Cross-Dialectal Perception of the Third-Tone Sandhi in Standard Chinese – Evidence from Eye Movements

Qian Li, Yingyi Luo, Aijun Li

Institute of Linguistics, Chinese Academy of Social Sciences, China

`li_qian@cass.org.cn, yingyi.luo@gmail.com, liaj@cass.org.cn`

Abstract

It is well-known that in Standard Chinese, the sandhi-derived T3 is realized with a rising f_0 pattern that is typically identical with the lexical T2. However, it is less understood on how dialect backgrounds play a role in T3 sandhi perception. The present study set out to shed some light on this issue with evidence with eye movements by examining the participants' looks to both targets and competitors within the Visual World Paradigm when they heard T3T3 stimuli as well as their T2T3 near-minimal-pair counterparts. We recruited participants from both Beijing and Nanjing. Our results show that, due to the modulation of their native tonal system, Nanjing listeners showed clear early advantages in the online processing of tones in Standard Chinese, as reflected with more looks to the target as compared to the Beijing group.

Index Terms: cross-dialectal, tone perception, tone sandhi, Standard Chinese, eye-tracking

1. Introduction

In Standard Chinese (hereafter as SC), lexical tones are realized with different f_0 patterns. Tone 1 (hereafter as T1) is realized with a high level f_0 contour; Tone 2 (T2) a rising f_0 contour; Tone 3 (T3) a dipping f_0; and Tone (T4) a falling contour. Various f_0 cues are known to be adequately employed by listeners to identify lexical tones in speech perception, such as the overall f_0 height (e.g., [1]-[4]), the direction of f_0 change (e.g., [5]-[7]), as well as the timing of the f_0 turning point (e.g., [2], [6], [7]). However, despite the linguistic diversity within China, these conclusions were generally made from experiments recruiting exclusively Beijing Mandarin (hereafter as BM) speakers; or alternatively, participants speaking different Northern Mandarin varieties were undifferentially treated as native speakers of SC. It is thus less studied on how dialectal backgrounds influence tone perception in SC. Another important fact that has been often neglected in the research of Mandarin spoken word recognition is the common phenomenon of diglossia. A large proportion of Chinese speakers are users of both SC and their own dialects, often with one variety used more dominantly than another [8].

So far, little has been known on how tonal representation in SC and dialects interplay with each other in the online process of spoken word recognition. On the one hand, previous studies on tonal awareness, typically with offline experimental tasks, have suggested that bi-dialectal users exhibited greater sensitivity to tones than those who speak only one dialect (e.g., [12]); on the other hand, the production of SC tones by speakers of dialects has been reported to deviate more or less from the standard variety, probably due to the influence from speakers' native dialects (e.g., [9]-[11]). A recent study on spoken word recognition using the eye-tracking technique has further suggest the possibility that speakers of dialects might be less sensitive to Mandarin tonal information than supposed. [13] investigated how the interaction between syllable frequency and syllable-specific tonal probability guides online lexical access in speakers of mutually unintelligible Chinese dialects with three different tonal systems. Mono-dialectal Mandarin speakers, bi-dialectal Shanghai–Mandarin speakers as well as bi-dialectal Cantonese–Mandarin speakers were included in this study. Although the end-state clicks were fastest for infrequent syllables with most probable tones and slowest for infrequent syllables with least probable tones across all groups, in online eye movement responses, only mono-dialectal Mandarin speakers showed an interaction between syllable frequency and tonal probability. Bi-dialectal speakers, however, showed different timing in their integration of tonal probabilities. This means that although bi-dialectal speakers can utilize Mandarin tonal probabilities, their sensitivity to Mandarin tonal information was delayed compared to mono-dialectals in online spoken word recognition. But note that, there are great differences in the segmental aspects among mutually unintelligible dialects such as Mandarin, Shanghainese, and Cantonese, which were investigated in [13]. This might impose extra influences on the process of tone recognition. So far, no consensus has been reached on whether native dialects facilitate or inhibit the processing of SC tones.

The current study thus aims to revisit this issue by including bi-dialectal speakers of SC and another closely related Mandarin variety. The target dialect in this study in Nanjing Mandarin (hereafter as NM), which is the southern variety of Mandarin spoken in the urban districts of Nanjing, the capital city of Jiangsu Province. It is similar with SC in terms of segments. Most consonantal contrasts in SC do exist in NM except for the lack of /n/-/l/ for onsets or /n/-/ŋ/ for coda. Vowels in NM are also comparable to those in SC except for the coalescence of diphthongs (see [12] for a brief summary of the phonological system of NM). Therefore the phonological system of NM differs from that in SC primarily in the tone system. There are five tones in NM, as illustrated in Figure 1. The f_0 values were normalized with z-score [14]. Each illustrated tonal contour was then based on the mean z-score averaged across 8 samples produced by 10 speakers in their 20s.

As can be seen from Figure 1, NM has a more complex tonal system than the four-tone system of SC. T1 in NM is a high-falling tone, T2 a rising tone; T3 a low-falling tone; T4 a mid tone. In addition, NM also has one checked tone T5, which is typically produced with shorter duration and a noticeable glottal syllable coda [15], as illustrated with a shorter f_0 contour in Figure 1.

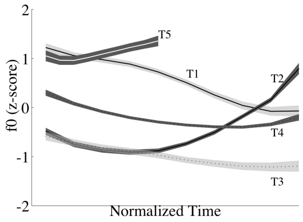

Figure 1: f_0 *realization of the five lexical tones in NM, produced in isolation. Lines stand for the mean, and gray areas for ±1 standard error of mean. T1 is illustrated with black solid line and light gray area; T2 with black solid line and dark gray area; T3 with dashed line and light gray area; T4 with dashed line and dark gray area; T5 with white solid line and dark gray area. Normalized time.*

When tones are combined, the third-tone sandhi takes place in both SC and NM. Figure 2 shows the f_0 realization of tonal combinations T3T3 vs. T2T3 in SC (see Figure 2*a*) vs. NM (see Figure 2*b*). The f_0 values were normalized with z-score [17]. The illustrated tonal contours for SC (in Figure 2*a*) were the mean z-score averaged across 30 samples produced by one male SC speakers. Those for NM (in Figure 2*b*) were based on 8 samples produced by 10 NM speakers. As can be seen from Figure 2a that in SC, when two T3s are combined, the first T3 is realized with a rising f_0 which is almost identical with the rising T2. Similar "rule" can also be found in NM (Figure 2*b*), as T3T3 is realized with a similar f_0 with that of T2T3, but the sandhi-derived T3 obviously shows lower overall f_0 than that of T2 in T2T3.

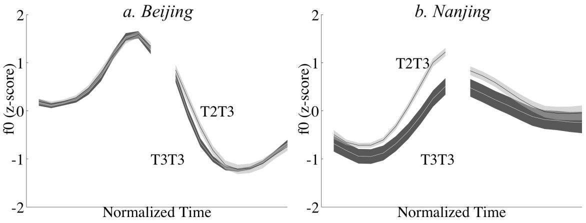

Figure 2: f_0 *realization of T3T3 vs. T2T3 in SC (a) vs. NM (b). Lines stand for the mean, and gray areas for ±1 standard error of mean. T3T3 is illustrated with white line and dark gray area; T2T3 with black solid line and light gray area. Normalized time.*

Therefore, the question arises as to how NM speakers process T3T3 sandhi in SC where the f_0 realization of T3T3 is almost identical to that of T2T3. Previous studies on the perception of T3 sandhi in SC have been interested in two main research questions: one concerns with whether native listeners could differentiate sandhi-T3 vs. T2 in SC (e.g., [16]), and the other mainly focused how sandhi-T3 is represented in the mental lexicon (e.g., [17]-[19]). No research effort thus far, has been put to investigate the effect of dialectal backgrounds on the perception of the third-tone sandhi in SC. This study therefore aims to fill this knowledge gap via an auditory word-recognition task within the Visual World Paradigm (hereafter as VWP) [20] using the technique of eye-tracking.

During a typical VWP experiment, participants are usually displayed with multiple pictures or written words on a computer screen, and simultaneously hear an auditory stimulus, which corresponds to the target while their eye movements are tracked [20]. Typically, a target is presented with a competitor (and optional distractors). The participants were asked to identify the words they have heard and click on the target with a mouse. Based on the assumption that eye movements are closely time-locked to the spoken-word processing, this paradigm makes it possible to tap into the time-course of auditory word recognition as well as the end-state responses in typically behavioral tasks.

The task involved in this paradigm is more natural speech perception experience than just asking the participants to make meta-linguistic judgments, which is typically used in traditional studies on tonal perception. To make inferences about how the target auditory stimuli are processed, previous studies have mainly looked at how fast listeners start to fixate upon the target words (or competitors) as well as how long they look at the targets (or competitors). To assess the effect of dialectal backgrounds, we included both native BM mono-dialectal listeners and SC-NM bi-dialectal listeners. Two hypotheses can be proposed based on the existing studies: 1) Given the fact that NM has a more complex tonal system, bi-dialectal speakers and also another more tonally complex dialect - i.e., the SC-NM bi-dialectal group may exhibit a greater processing sensitivity and faster word recognition when hearing SC tones compared to the BM mono-dialectal group. We would then expect to observe more and/or earlier looks to the target, as well as better performances in the behavioral results for the SC-NM group than the BM group. 2) Alternatively, if the SC-NM group is less sensitive and shows slower word recognition, we would then expect to see less and/or later looks to the target, as well as worse performances in the behavioral results for the SC-NM group.

2. Methods

2.1. Participants

29 native speakers of BM and 28 native speakers of NM were recruited. All participants were born in late 1980s or 1990s (BM: Mean=23±3; NM: Mean=22±3) and raised in the urban areas of Beijing and Nanjing, respectively. All NM speakers passed a word production test, in which their dialect fluency was judged by a linguist whose native dialect is NM. None of the participants had lived out of Beijing/Nanjing before 18. They were paid for their participation but unaware of the purpose of the experiment. All subjects had normal or corrected-to-normal vision.

2.2. Stimuli

The critical target stimuli consisted of 24 pairs of highly lexicalized disyllabic collocations. We included 12 T3T3-T2T3 minimal pairs differing only in the tone, as well as 12 T3T3-T2T3 near minimal pairs differing in the rhyme part of the second syllable. In both conditions, when T3T3 stimuli were used as targets (T3T3-Target condition), the competitors were their T2T3 counterparts within the pairs; and vice versa (T2T3-Target condition).

Critical stimuli within each pair were controlled to be closely matched in terms of lexical frequency based on [21] and orthographic complexity as we presented Chinese characters instead of pictures (following [4]), so that there was no significant difference between T3T3 and T2T3 stimuli for lexical frequency or visual complexity. Furthermore, we also matched the bigram mutual information which stands for the likelihood of two syllables co-occuring within a lexical item [22]. The overall mean bigram co-occuring frequency of all was not significantly different between each T3T3-T2T3 pair. In addition, in case that the participants were aware of the experimental purpose, we added 36 pairs of filler stimuli: 12 near minimal pairs of T4T4 vs. T1T4, 12 pairs of T1T1 vs. T3T1, and another 12 pairs of T2T2 vs. T4T2 sequences.

All auditory stimuli were pre-recorded by a female Beijing speaker who was born in the 1980s. All stimuli were produced with the same loudness and speaking rate. The mean duration of the first syllable of the critical stimuli was 340ms, and was not significantly different between T3T3 and T2T3 stimuli.

2.3. Procedure

Eye movements of both subject groups were recorded with an Eyelink 1000/1000+ system with a 35mm lens running at 500Hz. Visual stimuli were presented on a computer monitor. Participants were seated comfortably at a distance of 69cm from the screen. All recordings were done monocularly based on the pariticpants' left eyes and viewing was binocular.

The experiment began with a block of six training trials. Participants were tested individually. Each trial consisted of a preview screen (2000ms) and a fixation-cross screen (500ms) with only a fixation cross at the very center of the screen prior to the stimuli screen. The participants were asked to look at the fixation cross until it disappeared. Then the auditory stimulus was played through a headphone simultaneously with the presentation of the visual stimuli screen. The task was to click on the visual stimuli they had heard with a mouse as soon as possible.

In each trial, participants saw two visual stimuli on the screen, which consisted of a target corresponding to the auditory stimulus and a competitor. Each visual stimulus was 2.5cm in height and 5cm in width, corresponding to 5° visual angle. Two stimuli on both the preview screen and the visual stimuli screen approximately located at the centers of the left and right part of the screen. The order of the two visual stimuli was counter-balanced. The trial order was pseudorandomized so that auditory stimuli of the same condition were not presented in consecutive trials.

2.4. Eye Movement Data Analysis

The eye movement data were recorded at 2ms intervals. The proportion of looks at each time point was calculated. The eye movement data were reported from the onset of the auditory stimulus to 1000ms post stimulus onset. The upper limit was chosen at the point where the proportion of looks to target had reached the maximum (following [23]). Growth curve analysis was run for statistics (following [24]) with the package *lme4* [25] in R [26].

3. Results

3.1. Behavioral Data

The behavioral results show that, the overall mean rate of correct responses for all trials for Beijing subjects is 91.6% (SD=1.3%), and 92.73% (SD=1.53%) for Nanjing subjects. This thus suggests that at least in the behavioral part, our results are reliable. Figure 3 shows the mean rate of correct responses in both Minimal Pair condition and Near Minimal Pair condition.

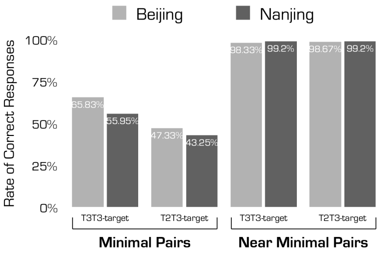

Figure 3: *The mean rate of correct responses in both Minimal Pair and Near Minimal Pair conditions across target types. Light gray bars for Beijing subjects and dark gray bars for Nanjing subjects.*

From Figure 3, we can see that in the rate of correct response in Minimal Pair part across subject groups and target types are generally much lower than those in Near Minimal Pair part. The mean rate of correct responses in the Minimal Pair part is basically around the chance level, while those in the Near Minimal part all hit the ceiling. We further separated these two part and ran linear-mixed effect models respectively. The results showed a significant main effect of TARGET TYPE in the Minimal Pair part (β=0.185, t=5.053, p<0.001), with neither significant main effect of DIALECT (β=-0.04, t=-1.065, n.s.) nor interaction with DIALECT (β=-0.06, t=-1.071, n.s.). For the Near Minimal Pair part, there was no significant difference across TARGET TYPE (β=-0.01, t=-0.38, n.s.) or DIALECT (β=0.01, t=0.54, n.s.).

3.2. Eye Movement Data (Near Minimal Pairs)

As the rates of correct response in the Minimal Pair part are around chance level, it is thus less meaning full to further look at the eye movements in the Minimal Pair part. We thus only show eye movement data in the Near Minimal Pair part. Figure 4 illustrates the mean proportion of looks to the target in T3T3-Target condition (i.e., when T3T3 is the target and its T2T3 counterpart as the competitor; left panel), as well as in T2T3-Target condition (i.e., when T2T3 is the target and its T3T3 counterpart as the competitor; right panel). X-axis stands for the time since auditory stimuli onset and y-axis for the proportion of looks to target. In general, two patterns can be observed from Figure 4.

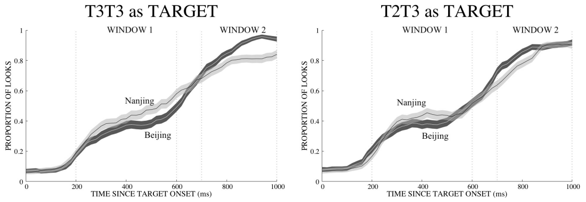

Figure 4: *The averaged proportion of looks to target over time when T3T3 is the target (left) compared to when T2T3 is the target (right). White lines stand for the Beijing group (with dark gray areas for ±1 standard error of mean); black lines for the Nanjing group (with light gray areas for ±1 standard error of mean).*

First, the proportion of looks to target for both subject groups remains at the bottom and does not start to increase (no more than 0.1) until around 200ms since auditory stimuli onset when the proportion starts to rise in both T3T3-Target condition (Figure 4; left panel) and T2T3-Target condition (Figure 4; right panel). This short period of time (i.e., about

200ms) relatively corresponds to the time that eyes' usually need to execute a saccadic action [20], during which participants were not looking at the target or competitor.

Second, after 200ms since auditory stimuli onset, the proportion of looks to target for different subject groups did not increase in parallel in either target conditions. The two growth curves in each target condition crossed over at around 600ms. The proportion of looks to the target for the Nanjing group seems higher than that for the Beijing group within the time window of around 200-600ms since auditory stimuli onset in both target conditions. The patterns were reversed within the time window of around 700-1000ms since auditory stimuli onset in both target conditions. A set of growth curve analyses were thus run for the comparison of proportion of looks to the target for the Beijing vs. Nanjing subject groups within the two specific time windows (i.e., 200-600ms and 700-1000ms). For our planned comparisons, the whole dataset was separated into two subsets according to the target condition, i.e., T3T3-Target condition and T2T3-Target condition. The estimated results were summarized in Tables 1 and 2.

Table 1: *Results of growth curve analyses with the 200-600ms time window.*

	T3T3-Target Condition	**T2T3-Target Condition**
Intercept	β=0.04, t=2.22, p<0.01	β=0.07, t=2.69, p<0.01
Slope	β=0.16, t=2.19, p<0.01	β=0.05, t=0.30, n.s.
Quadratic	β=0.11, t=1.10, n.s.	β=0.25, t=3.26, p<0.01

Table 2: *Results of growth curve analyses within the 700-1000ms time window.*

	T3T3-Target Condition	**T2T3-Target Condition**
Intercept	β=-0.04, t=-2.88, n.s.	β=-0.02, t=-0.32, n.s.
Slope	β=0.10, t=0.51, n.s.	β=-0.14, t=-0.71, n.s.
Quadratic	β=-0.06, t=0.71, n.s.	β=-0.24, t=-2.84, p<0.01

The results showed that within the 200ms-600ms time window, the proportions of looks to the target for the Nanjing subject group were significantly higher than that for the Beijing group in the first time window in both target conditions. However, at a later time window, although a significant effect of main effect of DIALECT was also observed, the directions were reversed in both target condition.

4. Discussion & Conclusions

While it has been well-known that dialect backgrounds might influence tone perception, what the effect is implemented in online speech recognition is still controversial. The present study set out to shed some light on this issue with evidence with eye movements in the VWP. We examined the participants' end-state responses as well as their looks to the targets for both Nanjing and Beijing subject groups within the VWP when they heard T3T3 stimuli as well as their T2T3 minimal-pair or near-minimal-pair counterparts. Our results yield obvious perceptual differences between the two subject groups, as reflected in the different eye movement patterns.

Our behavioral results showed that the rates of correct responses for both BM and NM listeners reached only around chance level in the Minimal Pair part, suggesting that neither group can reliably differentiate the two items within a T3T3-T2T3 minimal pair. In contrast, both groups can reach up to more than 90% for the near minimal pairs. This thus replicates previous studies on third-tone sandhi using exclusively Beijing Mandarin speakers within traditional offline tone perception experiments (e.g., [16]). We thus further examined the eye movement data for the near minimal pair part.

The eye movement data showed that, when the target was T3T3, NM listeners looked at the target with more proportions than BM listeners until 600ms since target onset. Taking into consideration of the saccadic execution time, this corresponds to 400ms into the auditory stimuli, which is about 60ms into the second syllable (mean duration of the first syllable is 340ms). This means within the first syllable, NM-speaking listeners take less time to show a preference to the target. However, the trends are not always so within the next time window. BM listeners in turn looked more at the target than NM listeners within the 700-1000ms time window. This relatively corresponds to the second syllable as well as after the auditory stimulus ends. This has indicated that once BM listeners recognize the target, they are more confident about their decisions than NM listeners at a later stage, which is closer to the stage when end-state decisions were made. The reserved patterns of eye movements thus explained the conflicting results in the body of literature on the effect of dialect backgrounds in tone perception (e.g., [9]-[13]). Listeners whose native dialects have a more complex tone system do exhibit a processing advantage and show greater tone sensitivity when hearing SC tones, especially at an early stage as they develop higher sensitivities to subtle f_0 differences as compared to listeners whose native dialects were simpler. The advantage is also related to their exposure to more than one tone system as tones in different dialects are represented separately. This is comparable to the findings which show bi-dialectal listeners have better tonal awareness than mono-dialectals. such advantage, however, disappeared at the decision-making stage later on. This is thus in line to previous studies where L2 language users show less well performance in both production and perception.

In addition, a further comparison between T3T3-target condition vs. T2T3-target condition suggested that, for both subject groups, the proportion of looks to target in both target type conditions show clear differences. This thus contribute further evidence to the fact that sandhi-T3 and T2 are processed differently, which indicate the non-neutralization of sandhi-T3 with the lexical T2.

In conclusion, this study showed an effect of dialectal backgrounds in SC tone perception. Listeners speaking a dialect with more complex tone systems showed clear early advantages in the online processing of SC tones, while this advantage due to tone experience had to give way to the native-speaker advantage in the end-state stage.

5. Acknowledgements

This research was financially supported by China Postdoctoral Science Foundation. We thank Ying Chen for making the eye-tracker available at Nanjing University of Science and Technology. Special thanks to Xiaoli Ji, Hanyu Xu and Wentao Xiao for their assistance during the experiment.

6. References

[1] H.-B. Lin, and B. H. Repp, "Cues to the Perception of Taiwanese Tones," *Haskins Laboratories Status Report on Speech Research* SR-99/100, pp. 137–47, 1989.

[2] X. S. Shen, M. Lin, and J. Yan, “F0 Turning Point as an f0 Cue to Tonal Contrast: A Case Study of Mandarin Tones 2 and 3,” *J. of the Acou. Soc. of America* 4, pp. 2241–43, 1993.

[3] A. L. Francis, V. Ciocca, and B. K. C. Ng, “On the (Non)categorical Perception of Lexical Tones,” *Perception & Psychophysics* 65, pp. 1029–44, 2003.

[4] J. Shen, D. Deutsch, and K. Rayner, “On-Line Perception of Mandarin Tones 2 and 3: Evidence from Eye Movements,” *J. of the Acou. Soc. of America* 133, pp. 3016–29, 2013.

[5] R. A. Fox, and Y.-Y. Qi, “Context Effect in the Perception of Lexical Tone,” *J. of Chinese Ling.* 18, pp. 261–84, 1990.

[6] X. S. Shen, and M. Lin, “A Perceptual Study of Mandarin Tones 2 and 3,” *Language and Speech* 34, pp. 145–56, 1991.

[7] C. B. Moore, and A. Jongman, “Speaker Normalization in the Perception of Mandarin Chinese Tones,” *J. of the Acou. Soc. of America* 102, pp. 1864–77, 1997.

[8] D. C. S. Li, and S. Lee, “Bilingualism in East Asia,” in T. K. Bhatia & B. C. Ritchie (eds.), *The Handbook of Bilingualism,* pp. 742–779. Blackwell, 2004.

[9] S. Jin, and F. Niu, “Changsha Difang Putonghua Guhua Yanjiu – Difang Putonghua Guhua de Gean Diaocha [A study on Changsha-accented Standard Chinese – A case study on accented Standard Chinese]”, *Applied Linguistics* 4, pp. 41-9, 2010.

[10] Z. Wang, “Daxuesheng Putonghua Ceshi Kefu ‘Xiangyin’ Yanjiu [A study on accented Standard Mandarin in college test of Standard Chinese]”, *Journal of Liaoning Normal University (Social Science Edition)* 37, pp. 564-9, 2014.

[11] J. Cheng, *Yixing Kouyin Putonghua Yuyin Yanjiu [A phonetic study on Yixing-accented Standard Chinese]*, MA thesis, Yangzhou Univeristy, 2017.

[12] X. Chen, X., R. C. Anderson, W. Li, M. Hao, X. Wu, and H. Shu, “Phonological Awareness of Bilingual and Monolingual Chinese Children”, *Journal of Educational Psychology* 96, pp. 142-151, 2004.

[13] S. Wiener and K. Ito, “Do Syllable-Specific Tonal Probabilities Guide Lexical Access? Evidence from Mandarin, Shanghai and Cantonese Speakers”, *Language, Cognition and Neuroscience, DOI:* 10.1080/23273798.2014.946934, 2014.

[14] P. Rose, “Considerations in the Normalisation of the Fundamental Frequency of Linguistic Tone”, *Speech Communication* 6, pp. 343–352, 1987.

[15] D. Liu, *Nanjinghua Yindang [Auditory recordings of Nanjing Mandarin]*. Shanghai Education Press.

[16] W. S.-Y. Wang, and K.-P. Li, “Tone 3 in Pekinese,” *J. of Speech, Language, and Hearing Res.* 10, pp. 629–36, 1967.

[17] X. Zhou, and W. Marslen-Wilson, “The Abstractness of Phonological Representation in the Chinese Mental Lexicon,” In H.-C. Chen (ed.) *Cognitive Processing of Chinese and Related Asian Languages*, pp. 3–26. The Chinese University Press, Hong Kong, 1997.

[18] Y.-F. Chien, J. A. Sereno, and J. Zhang, “Priming the Representation of Mandarin Tone 3 Sandhi Words,” *Language, Cognition and Neuroscience*, 2015.

[19] X. Li, and Y. Chen, “Representation and Processing of Lexical Tone and Tonal Variants: Evidence from the Mismatch Negativity,” *PLoS ONE* 10, 2015.

[20] M. K. Tanenhaus, M. J. Spivey-Knowlton, K. M. Eberhard, and J. C. Sedivy, “Integration of Visual and Linguistic Information in Spoken Language Comprehension,” *Science* 268, pp. 1632–34, 1995.

[21] Q. Cai, and M. Brysbaert, “SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles,” *PLoS ONE* 5, 2010.

[22] J. Da, “A Corpus-Based Study of Character and Bigram Frequencies in Chinese E-Texts and Its Implications for Chinese Language Instruction,” In P. Zhang, T. Xie, and J. Xu (eds.) *The Studies on the Theory and Methodology of the Digitalized Chinese Teaching to Foreigners: Proceedings of the Fourth International Conference on New Technologies in Teaching and Learning Chinese*, pp. 501–11, Tsinghua University Press, Beijing, 2004.

[23] J. G. Malins, and M. F. Joanisse, “The Roles of Tonal and Segmental Information in Mandarin Spoken Word Recognition: An Eyetracking Study.” *J. of Memory and Language* 62, pp. 407–20, 2010.

[24] Mirman, D., *Growth Curve Analysis and Visualization Using R*, CRC Press, 2014.

[25] D. Bates, M. Maechler, B. Bolker, and S. Walker, lme4: Linear Mixed-Effects Models Using Eigen and S4 (version 1.1-6), 2014. http://CRAN.R-project.org/package=lme4.

[26] R Core Team. R: A Language and Environment for Statistical Computing, 2014. http://www.R-project.org/.

[This paper was published at ISCSLP 2018]