Phonological Representation of Tone Sandhi in Nanjing Mandarin

Xin Li and Wentao Gu

School of Chinese Language and Literature, Nanjing Normal University, China
xinli@njnu.edu.cn, wtgu@njnu.edu.cn

Abstract

Although tone sandhi has been well studied in Chinese languages, how sandhi tones are represented in the mental lexicon remains controversial. In Nanjing Mandarin, there are two sandhi types: falling tone sandhi (HL.HL→HH.HL) and low tone sandhi (L.LL→LH.LL) - the latter is the same as in Standard Mandarin. We investigated the phonological representation of tone sandhi in Nanjing Mandarin using a cross-modal picture-word interference paradigm. A group of native speakers of Nanjing Mandarin were asked to name each visual stimulus with a disyllabic target word while listening to the audio stimulus which was a monosyllabic sound functioning as the distractor, and their reaction time was measured using a voice key. In terms of phonological relatedness of the distractor to the first syllable of the target word, three conditions were compared: surface tone, underlying tone, and control conditions. For low-frequency words in both sandhi types, results showed that the reaction time was significantly faster in the underlying tone condition, and was marginally significantly faster in the surface tone condition than the control condition. These suggest that for both sandhi types in Nanjing Mandarin the underlying tone is activated, while the surface tone may also be activated, but at a lesser degree.

Index Terms: tone sandhi, Nanjing Mandarin, phonological representation, cross-modal picture word interference

1. Introduction

Tone sandhi is a phonological process in which lexical tones exhibit contextually determined alternation [1]. Although tone sandhi has been well studied in Chinese languages, little has been known about how tone sandhi is stored in the mental lexicon.

In Standard Mandarin, T3 sandhi is the phenomenon that a low tone T3 preceding another becomes a rising tone similar to T2 (i.e., L.LL→LH.LL), though it remains controversial whether sandhi T3 is equal to T2 or not [2, 3]. The primary debate in the literature lies in whether the mental representation of a disyllabic T3 sandhi word is its underlying form T3T3, or its surface form similar to T2T3, or even both.

Chein et al. [4] adopted an auditory-auditory priming lexical decision paradigm in which each target word of T3T3 (e.g., “描号”/fu3 dao3/; coach) was preceded by three types of monosyllabic prime, i.e., a T2 prime (e.g., “冠”/fu2/; obey) sharing both segments and surface tone with the initial syllable of the target word, a T3 prime (e.g., “仚”/fu3/; assist) sharing both segments and underlying tone with the initial syllable, and a control prime (e.g., “敷”/fu1/; apply) shared only segments with the initial syllable. Because results showed that the response time elicited by the T3 prime was significantly faster than those elicited by the T2 and the control primes, Chein et al. [4] concluded that the mental representation of a T3 sandhi word in Mandarin is its underlying form T3T3. Studies [5,6] conducted in other experimental paradigms (i.e., cross-modal lexical decision, odd-man-out implicit priming) drew the same or similar conclusion.

The other account considered that sandhi T3 (similar to T2) is a tonal variant of T3 which is also stored in the mental lexicon, possibly under the node of an abstract T3 [7]. Thus, the tonal variant of T3 can be activated automatically when a monosyllabic T3 word is heard [8]. Nixon et al. [9] reported that naming of a sandhi picture (e.g., “钞表”/miao3 biao3/; stopwatch) was faster when there was a T2 (e.g., “描”/miao2/; trace) or T3 (e.g., “薇”/miao3/; despise) distractor in a picture-word interference experiment. In a follow-up experiment, Nixon et al. [9] found that visually presented sandhi distractors facilitated naming of T2 and T3 pictures. These indicate that multi-level phonological processing occurs not only in overt production, but also in visual processing of distractor words. It should be noted that they considered sandhi T3 and T2 to be different tone categories in spite of similar acoustic manifestations. The T2 facilitation effect in [9] was due to the activation of context-specific representation that is similar to T2. Their results suggest that both underlying and surface forms of T3T3 are stored in the mental lexicon. Additionally, some studies pointed out the possibility of surface tone representation [10, 11]. For example, a study based on an implicit priming paradigm suggested that T3T3 might be stored as T2T3 and retrieved as such during speech production [11].

Tone sandhi exists not only in Standard Mandarin but also in a number of Chinese dialects, showing differing phonological representations. For example, Shanghai Wu Chinese has a left- dominant sandhi pattern, i.e., the tone of the initial syllable of a disyllabic word spreads rightwards, e.g., 53 + X → 55 + 31, where X refers to any non-checked tone [12]. A study based on an auditory-auditory lexical decision paradigm [13] showed that Shanghai sandhi words were stored in surface form because only surface priming effect was observed. In contrast, Kunming Chinese has a right- dominant sandhi pattern, including both T1 sandhi (44 + X → 35 + X, where X refers to any tone other than 44) and T3 sandhi (53 + X → 55 + X, where X refers to any tone), for which the surface tone representation was supported using the same paradigm [14]. In Taiwan Southern Min Chinese, tone sandhi is in a circular chain-shift fashion, occurring on all non-checked tones at non-phrase-final syllables, for both underlying and surface priming effects were observed [15]: Underlying form is predominant for the 24→53 sandhi, while surface form is predominant for the 51→55 sandhi. It was argued that more productive sandhi (i.e., the sandhi rules that apply also frequently in novel words), elicited more underlying priming, while less productive sandhi yielded more surface priming [15].
In sum, the phonological representations of tone sandhi in the mental lexicon vary with dialects. These diverse results were interpreted in the way that ‘opacity’ and ‘locality’ of the phonological process play key roles in how words undergoing the process are represented and accessed [13, 16]. To have a more global view, however, more studies on other dialects need to be conducted. Nanjing Mandarin, as a right-dominant tone sandhi dialect, has up to six sandhi rules, among which T3 sandhi is almost the same as in Standard Mandarin (i.e., the initial low tone becomes a rising tone similar to T2, LL.LL→LH.LL), while T1 sandhi indicates that the falling tone T1 preceding another changes into a high tone similar to T4 (i.e., HL.HL→HH.HL) [17]. It is therefore of particular significance to investigate the phonological representation of sandhi tones in Nanjing Mandarin.

2. Experiment

This study adopted the picture-word interference paradigm [18], in which participants were asked to name the pictures while listening to auditory distractor words. Each picture name and the corresponding distractor word shared a certain phonological component or were unrelated to each other. In this way, whether the shared component was involved in the phonological encoding process or not could be examined.

2.1. Participants

Twenty-nine native speakers of Nanjing Mandarin were recruited. The data of four participants were excluded from analysis, because the parents of two speakers were not from Nanjing and the other two speakers showed atypical Nanjing Mandarin tone sandhi patterns. The remaining twenty-five speakers (15 female; mean age = 24.04), together with their parents, were born and raised in Nanjing, speaking fluent Nanjing Mandarin. None of them had a reported history of any visual or auditory impairment.

2.2. Materials

Since previous studies indicated that lexical frequency affects the processing of tone sandhi [15], lexical frequency was also considered in this study. The stimuli included both pictures and auditory distractors. For each of the two sandhi types, there were 40 targets pictures corresponding to disyllabic sandhi words, of which a half were high-frequency and the other half were low-frequency. Among the 80 target words, 78 were chosen from a dictionary of Nanjing dialect [19] and an online Chinese dictionary [20], and the other two were from daily conversation. Selected from the internet, all pictures were black line drawing on white background. For each target picture/word, there were monosyllabic auditory distractors in three conditions: surface (T2 for T3 sandhi; T4 for T1 sandhi), underlying (T3 for T3 sandhi; T1 for T1 sandhi) and control (in other tones). In addition, 80 filler stimuli (picture-distractor pairs) were used to conceal the purpose of the experiment. The examples of target stimuli are shown in Table 1.

<table>
<thead>
<tr>
<th>Target word (Chinese)</th>
<th>Auditory distractor</th>
<th>Underlying</th>
<th>Surface</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>T3T3: y3san3 (umbrella)</td>
<td>T3: y3 (rain)</td>
<td>T2: y2 (fish)</td>
<td>T1/T4: y4 (jade)</td>
<td></td>
</tr>
<tr>
<td>T1T1: t0q1kua1 (white gourd)</td>
<td>T1: t0q1 (winter)</td>
<td>T4: t0q4 (move)</td>
<td>T2/T3: t0q3 (understand)</td>
<td></td>
</tr>
</tbody>
</table>

All auditory distractors were recorded from a female native speaker of Nanjing Mandarin at a sound-proof booth in Nanjing Normal University. Speech audio was recorded at a 44.1kHz sampling rate using Adobe Audition 3.0. Using Praat [21], all auditory distractors were segmented and then normalized to an intensity of 85 dB to keep the volume constant.

2.3. Word frequency rating

Since there is no existing word frequency dictionary for Nanjing Mandarin, the frequencies of use for the T1 sandhi and T3 sandhi words were rated by 17 native speakers of Nanjing Mandarin via Wenjuanxing, an online questionnaire platform. A 5-point scale was adopted for scoring, with 1 as the least frequent and 5 as the most frequent in their daily use of Nanjing Mandarin. Because two raters did not give disperse scores as others, their ratings were excluded from data analysis. Hence, the average word frequencies of high- and low-frequency sandhi words were 3.84 (SD = 0.26) and 2.31 (SD = 0.40) for T1 sandhi, and 3.42 (SD = 0.32) and 1.98 (SD = 0.31) for T3 sandhi, respectively. T-tests showed that the high-frequency words had significantly higher frequency scores than their low-frequency counterparts (t (38) = 14.287, p < .001; t (38) = 14.325, p < .001).

2.4. Design

Altogether 480 trials corresponding to 480 stimuli (240 targets and 240 fillers) were divided into four blocks in terms of frequency (high vs. low) and sandhi type (T1T1 vs. T3T3). To maintain the participants’ attention, there was a break between the 2nd and 3rd blocks. For each participant, the order of four blocks was pseudo-randomized, with an additional constraint that the two succeeding blocks before/after the break differed in frequency and sandhi type. Stimuli were pseudo-randomly ordered within each block. We adopted a within-subject 3*2 factorial design, with distractor condition (surface, underlying, and control), sandhi type (T1T1, T3T3) and frequency (high, low) as three factors. For each combinatorial condition, there were 20 target stimuli and 20 filler stimuli. The task took approximately 45 minutes.

2.5. Procedure

Participants were tested in a sound-proof booth at Nanjing Normal University, using the E-Prime 3.0 software [22] to present stimuli and a voice key to collect response time. After familiarized with all target pictures and corresponding words, participants were seated in front of a computer monitor. They completed the practice trials before the actual test. Each trial started with a central crosshair on the computer screen for 500 ms to attract participants’ attention. Then, a picture stimulus was presented simultaneously with the auditory distractor. The inter-trial interval (ITI) was set at 500 ms. Participants were instructed to ignore the auditory distractors from the headphone and to name the pictures as quickly and accurately as possible. To validate the pronunciation of sandhi words, a Zoom H5 digital recorder was used to record the spoken responses of the participants. Response time (RT) was calculated starting from the time of presenting the target picture until the voice key was triggered by the participant’s response. The maximum duration of presenting a picture was set at 2000 ms.
3. Results

For response time (RT), pronunciation errors (including failure in sandhi), stuttering (9.4%) as well as outliers (5.3%) beyond the interval between the lower and upper Whiskers (Q1−1.5*QR ~ Q3+1.5*IQR) were removed. RT data of T1T1 from two participants were also excluded (3.7%) since they pronounced differently from others. Finally, 4,896 valid RTs were analyzed. Figure 1 shows the statistics of RTs for all combinations of distractor condition, tone sandhi type, and word frequency. Then, RT was log-transformed into logRT to reduce the skewness of data. The average RT was 784 ms, while the average duration of monosyllabic auditory distractors was 739 ms, verifying that the participants responded after hearing the complete stimuli.

![Figure 1: Means and 95% confidence intervals of response times for all combinations of distractor condition, tone sandhi type, and word frequency.](image)

Linear mixed-effects model was conducted using the afex package [23] in R 3.6.1 [24], with logRT as the dependent variable. Distractor condition, sandhi type, and frequency were fixed effects, while target picture and participant were random effects. With a backward stepwise method based on a likely ratio test (LRT) using the buildmer package [25], the random effects included the intercepts of target picture and participant, as well as the by-participant slopes on frequency, on sandhi type, and on the interaction between frequency and sandhi type. The degree of freedom $df$ of the model was derived on the basis of LRT to obtain the main and interaction effects. Post hoc test with a Bonferroni correction was conducted using the emmeans package [26].

Results showed a significant main effect of distractor ($\chi^2 (2) = 36.20, p < .001$) and a significant interaction effect between sandhi type and distractor condition ($\chi^2 (2) = 6.92, p < .05$). A three-way interaction between frequency, distractor condition and sandhi type was marginally significant ($\chi^2 (2) = 5.09, p = .079$), suggesting a possible effect of frequency on the interaction between sandhi type and distractor condition. Post hoc test showed that for the high-frequency T1 sandhi words, response times in both surface and underlying distractor conditions were significantly faster than in the control distractor condition ($t (4737) = 4.588, p < .0001$; $t (4738) = 4.862, p < .0001$), while for low-frequency words, naming latencies in the surface distractor condition were marginally significantly faster than in the control condition ($t (4740) = 2.288, p = .0665$), and response times in the underlying distractor condition were significantly shorter than in the control condition ($t (4737) = 2.819, p < .05$). For high-frequency T3 sandhi words, no significant difference was found between three distractor conditions, while for low-frequency words, a significant underlying distractor facilitation effect was observed — naming latencies in the underlying distractor condition were shorter than in the control condition ($t (4739) = 2.673, p < .05$). In addition, a trend of surface distractor facilitation was observed ($t (4739) = 2.173, p = .0896$).

4. Conclusions

For T1 sandhi words, both surface and underlying priming effects were observed regardless of word frequency, except that surface priming was only marginally significant in low-frequency words. A possible explanation for this is that the phonology of tone sandhi is shaped by exemplar, thus lower-frequency words show a weaker surface priming.

For T3 sandhi, no significant difference was found among three distractor conditions in high-frequency words, while a significant underlying priming effect and a marginally significant surface priming effect were found in low-frequency words. A possible account for different findings in the two sets of words may be due to the fact that the words associated with some pictures were more abstract than others — thus a longer response time is needed for naming these pictures due to the higher processing burden.

In sum, reaction time was significantly faster in the underlying tone condition and marginally significantly faster in the surface tone condition than in the control condition, suggesting that the underlying tone is activated in both T1 and T3 sandhi types in Nanjing Mandarin, while the surface tone may also be activated, but to a lesser degree. It can then be concluded that in Nanjing Mandarin the underlying forms of sandhi tones are stored in the mental lexicon, but whether the surface forms are also stored in the mental lexicon remains an open question. Another possible interpretation of surface distractor facilitation is not a storage of surface forms in the mental lexicon but an activation in articulation — in this case it is triggered only in overt speech. We are conducting more experiments to further clarify this issue.

5. Acknowledgements

This work is supported jointly by the National Social Science Fund of China (13&ZD189), and the project of Jiangsu Higher Institutions’ Excellent Innovative Team for Philosophy and Social Sciences (2017STD006). We would like to thank all participants in this study.

6. References