Mandarin third tone sandhi may be incompletely neutralizing in perception as well as production

Stephen Politzer-Ahles[1], Katrina Connell[2], Yu-Yin Hsu[1], & Lei Pan[1]

[1]The Hong Kong Polytechnic University, [2]Pennsylvania State University
sjpolit@polyu.edu.hk, kzc501@psu.edu, yu-yin.hsu@polyu.edu.hk, panlei.pl.pan@connect.polyu.hk

ABSTRACT

Mandarin third tone sandhi is traditionally assumed to be incompletely neutralizing in production but completely neutralizing in perception, based on metalinguistic judgment tasks in which participants cannot reliably identify the underlying tone of syllables neutralized by tone sandhi. We performed a visual world eye-tracking study to see if implicit sensitivity to the differences between the surface forms influences participants' eye movement patterns, even if they cannot consciously access this for identification tasks. We found a slight trend in this direction, with participants looking more towards orthographic representations that match the underlying form of the neutralized syllable they hear. The results are statistically inconclusive, but suggest that this paradigm may be able to provide evidence that Mandarin neutralized tones are indeed incompletely neutralized, and that further research along these lines is warranted.

Keywords: incomplete neutralization, Mandarin tone sandhi, third tone sandhi, visual world eye-tracking

1. INTRODUCTION

Incomplete neutralization is common cross-linguistically (e.g., [2]). Some neutralization patterns are argued to be incomplete in production but complete in perception: that is to say, two putatively neutralized sounds have reliable acoustic differences which can be detected with computer-assisted measurements statistical methods, but human listeners cannot reliably detect them when tested. A classic example of this case is third tone sandhi in Mandarin. Crucially, two tones, Rising and Low, are putatively neutralized in a certain context. When a Low tone (also known as a third tone or "tone three", hence the name "third tone sandhi") is preceded by another within the same intonational domain, it is instead pronounced as a Rising tone (see [3, 12], among others). For instance, the morpheme written 雨 is normally pronounced [ɥy˨˩], with Low tone; but when it appears before another Low tone, as in the compound word 雨伞 [ɥy˧˥ san˨˩] "umbrella", it is instead produced with a Rising tone, homophonic with the morpheme written 鱼 [ɥy˧˥], "fish". This alternation causes the distinction between Low and Rising tones to be neutralized in pre-Low positions that license third tone sandhi. Many acoustic studies, though, have found that the distinction is not completely neutralized: a "Rising" tone derived via tone sandhi from an underlying Low tone tends to be slightly lower and have a slightly turning point in its tonal contour (can lose some). On the other hand, several studies have also shown that speakers cannot reliably identify the underlying tone in these incompletely neutralized contexts tone [4, 7, 11, 13].

Nonetheless, some results in the extant literature do suggest that the subtle contrast between underlying and sandhi-derived Rising tones may influence listeners' perception. Zhou and Marslen-Wilson [14] report auditory-auditory priming effects on Rising versus sandhi-derived targets, but this may be due to lexical awareness of the words (as the experiment used unambiguous bisyllables not designed to test incomplete neutralization). Liu [4] found that listeners can accurately discriminate sandhi-derived and underlyingly Rising tokens in an AXB task, but accurate discrimination of sounds could be based on irrelevant low-level or token-specific features rather than a categorical difference between the two (e.g., someone may also be able to accurately discriminate any two random tokens within the same category as well). Finally, Speer and Xu [10] report a visual world eye-tracking experiment in which participants heard underlyingly or sandhi-derived Rising tone syllables in a sentence context, and looked at written characters on the screen. Counterintuitively, when participants heard the sentence with a sandhi-derived Rising tone in it, they initially looked at the character corresponding to the Rising tone more than the one corresponding to the Low tone; also counterintuitively, when they heard the sentence with an underlyingly Rising tone in it, they initially looked at the Low-tone character more than the Rising tone character. These results are opposite what one would initially predict (it makes the most sense to assume that if participants can recognize the difference, they would look at the appropriate character more than the inappropriate character). Nonetheless, they are potentially consistent with the notion that participants

are subtly sensitive to the difference between the incompletely neutralized tones.

Overall, there is weak evidence that listeners may be sensitive to the difference between incompletely neutralized Mandarin tones. Importantly, all the studies suggesting that listeners are not sensitive to this difference are based on explicit metalinguistic judgment tasks, whereas some of the potential evidence that listeners are sensitive to the difference comes from online measures like eye movements and reaction times, which participants do not have direct control over and which may reflect processes that participants are not consciously aware of. Because of these conflicting sets of results, we hypothesized that listeners may be able to hear the difference between sandhi-derived and underlying Rising tones at the unconscious, automatic level, but not able to consciously access that for a metalinguistic judgment. We test this with a visual world eye-tracking experiment, using a design that is essentially a simplification of that used by Speer and Xu [10]. Participants heard ambiguous bisyllabic words, without a sentence context, like [duˊ pən˨˩], which might correspond to the word 读本 "reading book", where the citation form of the first syllable has Rising tone, or to the word 赌本 "bookie", where the citation form of the first syllable has Low tone. If participants are somewhat sensitive to the difference between the incompletely neutralized tones, they should look more at the word with an underlyingly Low first syllable when they hear a token that was spoken as a production of the word with underlying Low tone, compared to when they hear a token that was spoken as a production of the word with underlying Rising tone. By using single words rather syllables embedded in sentences, we aimed to reduce potential complications related to sentence processing and the plausibility or semantic fit between critical words and the rest of the sentence.

2. METHODS

All experiment materials, data, de-identified participant demographic information, and analysis scripts are available at https://osf.io/ursh9/. Experiment methods were pre-registered at https://osf.io/35ang/register/5771ca429ad5a1020de2872e.

2.1. Participants

60 native speakers of Mandarin (mean age 23.55, age range 18-34, 48 women and 12 men) with normal or corrected-to-normal vision and hearing participated. All experiment procedures were approved by the Human Subjects Ethics Sub-committee at the Hong Kong Polytechnic University (project reference # HSEARS20171012002). Two additional volunteers participated in the experiment but their data were not included in the analysis due to losing track of the eyes and failure of validation during the experiment.

2.2. Materials

The critical stimuli consisted of 14 pairs of disyllabic words that were identical in their segmental structure and differed in the underlying tone of the initial syllable: the second syllable was always Low, and the first was underlyingly Low or Rising, but was always pronounced with surface Rising tone. A female native Mandarin speaker produced the words in isolation.

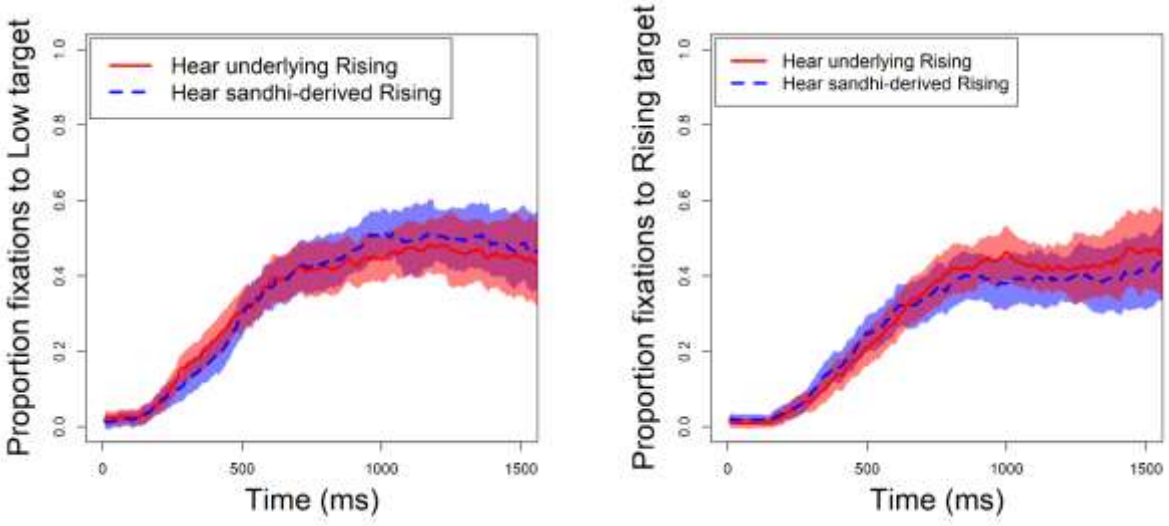
In the eye-tracking experiment, each critical display comprised four two-character words. Two words were the target pair, corresponding to the underlyingly Low-Low and the underlyingly Rising-Low interpretations of the auditory stimulus. The other two words were distractors. In addition to the 14 critical sets, there were 52 filler sets of various types, designed to prevent participants from deducing the aim of the experiment or from being able to anticipate which words in the display would be targeted before they hear the auditory stimulus.

The critical items were arranged into four lists in a Latin square design.

2.3. Procedure

The experiment was compiled using Experiment Builder software (SR Research). Participants' eye-movements were recorded with a desktop EyeLink 1000 Eye Tracker recording at 1000 Hz (1 gaze position sample recorded every millisecond). The experiment began with a calibration of the participants' pupil and corneal reflection. This calibration was followed by the practice session of 4 trials. After any questions were answered, the experiment began. A trial began with four words appearing on the screen in Times New Roman size 80 font in white on a black background in a non-displayed 2x2 grid. The words remained on the screen for 3000ms (preview time). This time allowed participants to pre-activate the pronunciations of each of the words and to familiarize themselves with their locations. No auditory stimulus was heard during this presentation. After the 3000ms preview, the images disappeared, and a fixation cross appeared in the middle of the screen for 500ms to return the participant's gaze to a neutral starting point. As the fixation cross disappeared, the words reappeared on the screen in the same locations as during the preview, and an auditory stimulus was heard though headphones. This auditory stimulus was the target

Figure 1: Proportion of looks to target. Shaded area represents a difference-adjusted by-participant Cousineau-Morey interval [1, 6].



word for that trial, heard in isolation. Participants were instructed to click on the word spoken as quickly as possible. Once the participant clicked, a blank screen appeared for 700ms, after which the next trial began. Both eye-movements (recorded from the target-word onset in the auditory stimulus) and selection accuracy were recorded.

2.4. Analysis

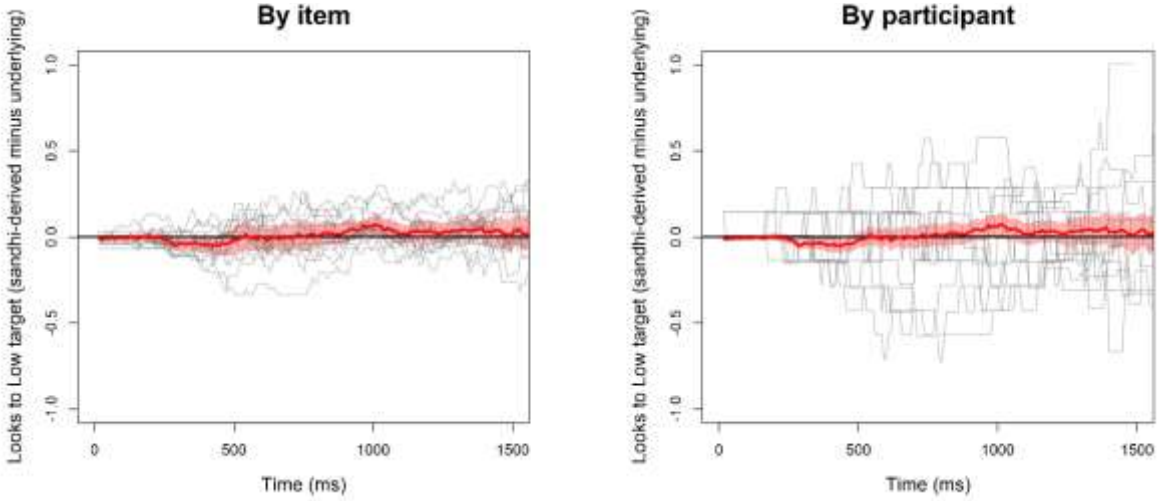
Trials in which the participant did not click on either the Low- or the Rising-tone visual target word were excluded from further analysis. The hypothesis that there were more looks to the Low target when hearing the sandhi-derived than the underlying Rising stimulus was evaluated with a pre-registered cluster-based permutation test [5]. This involves conducting a statistical test (in our case, a dependent-samples t-test between the two conditions) at each sample, forming clusters of adjacent samples that all pass a pre-determined significance threshhold, and then evaluating the significance of a test statistic (in our case, the sum of the test statistics within the largest cluster) non-parametrically by permuting the data and condition labels many times [within each participant]. This allows for testing the whole timecourse for the hypothesis that the two conditions differ, while controlling for multiple comparisons. For an explanation of this test see [5]; the R code we used to implement the test is available in our data analysis script at https://osf.io/ursh9/. In the Results section below, all *p*-values reported are one-tailed *p*-values for this cluster-based permutation statistic testing whether the proportion of fixations to the Low target was higher when hearing a sandhi-derived Rising tone than when hearing an underlyingly Rising tone.

3. RESULTS

Overall, participants clicked on the character corresponding to the Low-tone target on 56% of trials when they heard the stimulus with the sandhi-derived Rising tone, and on 52.6% of trials when they heard the stimulus with the underlyingly Rising tone. This is in the direction one might expect if listeners are sensitive to the difference, but this difference was not significant in a logistic mixed-effects model with maximal random effects for participants and items (b=0.28, z=1.15, p=.249).

The left side of Figure 1 shows the proportion of looks to the Low target over time, as a function of which auditory stimulus was heard. The right side shows the proportion of looks to the Rising target; while we did not perform statistical analyses on these, since we only pre-registered analysis for the looks to the Low target, we show the data here for completeness. For each condition, the proportion of fixations peaks around 50% rather than reaching 100%; this is not surprising, since the stimuli are ambiguous. As in [10], the early portion of the time window shows a counterintuitive pattern, with participants looking more at the putatitively inappropriate target. Crucially, in the latter portion of the window, participants appear to look more at the putatively appropriate target: i.e., for targets corresponding to a word whose first syllable is Low tone, participants look more at these targets when hearing a Rising tone that was derived from a Low tone via tone sandhi, compared to when hearing an underlyingly Rising tone. This trend was not significant, however, in our pre-registered statistical analysis (p=.351). We performed an additional exploratory cluster-based permutation test, relaxing the cluster threshhold to p<.3 (as looser threshholds

Figure 2: Difference between looks to Low target in the two conditions, by participant and by item. Shaded interval represents a two-tailed 95% confidence interval based on the *t* statistic.



are more sensitive for detecting weak but long-lasting effects [5], like the one observed here) and calculating the fixation proportions by item rather than by participant, since the effects for items were more stable. This yielded a p=.098 effect, although p-values are not interpretable for this test since it is not confirmatory.

Figure 2 shows the pattern across participants and items. The pattern is much clearer for items, given that each item had observations from as many as 30 participants per condition, whereas each participant had observations from only 7 or fewer items per condition.

4. DISCUSSION

In a visual world eye-tracking we observed suggestive evidence that participants may be implicitly sensitive to incompletely neutralized.

Overall, the results are inconclusive, consistent with both the presence and the absence of a difference between conditions. On the one hand, the pre-registered analysis did not yield a statistically significant effect, so we are not able to reject the possibility that the tones are completely neutralized in perception. On the other hand, visual inspection of the data suggest that it may not be reasonable to conclude that they are completely neutralized either (keeping in mind that failure to reject the null hypothesis is not the same as acceptance of the null hypothesis), and exploratory analysis suggests that our pre-registered analysis plan may not have been as sensitive as it could have; i.e., if we replicate the study with more items, and perform confirmatory analysis on a more stable proportion measure and use a more appropriate clustering threshhold, then it may be possible to detect a significant effect. While it would be premature to conclude that there is a reliable difference, given that this analysis relied on researcher degrees of freedom [8, 9], the results at least suggest that this paradigm may be capable of providing evidence for incomplete neutralization in the perception of Mandarin Low and Rising tones, and they demonstrate that this issue is worth further investigation and replication.

A limitation of the study is that, while we used a large number of listeners and items, the experiment only used one speaker. Using one speaker is currently standard for studies in this area (Peng [7], Liu [4], and Zhang & Peng [13] each used one speaker in their perceptual tasks, and Wang and Li [11] used two). Nonetheless, acoustic differences between underlying and sandhi-derived Rising tones may not be constant. Therefore, examining the extent to which incompletely neutralized perception generalizes across speakers (if it occurs at all) is a valuable question for future study.

5. REFERENCES

[1] Cousineau, D. (2005). Confidence intervals in within-subject designs: a simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*, 42-45.

[2] Kim, H., & Jongman, A. (1996). Acoustic and perceptual evidence for complete neutralization of manner of articulation in Korean. *Journal of Phonetics, 24*, 295-312.

[3] Kuo, Y., Xu, Y., & Yip, M. (2007). The phonetics and phonology of apparent cases of iterative tone change in Standard Chinese. In *Phonology and Phonetics: Tones and Tunes, volume 2*, 212-237. Berlin: Mouton de Gruyter.

[4] Liu, X. (2013). 上声变调的声学与感知实验研究 [Acoustic and perceptual research on third tone sandhi]. 文教资料 [*Culture and Education Data*], *29*, 139-142. (In Chinese)

[5] Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177-190.

[6] Morey, R. (2008). Confidence intervals from normalized data: a correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*, 61-64.

[7] Peng, S. (2000). Lexical versus phonological representations of Mandarin sandhi tones. In *Language acquisition and the lexicon: Papers in laboratory phonology V*, Michael Broe and Janet Pierrehumbert [Eds.]. pp. 152-167. Cambridge, UK: Cambridge University Press.

[8] Roettger, T. (ms.). Researcher degrees of freedom in phonetic research. https://psyarxiv.com/fp4jr

[9] Simonsohn, U., Nelson, L., & Simmons, J. (2014). P-curve: a key to the file drawer. *Journal of Experimental Psychology: General*, 143, 534-547.

[10] Speer, S., & Xu, L. (2008). Processing lexical tone in third-tone sandhi. Talk presented at *Laboratory Phonology* 11.

[11] Wang, W., & Li, K. (1967). Tone 3 in Pekinese. *Journal of Speech and Hearing Research*, *10*, 629-636.

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

[13] Zhang, C., & Peng, G. (2013). Productivity of Mandarin Third Tone sandhi: A wug test. In Peng G. and Shi F. (Eds.) *Eastward Flows the Great River: Festschrift in Honor of Prof. William S-Y. Wang on his 80thBirthday*, pp. 256-282. Hong Kong: City University of Hong Kong Press.

[14] Zhou, X., & Marslen-Wilson, W. (1997). The abstractness of phonological representation in the Chinese mental lexicon. In *Cognitive Processing of Chinese and Related Asian Languages*, Hsuan-Chih Chen [Ed.], 32-27. Hong Kong, Hong Kong: The Chinese University Press.