first place? In context, the ambiguity could be resolved with phonological, semantic, and syntactic information (Speer, Shih, & Slowiaczek, 1989, 2016), similar to the disambiguation of homophones. It is possible that tone sandhi increased the ease

¹We use international phonetic symbol (IPA) to transcribe syllable throughout this chapter.

of articulation or perception in the past but became overgeneralized and interpreted to be a categorical rule over time (Anderson, 1981; Blevins, 2006; Ohala, 1993). The pitch patterns of lexical tones in East Asian languages have undergone diachronic change and varied between dialects. Tone sandhi might have remained as a categorical phonological rule even after losing its phonetic function because of tone pattern change (Zhang & Lai, 2010).

It is worth noticing that not all sandhi rules involve the substitution of phonological representations. For example, the half Tone 3 rule in Mandarin simplifies the pitch contour of T3 but does not result in categorical change or morpheme/word ambiguity. The half Tone 3 rule is believed to reflect the universal demand on the ease of articulation (Xu, 2004), whose application is less dependent on language experience. Indeed, the application of Tone 3 sandhi has been reported to appear later and less accurate during development and it is also hard for second language learners (Chen, Wee, Tong, Ma, & Li, 2016).

In this chapter, we focus on Mandarin Tone 3 sandhi, one of the most studied sandhi rules. Speech production and perception involve different neural processing. Therefore, we discuss tone sandhi in production and perception respectively. A rough delineation of the processing of speech production and perception according to current speech models is as below (Golfinopoulos, Tourville, Guenther, & Gol, 2010; Hickok & Poeppel, 2007a; Indefrey & Levelt, 2004; Price, 2010). In speech production, the motor representations of speech sounds are activated and sequenced in the posterior inferior frontal gyrus (pIFG) and premotor areas and executed in the motor cortex. The auditory feedback of the articulation is then processed in superior temporal gyrus (STG) as part of the self-monitoring process. In speech perception, the auditory inputs activate the categorical auditory representations in the STG/STS, which in turn lead to the retrieval of the lexical representations in the lower part of the temporal lobe. Motor representations are not necessarily involved in speech perception (Scott, McGettigan, & Eisner, 2009).

The traditional description of Tone 3 sandhi (33 → 23) is more from the production perspective. If Tone 3 sandhi does involve the substitution of motor representations of tones, the literature suggests that pIFG/premotor areas should be engaged, since they are responsible for the storage and sequencing of categorical motor representations (Golfinopoulos et al., 2010; Hickok & Poeppel, 2007a; Indefrey & Levelt, 2004; Price, 2010). In this case, the next question is how the discrepancy between the underlying and surface tones escapes self-monitoring. Concerning tone perception, the behavioral finding that native Mandarin speakers were prone to confuse T2 and T3 even under monosyllable condition raises the question of whether Tone 3 sandhi, a high-level phonological rule, can modulate early auditory processing. Furthermore, the morpheme/word ambiguity resulted from the application of Tone 3 sandhi must be resolved in the later stage based on contextual information, including the following tone, word boundary, phrase structure, etc. (see Chap. 3 in this volume for a discussion on the role of linguistic context on tone perception), and we still know very little about the neural mechanism underlying this disambiguation process. These issues are discussed in the following sections.

7.2 Tone 3 Sandhi in Speech Production

7.2.1 Behavioral Studies

The claim that Tone 3 sandhi is a language-specific phonological rule that substitutes the underlying Tone 3 by the surface Tone 2 is supported by the finding that T3, but not T2, primed targets carrying tone sequence 33 in the lexical decision task (Chien, Sereno, & Zhang, 2016), while in the picture-naming task, T2 and T3 both induced a facilitation effect (Nixon, Chen, & Schiller, 2015). Chien et al. (2016) conducted an auditory-auditory priming lexical decision experiment using disyllabic word targets and legal monosyllable primes of T1, T2, or T3 (e.g., /fu1/, /fu2/, and /fu3/). The prime preceded the target by 250 ms. The critical targets consisted of two Tone 3 syllables (e.g., /fu3 tao3/ 輔導). They demonstrated that T3 significantly facilitated targets carrying tone sequence 33. Namely, these targets had shorter reaction times (RTs) with T3 prime than T1 prime. No facilitation effect was found for T2 prime. These findings indicated that only the underlying T3 but not the surface T2 was involved in the lexical decision task. A similar effect has also been reported for Taiwanese tone sandhi pair (Chien, Sereno, & Zhang, 2017).

In contrast, Nixon et al. (2015) adopted the picture naming instead of the lexical decision task. The participants were asked to name a picture, and a word distractor was presented visually 0 ms or 83 ms after the picture. The target pictures had disyllable names. The distractors were semantically and orthographically unrelated to the targets, while the phonological relationship between the picture names and the distractor words was manipulated. Experiment 1 used monosyllable distractors (e.g., 驢、屨、綠). For picture names consisting of two T3 syllables, a facilitation effect was found for both T2 and T3 distractors. Namely, naming RTs were shorter in trials with T2 and T3 distractors than trials with control (T1/T4) distractors, indicating that the production of tone sequence 33 involved the phonological representations of both T2 and T3. In Experiment 2, the first syllable of the picture names carried either T2 or T3 and the tone of the second syllable was not limited to Tone 3 (2X vs. 3X, e.g., 浮標 vs. 武器). The distractors were disyllabic words carrying tone sequences 33 (e.g., 雨傘) or control sequences (1X or 4X, e.g., 夫婦 and 噪音). Distractors carrying tone sequence 33 facilitated the naming of both sequence 2X and 3X, indicating that distractor words carrying tone sequence 33 activated the phonological representations of both T2 and T3. The effect of distractor type (Exp. 1) or target type (Exp. 2) did not interact with the onset time of the distractors. Taken together, these findings indicated that for words carrying tone sequence 33, only T3 is stored in the lexicon, while the phonological representations of T2 and T3 were both activated for the production of tone sequence 33.

7.2.2 *Neuroimaging and Electrophysiological Studies*

Where is Tone 3 sandhi implemented in the brain during speech production? Using functional magnetic imaging (fMRI), Chang and Kuo (2016) and Chang et al. (2014) examined the production of sequences of the four Mandarin lexical tones. The participants were required to pronounce visually displayed phonetic symbols in the scanning sessions. Sixteen tonal syllables were used (four tones x four vowels /a/, /i/, /u/, and /y/). Tones in one sequence were borne by the same vowel. Larger brain activations in the right pIFG for Tone 3 sequence (e.g., 33 > 11, 22, 44) was found. It was suggested that right pIFG was involved in the implementation of Tone 3 sandhi.

It has been debated whether the underlying and the surface tones are both stored for words involving tone sandhi (e.g., Hsieh, 1970; Tsay & Myers, 1996) or only the underlying tone is stored, which is substituted by the surface tone on-line before articulation. Brain imaging literature suggests that the phonological representations of words reside in the temporal lobe (Hickok & Poeppel, 2007a; Indefrey & Levelt, 2004), while the frontal lobe is engaged in on-line phonological processing and articulation. Therefore, the finding of higher IFG activation during sandhi tone production supports that Mandarin Tone 3 sandhi requires on-line tone substitution, consistent with recent behavioral studies (Chien et al., 2016; Nixon et al., 2015).

One concern with this interpretation is that T3 might be physically harder to pronounce because it has the most complicated contour (falling–rising) among the four Mandarin lexical tones, at least in standard Mandarin. However, in that case, extra right IFG activation for Tone 3 should also be observed with monosyllable stimuli. Chang et al. included both monosyllable and disyllable conditions. Tone 3 sandhi only applied under the disyllable condition. They found higher brain activations for Tone 3 only under the disyllable condition, indicating that the higher activation in the right IFG for sequence 33 did not only reflect the inherent physical difficulty in producing Tone 3.

Because repeated sequence 33 was pronounced as mixed sequence 23 on the surface, another concern is that right pIFG is involved in the production of any mixed sequence, no matter whether Tone 3 sandhi is applied or not. Mixed sequences might increase the processing loading for tone retrieval and sequencing. Mixed sequences might also require extra computation for co-articulation and change of pitch direction (Xu & Emily Wang, 2001; Xu & Xu, 2005). Chang et al. (2014) contrasted “genuine” mixed sequences (twelve of them, e.g., 2413) and sandhi sequence 3333 against repeated sequences (1111, 2222, and 4444) respectively. Additional activation in the right posterior IFG was only observed for sequence 3333. Chang et al. also manipulated the requirement on the overt oral response in order to distinguish the pre-articulatory planning and the motor execution stages of speech production. Higher right pIFG response to sequence 33 was observed only under overt production condition, indicating that the application of Tone 3 sandhi depends on overt production.

The implementation of Tone 3 sandhi during speech production has also been investigated with event-related potential (ERP) technique. Zhang et al. (2015) directly

compared the production of tone sequence 23 and 33. Since both sequences were pronounced as 23 on the surface, the difference between them cannot be due to articulatory or acoustic difference and is more likely to reflect the implementation of Tone 3 sandhi. It was reported that sequence 33 elicited larger P2 (230–320 ms) than sequence 23, consistent with the claim that Tone 3 sandhi requires additional processing. Furthermore, this effect was found under both real word, and pseudoword conditions (legal vs. illegal syllable), supporting that Tone 3 sandhi involves on-line computation instead of the retrieval of an alternative phonological representation of a word. One advantage of the ERP method is its higher temporal resolution. However, in this study, the participants were required to repeat the auditorily presented stimuli covertly upon hearing the second syllable and to produce them overtly upon seeing a visual cue 1000–1600 ms after the offset of the auditory stimuli. The ERPs were time-locked to the onset of the second syllable of the stimuli. Because of the experimental procedure used, this study might be less informative about the time course of natural speech production.

The right auditory cortex is known to be specialized in pitch perception (Jamison, Watkins, Bishop, & Matthews, 2006; Poeppel, 2003; Schönwiesner, 2005; Shtyrov, Kujala, Palva, Ilmoniemi, & Näätänen, 2000; Zatorre, 2001). The right IFG could be recruited for tone processing through its interaction with the right auditory cortex (Kell, Morillon, Kounieher, & Giraud, 2011; Pulvermüller, Kiff, & Shtyrov, 2012). Based on findings in pitch without linguistic function (Jamison et al., 2006; Poeppel, 2003; Schönwiesner, 2005; Shtyrov et al., 2000; Zatorre, 2001), a functional asymmetry between the left and right auditory cortices has been proposed. Zatorre (2001) suggested that the left auditory areas have a better temporal resolution, while the right auditory areas have a better spectral resolution. The asymmetric sampling in time hypothesis, on the other hand, proposed that the left auditory areas extract information from short (~20–40 ms) temporal integration windows, while the right auditory areas extract information from long (~150–250 ms) integration windows (Poeppel, 2003).

During speech production, the interaction between the frontal and temporal regions is necessary for self-monitoring and error correction (Guenther, Ghosh, & Tourville, 2006; Hickok, 2012; Hickok & Poeppel, 2007b), namely, to identify the discrepancy between the expected output and the auditory feedback. If the auditory feedback deviates from the expectation, the mapping between phonological representations and motor commands needs to be adjusted accordingly. Therefore, the interaction between the motor system in the frontal areas and the auditory system in the temporal areas is crucial for speech production, especially during development or when speech production is perturbed (Flagmeier et al., 2014).

Since right auditory cortex specializes in pitch perception, right IFG might be recruited for tone processing through its interaction with right auditory cortex via right arcuate fasciculus. Using fMRI, Liu et al. (2006) compared the production of Mandarin tones and vowels in the character-naming and the pinyin-naming tasks. Both included sixteen tonal syllables (4 tones \times 4 vowels /a/, /ə/, /i/, and /u/ for the pinyin-naming task/ɿ a/, /ɿ ə/, /ɿ ɿ /, and /ɿ u/ for the character-naming task). Higher brain activations in the right IFG for tone production than vowel production

were found in both tasks, while higher activations for vowel than tone were found exclusively in the left hemisphere. These findings support that right IFG is more important for tone production. Further, structural and functional anomalies in right IFG (Albouy et al., 2013; Hyde et al., 2007; Hyde, Zatorre, & Peretz, 2011), right STG (Albouy et al., 2013; Zhang, Peng, Shao, & Wang, 2017), and the right frontal–temporal pathway (Loui, Alsop, & Schlaug, 2009; Wang, Zhang, Wan, & Peng, 2017) have been reported in patients with congenital amusia (Peretz, 2013), an impairment to process music melody as well as lexical tone (Jiang, Hamm, Lim, Kirk, & Yang, 2012; Liu et al., 2012, 2016; Nan, Sun, & Peretz, 2010; Tillmann et al., 2011).

In the case of Tone 3 sandhi, the updated phonological representation/motor command must help to generate the prediction on auditory feedback, so the discrepancy between underlying and surface tones would not alert the self-monitoring system during speech production. This scenario is consistent with the finding of a larger right pIFG response to 33 sequence only when overt production was required (Chang & Kuo, 2016). In parallel to the finding in tone, Loui, Li, & Schlaug, (2011) have created a pitch-based artificial rule and found that the participants' learning performance positively correlated with the volumes of the right arcuate fasciculus connecting the right IFG and the superior temporal lobe.

In sum, studies in Tone 3 sandhi production suggest that Tone 3 sandhi implementation requires extra on-line computation in the right IFG, which might be involved through interaction with the right auditory cortex.

7.3 Tone 3 Sandhi in Tone Perception

7.3.1 Behavioral Studies

Tone 3 sandhi results in a discrepancy between the pronounced and perceived tones without alerting the attention and the self-monitoring systems. That raises a naïve question: are the auditory representations of T2 and T3 less distinctive from each other after the acquisition of Tone 3 sandhi? Among the six possible tone pairs in Mandarin, T2-T3 was often reported to be one of the most difficult pairs to distinguish even for non-native speakers (Hao, 2018; Huang & Johnson, 2011; So & Best, 2014). Because non-native speakers do not know the sandhi rule, acoustic similarity is more likely the reason for their difficulty.

However, several behavioral studies have demonstrated that T2 and T3 were even more similar to each other for native speakers than for non-native speakers (Chen, Liu, & Kager, 2015, 2016; Huang & Johnson, 2011). Huang & Johnson (2011) recruited both Chinese and English speakers in two Mandarin tone discrimination experiments. One used speech sound stimuli (legal Mandarin monosyllable /pa/) and the other used sine-wave stimuli. All six possible tone pairs were included. Using speech sound, Chinese speakers generally discriminated Mandarin tones faster than English speakers. They were significantly slower than the English group only

at discriminating T2 and T3. T2-T3 also elicited the longest RT among all tone pairs in the Chinese but not the English group. Further, such group difference in discriminating T2 and T3 was not observed under non-speech condition.

Chen, Liu et al. (2016) recruited Dutch and Chinese speakers in a Mandarin tone discrimination experiment. Their task was to discriminate T3 from T2 and T4 from T1 under the monosyllable and the disyllable conditions. The monosyllable stimuli were legal Mandarin syllables. The disyllable stimuli consisted of two legal monosyllables that did not form a real word. The results showed that Dutch speakers outperformed Chinese speakers at discriminating tone sequence 33 from sequences containing T2 (23, 32, and 22) (77% accuracy for the Chinese group and 82% for Dutch). Such group difference was not found under the monosyllable condition (2 vs. 3) or in the T1-T4 pair (1 vs. 4 under the monosyllable condition. 44 vs. 41, 11, and 14 under the disyllable condition). A similar result was also reported in Chen et al. (2015).

The results of Huang and Johnson (2011) and Chen et al. (2015, 2016) are surprising because acquiring a language unusually leads to better performance in discriminating acoustically similar but linguistically distinctive sounds. These results indicate that acoustic similarity and Tone 3 sandhi might both account for the difficulty of discriminating T2 and T3 (Hume & Johnson, 2001). Huang and Johnson (2011) used monosyllable stimuli, implying that the confusion between T2 and T3 occurred in the early context-independent stage of auditory processing, while Chen, Liu et al. (2016) reported lower accuracy in native speakers than non-native speakers at discriminating T2 and T3 only under the disyllable condition, indicating that a viable context for Tone 3 sandhi is critical. The early automatic stage of auditory processing can be examined by the mismatch negativity (MMN) paradigm, which is discussed in the next section.

7.3.2 *Neuroimaging and Electrophysiological Studies*

MMN is an ERP component often elicited using the oddball paradigm, in which a standard sound is displayed with higher probability and a deviant sound with lower probability (Näätänen, Paavilainen, Rinne, & Alho, 2007). MMN is found around 100–300 ms after stimulus onset in the difference waveform of the deviant minus the standard and believed to reflect the automatic detection of sound change, namely the difference between the memory trace of the standard and the current deviant input. Phonological rules of phoneme change such as place assimilation (e.g., /d/ to /b/ in “bad boy”) have been reported to modulate MMN (Mitterer & Blomert, 2003; Mitterer, Csépe, Honbolygo, & Blomert, 2006; Sun et al., 2015; Tavabi, Elling, Dobel, Pantev, & Zwitserlood, 2009). Namely, MMN elicited by phoneme change was reduced if the change could be explained by place assimilation rule.

Previous studies in Mandarin have demonstrated that MMN elicited by the contrast between T2 and T3 was lower in amplitude and longer in peak latency than the non-sandhi tone pairs (e.g., T1-T3) (Chandrasekaran, Gandour, & Krishnan, 2007;

Chandrasekaran, Krishnan, & Gandour, 2007; Cheng et al., 2013; Li & Chen, 2015; see also Chap. 6 in this volume). Similar results were also reported in the magnetoencephalographic counterpart of MMN (Hsu, Lin, Hsu, & Lee, 2014). Chandrasekaran, Gandour, et al. (2007) recruited both English and Chinese speakers and included three Mandarin tone pairs (T1-T3, T2-T3, and T1-T2). The Chinese group showed larger MMN amplitude than the English group for T1-T2 and T1-T3, indicating higher sensitivity to tone difference in native speakers. As for T2-T3, no language group effect was found. Further, for the Chinese group, the MMN amplitude of T2-T3 was significantly smaller than T1-T2 and T1-T3, while no tone pair difference was found for the English group. Similar findings were also reported in Chandrasekaran, Krishnan, et al. (2007), which compared T2-T3 and T1-T3.

Taking the MMN amplitude as an index of the dissimilarity between tones, these results indicated that T2 and T3 were more similar to each other than they were to T1 for the Chinese but not the English group. T1 has a flat pitch contour, while T2 and T3 both have non-flat pitch contours. Chandrasekaran, Gandour, et al. (2007) thus suggested that native speakers are more sensitive to the distinction between flat and non-flat tones and that explained their findings. This acoustic similarity account is the most commonly held one for the weaker MMN elicited by the contrast between T2 and T3 (Chandrasekaran, Gandour, et al. 2007; Chandrasekaran, Krishnan, et al. 2007; Cheng et al., 2013; Hsu et al., 2014; Yu, Shafer, & Sussman, 2017).

However, although T2 and T3 both have non-flat pitch contours, they differ in the direction (Fig. 7.1) and previous studies have reported that Mandarin speakers were more sensitive to pitch direction than English speakers (Gandour, 1983, 1984). In addition, acoustic similarity can barely explain the behavioral findings that T2 and T3 were perceptually more similar to native speakers than to non-native speakers (A. Chen et al., 2015, 2016; Huang & Johnson, 2011). No matter how similar two speech sounds in a language are along a specific acoustic dimension, it is unlikely that learning the language could increase the difficulty in distinguishing them. Therefore, the alternative Tone 3 sandhi account is worth more consideration and examination (Li & Chen, 2015). The MMN response has been proposed to reflect the discrepancy between the deviant sound and the short-term memory trace of the standard sound (Näätänen et al., 2007). If the Mandarin T3 standard activates the phonological representations of both T3 and T2, then the deviant T2 may result in less discrepancy.

Yet another explanation for the reduced MMN elicited by the contrast between T2 and T3 comes from the underspecification theory (Archangeli, 1988). According to this theory, some phonemes are not fully represented in memory, and that is why they are often replaced or assimilated by other phonemes. Reduced MMN has been reported using underspecified vowel as the standard sound and suggested to reflect less conflict at the phonological level (Cornell, Lahiri, & Eulitz, 2011; Eulitz & Lahiri, 2004; Scharinger, Monahan, & Idsardi, 2016). Politzer-Ahles et al. (2016) recruited both native and non-native Mandarin speakers. Hypothesizing that T3 is phonological underspecified, they predicted reduced MMN when T3 served as the standard compared to when it served as the deviant in the Mandarin group. They reasoned that the phonological representation of a standard sound lasts longer than its surface features. Therefore, when an underspecified sound serves as the standard,

its phonological representation conflicts less with the incoming deviant sound. On the other hand, when an underspecified sound serves as the deviant, its acoustic features conflict with the fully specified phonological representation of the standard sound. The predicted effect was observed in Experiment 3, which included all six possible Mandarin tone pairs. However, a closer examination into all the tone pairs containing T3 (T1-T3, T2-T3, T4-T3) showed a significant asymmetry only in the T2-T3 pair. Namely, standard T3 and deviant T2 elicited smaller MMN than standard T2 and deviant T3. In addition, such asymmetry was also reported in non-native speakers in Experiment 1 & 2 and tone pairs without T3 (for pair T2-T4, smaller MMN was found when T2 served as the standard) in Experiment 1. Therefore, the interpretation of these results is not clear.

As far as we know, none of the previous imaging studies in tone perception has directly compared sandhi and non-sandhi conditions. Nevertheless, studies focusing on the lateralization of tone perception serve to clarify the role of right IFG. It has been suggested that speech production and perception involve similar neural circuits (D'Ausilio et al., 2009; Galantucci, Fowler, & Turvey, 2006; Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Scott et al., 2009). Since right IFG has been reported to engage in tone production (Chang & Kuo, 2016; Chang et al., 2014; Liu et al., 2006), its role in tone perception is worth examining. The fMRI study of Li et al. (2010) adopted an auditory matching task. The participants were presented with a sequence of three legal Mandarin syllables and asked to judge whether any of them matches the following monosyllable probe, e.g., /pau1 xuən4 mu2/-/tʂɿ1/ (a yes trial in the tone matching task). The position of the target within the trisyllable sequence was randomly assigned, in order to increase the processing loading of brain regions involved in phonological encoding and working memory. Taking fixed target position condition as the baseline, they found higher activations in the right pIFG and right inferior parietal lobule in tone matching task than in consonant or rime matching task. Right IFG activation has also been reported in tone judgment task with visually presented Chinese characters (whether the reading of the character has Tone 4), using arrow judgment task as the baseline condition (Kwok et al., 2015). These findings showed that right IFG also plays a role in tone perception tasks.

It is worth noticing that the finding that right hemisphere is more important for the processing of tone than the other phonological units (Li et al., 2010; Liu et al., 2006; Luo et al., 2006) does not necessarily contradict with the argument that experience in tone language leads to more reliance on the left hemisphere and the left-lateralization of tone processing (Zatorre & Gandour, 2008; see also Chap. 5 in this volume). Increased activation in left frontal, parietal, and insular regions have been reported in studies comparing native versus non-native speakers (Gandour et al., 2003, 2000; Hsieh, Gandour, Wong, & Hutchins, 2001; Klein, Zatorre, Milner, & Zhao, 2001; Wong, Parsons, Martinez, & Diehl, 2004) and tone vs. non-speech pitch (Gandour et al., 2000; Hsieh et al., 2001; Wong et al., 2004) in auditory discrimination task. Here we point out that such results are not incompatible with the finding of higher reliance on the right hemisphere in the processing of tone than the other phonological units, as demonstrated in Fig. 7.2.

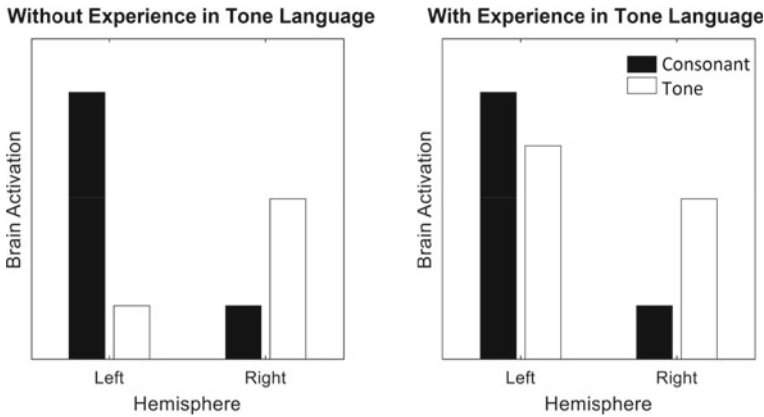


Fig. 7.2 Hypothetical brain responses to consonant and tone in the left and right hemisphere with and without experience in tone languages

To the aim of examining the phonological processing of tone in natural speech perception, most existing neuroimaging studies suffered from the confound of lexical processing or task-relevant effect. Because all legal monosyllables in Mandarin have corresponding words/morphemes, using legal monosyllable stimuli inevitably introduced the confound of lexical processing in the contrast between native and non-native speakers and the contrast between tone versus non-speech pitch (Gandour et al., 2003, 2000; Hsieh et al., 2001; Klein et al., 2001; Kwok et al., 2015; Nan & Friederici, 2013; Wong et al., 2004). Further, all active tasks introduced task-specific effect, e.g., verbal working memory and selective attention, especially when using passive listening condition as the baseline (Gandour et al., 2003; Hsieh et al., 2001; Wong et al., 2004), in which case task-specific component was more likely to survive baseline subtraction. Future imaging studies need to take these issues into consideration.

In brief, existing behavior and ERP evidence imply that T2 and T3 are less distinct from each other in the pre-attentive stage of auditory processing, but more researches are needed to better disentangle the acoustic similarity account from the Tone 3 sandhi account. In the future, how the listeners overcome the discrepancy between surface and underlying tones based on contextual information in the later stage of auditory processing, so to retrieve the right word/morpheme, needs to be investigated for a deeper understanding of Tone 3 sandhi.

7.4 General Discussion

This chapter reviews our current understanding of Tone 3 sandhi, including its implementation during speech production and, regarding tone perception, whether the acquisition of Tone 3 sandhi affects the pre-attentive auditory processing of tone.

The results of existing behavioral studies supported that the underlying T3 is stored in the lexicon (Chien et al., 2016) and the representations of T2 and T3 are both activated for the production of T3 sequences (Nixon et al., 2015). fMRI studies of tone production (Chang & Kuo, 2016; Chang et al., 2014; Liu et al., 2006) have demonstrated that right pIFG was involved in the processing of tone, supporting that Tone 3 sandhi involves on-line substitution of neural representations. Since the right auditory cortex is known to be specialized in pitch perception, right IFG might be recruited for tone processing through its interaction with the right auditory cortex.

One way to further examine the frontal–temporal interaction in tone production is to perturbate the auditory feedback, which supposedly increases the loading on the self-monitoring system. Larger activation in the right IFG activation and bilateral temporal cortices (Fu et al., 2006) and increased functional connectivity in the right temporal–frontal loop (Flagmeier et al., 2014) have been reported with pitch-shifted auditory feedback in English. In the fMRI study of Fu et al. (2006), the participants were asked to pronounce visually presented real words. Their speech was lowered in pitch by 4 semitones under self-distorted condition. Compared to self-undistorted condition, distorted feedback elicited higher activations in bilateral temporal cortices and right IFG. However, perturbation involving vowel change was also reported to increase IFG activation bilaterally (Niziolek & Guenther, 2013; Zheng et al., 2013) or in the right hemisphere (Tourville, Reilly, & Guenther, 2008). Direct comparison of different types of perturbation, e.g., consonant, vowel, non-lexical pitch, lexical tone, etc., might help to clarify whether right IFG is more engaged in self-monitoring during tone production.

As for tone perception Huang & Johnson (2011), and Chen (2015, 2016) demonstrated that native speakers were slower or less accurate in discriminating T2 and T3 than non-native speakers. One explanation is that acquiring Tone 3 sandhi leads to the co-activation of T2 and T3, which is consistent with the finding of reduced MMN elicited by the contrast between T2 and T3 in the native speakers (Chandrasekaran, Gandour, et al., 2007; Chandrasekaran, Krishnan, et al., 2007; Cheng et al., 2013; Hsu et al., 2014; Li & Chen, 2015). However, in studies comparing tone pairs, the effect of Tone 3 sandhi could hardly be disentangled from that of acoustic similarity or under-specified phonological representation since Mandarin only has four tones and six possible tone pairs. To further investigate how language-specific phonological rule modifies auditory processing, alternative solutions include the comparison between participants with different language backgrounds (Chang, Lin, & Kuo, 2019) and systematic manipulation of linguistic context and inter-stimulus-interval (ISI).

Previous behavioral studies suggested that the influence of language experience on tone perception might be context-dependent. English speakers discriminated Mandarin tones carried by sine waves (non-linguistic context) better than Chinese speakers (Huang & Johnson, 2011). Chen, Liu et al. (2016) reported that Dutch speakers outperformed Chinese speakers at discriminating T2 and T3 carried by disyllabic stimuli, which provided a viable context for Mandarin Tone 3 sandhi (33 → 23). The interaction between linguistic context and phonological rule in the MMN paradigm has been studied using segments. Sun et al. (2015) examined the MMN elicited by /f/ to /v/ change in French. French /f/ is a voiceless sound, while /v/ is a

voiced one. The change from /f/ to /v/ is legal when a voiced obstruent consonant follows /f/. Utilizing this optional but language-specific voicing assimilation rule, Sun et al. (2015) compared ERP elicited by /f/ to /v/ change under viable (/ofbe/ → /ovbe/) and unviable context (/ofne/ → /obne/). The ERP analysis was time-locked to the onset of /f/ or /v/. They found MMN and P300 only for voicing change under context unviable for the voicing assimilation rule, supporting that the representations of /f/ and /v/ were both activated when the context was viable for the voicing assimilation rule. These results demonstrated that linguistic context influenced the effect of phonological rule on MMN.

ISI has also been proposed to influence the effect of language experience on MMN. Yu et al. (2017) manipulated ISI and suggested that long ISI could diminish the effect of short-term sensory memory trace and thus reveal the processing of the long-term phonological representations. They used disyllable stimuli that differed only in the first tone and reported that MMN elicited by tone change was evident in both Chinese and English groups under short ISI condition, while under long ISI condition, only the Chinese group showed the MMN response. These findings supported that ISI could be used to examine the influence of language experience and to disentangle the acoustic/phonetic and the phonological stages of auditory processing.

Another interesting result from Yu et al. (2017) is that, unlike previous MMN studies using monosyllable stimuli (Chandrasekaran, Gandour, et al., 2007; Chandrasekaran, Krishnan, et al., 2007; Cheng et al., 2013; Hsu et al., 2014; Li & Chen, 2015), the contrast between T2 and T3 did not yield reduced MMN or lower discrimination accuracy, which might result from the inviable context for Tone 3 sandhi, i.e., tone sequence 31. Yu et al. (2017) compared the discrimination of tone sequence 31 from sequence 21 and 11. The Chinese group showed similar accuracies (both above 90%) and outperformed the English group under both conditions. When using sequence 31 as the standard, MMN elicited by deviant 21 was as strong as MMN elicited by deviant 11, with either short ISI or long ISI. Such results are in line with the idea that viable linguistic context is crucial for Tone 3 sandhi effect.

In the future, the role of linguistic context on the production and perception of Tone 3 sandhi needs more systematic investigations. Furthermore, the nature of tone sandhi depends on the exact rule in question (Chien et al., 2017; Myers & Tsay, 2003; Xu, 2004; Zhang & Lai, 2010; Zhang & Liu, 2016) and varies between languages (Chen, 2000; Tsay & Myers, 1996). This chapter focuses on Mandarin Tone 3 sandhi. Whether a general neural mechanism is shared across sandhi rules and tone languages requires further tests in the future (Chang et al., 2019; Chien et al., 2017).

References

- Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P. E., Tillmann, B. (2013). Impaired pitch perception and memory in congenital amusia: The deficit starts in the auditory cortex. *Brain*, 136(5), 1639–1661. <https://doi.org/10.1093/brain/awt082>.
- Anderson, S. R. (1981). Why phonology isn't "natural." *Linguistic Inquiry*, 12(4), 493–539. Retrieved from <https://escholarship.org/uc/item/7b6962b4#page-39>.
- Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, 5(02), 183–207. <https://doi.org/10.1017/S0952675700002268>
- Blevins, J. (2006). A theoretical synopsis of evolutionary phonology. *Theoretical Linguistics*, 32(2), 117–166. <https://doi.org/10.1515/TL.2006.009>
- Chandrasekaran, B., Gandour, J. T., & Krishnan, A. (2007). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative Neurology and Neuroscience*, 25(3–4), 195–210. <https://doi.org/10.1016/j.ygyno.2014.12.035>. <https://doi.org/10.1016/j.pharmac.2014.12.035>
- Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research*, 1128(1), 148–156. <https://doi.org/10.1016/j.brainres.2006.10.064>
- Chang, C. Y. (2010). *Dialect differences in the production and perception of Mandarin Chinese tones*. The Ohio State University.
- Chang, C. H. C., & Kuo, W. J. (2016). The neural substrates underlying the implementation of phonological rule in lexical tone production: An fMRI study of the tone 3 sandhi phenomenon in Mandarin Chinese. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0159835>
- Chang, C. H. C., Lee, H. J., Tzeng, O. J. L., & Kuo, W.-J. (2014). Implicit target substitution and sequencing for lexical tone production in Chinese: An fMRI study. *PLoS ONE*, 9(1). <https://doi.org/10.1371/journal.pone.0083126>.
- Chang, C. H. C., Lin, T. H., & Kuo, W. J. (2019). Does phonological rule of tone substitution modulate mismatch negativity? *Journal of Neurolinguistics*, 51, 63–75. <https://doi.org/10.1016/j.jneuroling.2019.01.001>
- Chao, Y. R. (1948). *Mandarin primer*. Cambridge (UK): Harvard University Press. <https://doi.org/10.4159/harvard.9780674732889>.
- Chen, A., Liu, L., & Kager, R. (2015). Cross-linguistic perception of Mandarin tone sandhi. *Language Sciences*, 48, 62–69. <https://doi.org/10.1016/j.langsci.2014.12.002>
- Chen, A., Liu, L., & Kager, R. (2016). Cross-domain correlation in pitch perception, the influence of native language. *Language, Cognition and Neuroscience*, 31(6), 751–760. <https://doi.org/10.1080/23273798.2016.1156715>
- Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects*. Cambridge University Press.
- Chen, N. F., Wee, D., Tong, R., Ma, B., & Li, H. (2016). Large-scale characterization of non-native Mandarin Chinese spoken by speakers of European origin: Analysis on iCALL. *Speech Communication*, 84, 46–56. <https://doi.org/10.1016/j.specom.2016.07.005>
- Cheng, Y.-Y., Wu, H.-C., Tzeng, Y.-L., Yang, M.-T., Zhao, L.-L., & Lee, C.-Y. (2013). The development of mismatch responses to Mandarin lexical tones in early infancy. *Developmental Neuropsychology*, 38(5), 281–300. <https://doi.org/10.1080/87565641.2013.799672>
- Chien, Y.-F., Sereno, J. A., & Zhang, J. (2016). Priming the representation of Mandarin tone 3 sandhi words. *Language, Cognition and Neuroscience*, 31(2), 179–189. <https://doi.org/10.1080/23273798.2015.1064976>
- Chien, Y.-F., Sereno, J. A., & Zhang, J. (2017). What's in a word: Observing the contribution of underlying and surface representations. *Language and Speech*, 60(4), 643–657. <https://doi.org/10.1177/0023830917690419>
- Cornell, S. A., Lahiri, A., & Eulitz, C. (2011). "What you encode is not necessarily what you store": Evidence for sparse feature representations from mismatch negativity. *Brain Research*, 1394, 79–89. <https://doi.org/10.1016/J.BRAINRES.2011.04.001>

- D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, *19*(5), 381–385. <https://doi.org/10.1016/j.cub.2009.01.017>
- Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience*, *16*, 577–583. <https://doi.org/10.1162/089892904323057308>
- Flagmeier, S. G., Ray, K. L., Parkinson, A. L., Li, K., Vargas, R., Price, L. R., Robin, D. A. (2014). The neural changes in connectivity of the voice network during voice pitch perturbation. *Brain and Language*, *132*, 7–13. <https://doi.org/10.1016/j.bandl.2014.02.001>
- Fu, C. H. Y., Vythelingum, G. N., Brammer, M. J., Williams, S. C. R., Amaro, E., Andrew, C. M., McGuire, P. K. (2006). An fMRI study of verbal self-monitoring: Neural correlates of auditory verbal feedback. *Cerebral Cortex*, *16*(7), 969–977. <https://doi.org/10.1093/cercor/bhj039>
- Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–377. <https://doi.org/10.3758/BF03193857>
- Gandour, J. T. (1983). Tone perception in far eastern-languages. *Journal of Phonetics*, *11*(2), 149–175.
- Gandour, J. T. (1984). Tone dissimilarity judgments by Chinese Listeners. *Journal of Chinese Linguistics*, *12*(2), 235–261. Retrieved from <https://www.jstor.org/stable/23767002>
- Gandour, J. T., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., Lurito, J. (2003). Temporal integration of speech prosody is shaped by language experience: an fMRI study. *Brain and Language*, *84*(3), 318–336. [https://doi.org/10.1016/S0093-934X\(02\)00505-9](https://doi.org/10.1016/S0093-934X(02)00505-9)
- Gandour, J. T., Wong, D., Hsieh, L., Weinzapfel, B., Lancker, D. V., & Hutchins, G. D. (2000). A crosslinguistic PET study of tone perception. *Journal of Cognitive Neuroscience*, *12*(1), 207–222. <https://doi.org/10.1162/089892900561841>
- Golfinopoulos, E., Tourville, J. A. A., Guenther, F. H. H., & Gol, E. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *NeuroImage*, *52*(3), 862–874. <https://doi.org/10.1016/j.neuroimage.2009.10.023>
- Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, *96*(3), 280–301. <https://doi.org/10.1016/j.bandl.2005.06.001>
- Hao, Y.-C. (2018). Second language perception of Mandarin vowels and tones. *Language and Speech*, *61*(1), 135–152. <https://doi.org/10.1177/0023830917717759>
- Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135–145. <https://doi.org/10.1038/nrn3158>
- Hickok, G., & Poeppel, D. (2007a). The cortical organisation for speech processing. *Nature*, *8*(May), 393–402. <https://doi.org/10.7554/eLife.14521>
- Hickok, G., & Poeppel, D. (2007b). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. <https://doi.org/10.1038/nrn2113>
- Hsieh, H.-I. (1970). The psychological reality of tone sandhi rules in Taiwanese. In *Papers From the 6th Annual Regional Meeting of the Chicago Linguistic Society* (pp. 489–503). Retrieved from <https://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+psychological+reality+of+tone+sandhi+rules+in+Taiwanese#0>
- Hsieh, L., Gandour, J. T., Wong, D., & Hutchins, G. D. (2001). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain and Language*, *76*(3), 227–252. <https://doi.org/10.1006/brln.2000.2382>
- Hsu, C. H., Lin, S. K., Hsu, Y. Y., & Lee, C. Y. (2014). The neural generators of the mismatch responses to Mandarin lexical tones: An MEG study. *Brain Research*, *1582*, 154–166. <https://doi.org/10.1016/j.brainres.2014.07.023>
- Huang, T., & Johnson, K. (2011). Language specificity in speech perception: Perception of mandarin tones by native and nonnative listeners. *Phonetica*, *67*(4), 243–267. <https://doi.org/10.1159/000327392>
- Hume, E., & Johnson, K. (2001). A model of the interplay of speech perception and phonology. *Studies on the Interplay of Speech Perception and Phonology*, *55*, 1–22.

- Retrieved from https://corpus.linguistics.berkeley.edu/~kjohnson/papers/Hume_Johnson2001.pdf%5Cnpapers://e7d065ae-9998-4287-8af0-c9fa85af8e96/Paper/p23053.
- Hyde, K. L., Lerch, J. P., Zatorre, R. J. R., Griffiths, T. D., Evans, A. C., & Peretz, I. (2007). Cortical thickness in congenital Amusia: When less is better than more. *Journal of Neuroscience*, 27(47), 13028–13032. <https://doi.org/10.1523/JNEUROSCI.3039-07.2007>
- Hyde, K. L., Zatorre, R. J., & Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital Amusia. *Cerebral Cortex*, 21(2), 292–299. <https://doi.org/10.1093/cercor/bhq094>
- Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1–2), 101–144. <https://doi.org/10.1016/j.cognition.2002.06.001>
- Jamison, H. L., Watkins, K. E., Bishop, D. V. M., & Matthews, P. M. (2006). Hemispheric specialization for processing auditory nonspeech stimuli. *Cerebral Cortex*, 16(9), 1266–1275. <https://doi.org/10.1093/cercor/bhj068>
- Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2012). Impaired categorical perception of lexical tones in Mandarin-speaking congenital amusics. *Memory & Cognition*, 40(7), 1109–1121. <https://doi.org/10.3758/s13421-012-0208-2>
- Kell, C. A., Morillon, B., Kounieher, F., & Giraud, A. L. (2011). Lateralization of speech production starts in sensory cortices—A possible sensory origin of cerebral left dominance for speech. *Cerebral Cortex*, 21(4), 932–937. <https://doi.org/10.1093/cercor/bhq167>
- Klein, D., Zatorre, R. J. R., Milner, B., & Zhao, V. (2001). A Cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *NeuroImage*, 13(4), 646–653. <https://doi.org/10.1006/nimg.2000.0738>
- Kwok, V. P. Y., Wang, T., Chen, S., Yakpo, K., Zhu, L., Fox, P. T., & Tan, L.-H. (2015). Neural signatures of lexical tone reading. *Human Brain Mapping*, 36(1), 304–312. <https://doi.org/10.1002/hbm.22629>
- Li, A., Xiong, Z., & Wang, X. (2006). Contrastive study on tonal patterns between accented and standard Chinese. In *Proceedings of the 5th International Symposium on Chinese Spoken Language 2006* (pp. 157–168).
- Li, X., & Chen, Y. (2015). Representation and processing of lexical tone and tonal variants: Evidence from the mismatch negativity. *PLoS ONE*, 10(12), 1–24. <https://doi.org/10.1371/journal.pone.0143097>
- Li, X., Gandour, J. T., Talavage, T., Wong, D., Hoffa, A., Lowe, M., & Dziedzic, M. (2010). Hemispheric asymmetries in phonological processing of tones versus segmental units. *NeuroReport*, 21(10), 690–694. <https://doi.org/10.1097/WNR.0b013e32833b0a10>
- Liu, F., Chan, A. H. D., Ciocca, V., Roquet, C., Peretz, I., & Wong, P. C. M. (2016). Pitch perception and production in congenital amusia: Evidence from Cantonese speakers. *The Journal of the Acoustical Society of America*, 140(1), 563. <https://doi.org/10.1121/1.4955182>
- Liu, F., Jiang, C., Thompson, W. F., Xu, Y., Yang, Y., & Stewart, L. (2012). The mechanism of speech processing in congenital amusia: Evidence from Mandarin speakers. *PLoS ONE*, 7(2), e30374. <https://doi.org/10.1371/journal.pone.0030374>
- Liu, L., Peng, D., Ding, G., Jin, Z., Zhang, L., Li, K., & Chen, C. (2006). Dissociation in the neural basis underlying Chinese tone and vowel production. *NeuroImage*, 29(2), 515–523. <https://doi.org/10.1016/j.neuroimage.2005.07.046>
- Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *Journal of Neuroscience*, 29(33), 10215–10220. <https://doi.org/10.1523/JNEUROSCI.1701-09.2009>
- Loui, P., Li, H. C., & Schlaug, G. (2011). White matter integrity in right hemisphere predicts pitch-related grammar learning. *NeuroImage*, 55(2), 500–507. <https://doi.org/10.1016/j.neuroimage.2010.12.022>
- Luo, H., Ni, J.-T., Li, Z.-H., Li, X.-O., Zhang, D.-R., Zeng, F.-G., & Chen, L. (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. <https://doi.org/10.1073/pnas.0607065104>

- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, 17(19), 1692–1696. <https://doi.org/10.1016/j.cub.2007.08.064>
- Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics*, 65(6), 956–969. <https://doi.org/10.3758/BF03194826>
- Mitterer, H., Csépe, V., Honbolygo, F., & Blomert, L. (2006). The recognition of phonologically assimilated words does not depend on specific language experience. *Cognitive Science*, 30(3), 451–479. https://doi.org/10.1207/s15516709cog0000_57
- Myers, J., & Tsay, J. (2003). Investigating the phonetics of Mandarin tone sandhi. *Taiwan Journal of Linguistics*, 1(1), 29–68. [https://doi.org/10.6519/TJL.2003.1\(1\).2](https://doi.org/10.6519/TJL.2003.1(1).2)
- Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, 118(12), 2544–2590. <https://doi.org/10.1016/j.clinph.2007.04.026>
- Nan, Y., & Friederici, A. D. (2013). Differential roles of right temporal cortex and broca's area in pitch processing: Evidence from music and mandarin. *Human Brain Mapping*, 34(9), 2045–2054. <https://doi.org/10.1002/hbm.22046>
- Nan, Y., Sun, Y., & Peretz, I. (2010). Congenital amusia in speakers of a tone language: Association with lexical tone agnosia. *Brain*, 133(9), 2635–2642. <https://doi.org/10.1093/brain/awq178>
- Nixon, J. S., Chen, Y., & Schiller, N. O. (2015). Multi-level processing of phonetic variants in speech production and visual word processing: Evidence from Mandarin lexical tones. *Language, Cognition and Neuroscience*, 30(5), 491–505. <https://doi.org/10.1080/23273798.2014.942326>
- Niziolek, C. A., & Guenther, F. H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech FEEDBACK alterations. *Journal of Neuroscience*, 33(29), 12090–12098. <https://doi.org/10.1523/JNEUROSCI.1008-13.2013>
- Ohala, J. J. (1993). *Coarticulation and phonology*. *Language and speech* (Vol. 36). <https://doi.org/10.1177/002383099303600303>.
- Peng, S.-H. (2000). Lexical versus “phonological” representations of Mandarin sandhi tones. In M. B. Broe & J. B. Pierrehumbert (Eds.), *Papers in laboratory phonology 5: Acquisition and the lexicon* (1st ed., pp. 152–167). Cambridge (UK): Cambridge University Press.
- Peretz, I. (2013). *The biological foundations of music: Insights from congenital amusia*. *The psychology of music* (3rd ed.). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-381460-9.00013-4>.
- Poeppl, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as “asymmetric sampling in time.” *Speech Communication*, 41(1), 245–255. [https://doi.org/10.1016/S0167-6393\(02\)00107-3](https://doi.org/10.1016/S0167-6393(02)00107-3)
- Politzer-Ahles, S., Schluter, K., Wu, K., & Almeida, D. (2016). Asymmetries in the perception of Mandarin tones: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance*, 42(10), 1547–1570. <https://doi.org/10.1037/xhp0000242>
- Price, C. J. (2010). The anatomy of language: A review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, 1191(1), 62–88. <https://doi.org/10.1111/j.1749-6632.2010.05444.x>
- Pulvermüller, F., Kiff, J., & Shtyrov, Y. (2012). Can language-action links explain language laterality? An ERP study of perceptual and articulatory learning of novel pseudowords. *Cortex*, 48(7), 871–881. <https://doi.org/10.1016/j.cortex.2011.02.006>
- Scharinger, M., Monahan, P. J., & Idsardi, W. J. (2016). Linguistic category structure influences early auditory processing: Converging evidence from mismatch responses and cortical oscillations. *NeuroImage*, 128, 293–301. <https://doi.org/10.1016/j.neuroimage.2016.01.003>
- Schönwiesner, M. (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *European Journal of Neuroscience*, 22, 1521–1528. Retrieved from <https://onlinelibrary.wiley.com/doi/10.1111/j.1460-9568.2005.04315.x/full>.

- Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, *10*(4), 295–302. <https://doi.org/10.1038/nrn2603>
- Shtyrov, Y., Kujala, T., Palva, S., Ilmoniemi, R. J., & Näätänen, R. (2000). Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. *NeuroImage*, *12*(6), 657–663. <https://doi.org/10.1006/nimg.2000.0646>
- So, C. K., & Best, C. T. (2014). Phonetic influences on english and french listeners' assimilation of mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, *36*(2), 195–221. <https://doi.org/10.1017/S0272263114000047>
- Speer, S. R., Shih, C.-L., & Slowiaczek, M. L. (2016). *Prosodic structure in language understanding: Evidence from tone sandhi in Mandarin*. <https://doi.org/10.1177/002383098903200403>.
- Speer, S. R., Shih, C. L., & Slowiaczek, M. L. (1989). Prosodic structure in language understanding: Evidence from tone sandhi in mandarin. *Language and Speech*, *32*(4), 337–354. <https://doi.org/10.1177/002383098903200403>
- Sun, Y., Giavazzi, M., Adda-decker, M., Barbosa, L. S., Kouider, S., Bachoud-Lévi, A. C., Peperkamp, S. (2015). Complex linguistic rules modulate early auditory brain responses. *Brain and Language*, *149*(2009), 55–65. <https://doi.org/10.1016/j.bandl.2015.06.009>.
- Tavabi, K., Elling, L., Dobel, C., Pantev, C., & Zwitserlood, P. (2009). Effects of place of articulation changes on auditory neural activity: A magnetoencephalography study. *PLoS ONE*, *4*(2). <https://doi.org/10.1371/journal.pone.0004452>.
- Tillmann, B., Burnham, D., Nguyen, S., Grimault, N., Gosselin, N., & Peretz, I. (2011). Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages. *Frontiers in Psychology*, *2*(JUN), 120. <https://doi.org/10.3389/fpsyg.2011.00120>.
- Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, *39*(3), 1429–1443. <https://doi.org/10.1016/j.neuroimage.2007.09.054>
- Tsay, J., & Myers, J. (1996). Taiwanese tone sandhi as allomorph selection. In *Proceedings of Annual Meeting of the Berkeley Linguistics Society*.
- Wang, J., Zhang, C., Wan, S., & Peng, G. (2017). Is congenital amusia a disconnection syndrome? A study combining tract- and network-based analysis. *Frontiers in Human Neuroscience*, *11*(September), 1–11. <https://doi.org/10.3389/fnhum.2017.00473>
- Wang, W. S.-Y., & Li, K.-P. (1967). Tone 3 in Pekinese. *Journal of Speech and Hearing Research*, *10*(3), 629–636. Retrieved from <https://jslhr.asha.org/cgi/content/abstract/10/3/629>.
- Wong, P. C. M., Parsons, L. M., Martinez, M., & Diehl, R. L. (2004). The role of the insular cortex in pitch pattern perception: The effect of linguistic contexts. *Journal of Neuroscience*, *24*(41), 9153–9160. <https://doi.org/10.1523/JNEUROSCI.2225-04.2004>
- Xu, Y. (2004). Understanding tone from the perspective of production and perception. *Language and Linguistics*, *5*(4), 757–797.
- Xu, Y., & Emily Wang, Q. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, *33*(4), 319–337. [https://doi.org/10.1016/S0167-6393\(00\)00063-7](https://doi.org/10.1016/S0167-6393(00)00063-7)
- Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, *33*(2), 159–197. <https://doi.org/10.1016/j.wocn.2004.11.001>
- Yu, Y. H., Shafer, V. L., & Sussman, E. S. (2017). Neurophysiological and behavioral responses of Mandarin lexical tone processing. *Frontiers in Neuroscience*, *11*, 95. <https://doi.org/10.3389/fnins.2017.00095>
- Zatorre, R. J. R. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, *11*(10), 946–953. <https://doi.org/10.1093/cercor/11.10.946>
- Zatorre, R. J. R., & 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. <https://doi.org/10.1098/rstb.2007.2161>

- Zhang, C., Peng, G., Shao, J., & Wang, W. S.-Y. (2017). Neural bases of congenital amusia in tonal language speakers. *Neuropsychologia*, 97(July 2016), 18–28. <https://doi.org/10.1016/j.neuropsychologia.2017.01.033>.
- Zhang, C., Xia, Q., & Peng, G. (2015). Mandarin third tone sandhi requires more effortful phonological encoding in speech production: Evidence from an ERP study. *Journal of Neurolinguistics*, 33, 149–162. <https://doi.org/10.1016/j.jneuroling.2014.07.002>
- Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology*, 27(01), 153. <https://doi.org/10.1017/S0952675710000060>
- Zhang, J., & Liu, J. (2016). The productivity of variable disyllabic tone sandhi in Tianjin Chinese. *Journal of East Asian Linguistics*, 25(1), 1–35. <https://doi.org/10.1007/s10831-015-9135-0>
- Zheng, Z. Z., Vicente-Grabovetsky, A., MacDonald, E. N., Munhall, K. G., Cusack, R., & Johnsrude, I. S. (2013). Multivoxel patterns reveal functionally differentiated networks underlying auditory feedback processing of speech. *Journal of Neuroscience*, 33(10), 4339–4348. <https://doi.org/10.1523/JNEUROSCI.6319-11.2013>