

The Role of Tone Sandhi in Speech Production: Evidence for Phonological Parsing

Stephen Politzer-Ahles¹, Jie Zhang¹

¹Linguistics Department, University of Kansas, United States

sjpa@ku.edu, zhang@ku.edu

Abstract

The present study investigates the role of phonological alternation during speech production in Mandarin, using the odd-man-out implicit priming paradigm, a task in which participants respond faster to words in sets that are phonologically homogeneous in some respect than in sets that are phonologically heterogeneous. We test whether priming is obtained when words in a set share the same tones at the underlying level but have different tones at the surface level—i.e., when the set includes a word that undergoes third tone sandhi. We find that sets of words that are heterogeneous at the surface level (i.e., the heterogeneity is due to application of tone sandhi) failed to elicit priming, just as sets of words that are heterogeneous at the underlying level (i.e., the heterogeneity is due to lexical tone). This finding suggests that the phonological alternation was computed abstractly before the initiation of articulation, offering evidence that the progression from underlying phonological representations to articulatory execution may be mediated online by a level at which abstract phonological alternations are processed.

Index Terms: tone sandhi, speech production, implicit priming, Mandarin

1. Introduction

Traditional generative approaches to phonology explain contextual alternations by assuming a mapping between input and output (underlying and surface) forms and a system of predictable changes that may be applied to input forms. This is only intended to be an account of speakers' competence, however, rather than a performance model; there is little direct evidence that speakers must constantly "do phonology" as they speak and listen [1,2], and current models of speech production have little to say about when and how phonological alternation happens. For instance, Levelt and colleagues' model of speech production [3], one of the most explicit models to date, assumes that coarticulatory variation mostly falls out from overlap in motor gestures—in other words, the model treats most alternation as an unconscious reflex of articulation, and does not assume a separate cognitive level for the computation of phonological alternation.

There are clearly, however, many kinds of alternations that cannot be handled by articulatory heuristics alone—alternations that are not coarticulatory or phonetically natural in nature, that differ across languages or registers, or that interact with morphosyntactic structure. It seems reasonable, then, to suppose that some phonological alternations are computed prior to being translated into motor commands.

To test whether phonological alternation is computed before or during articulatory preparation, we adopted the *implicit priming* (also known as *form preparation*) paradigm, which allows the experimenter to see what units are active during phonological encoding prior to the initiation of articulation [4]. In implicit priming, participants memorize small sets of words (e.g. {loner, local, lotus} or {loner,

beacon, major}) that are paired with various cues, and are asked to say the words as quickly and accurately as possible when they see the cues. Reaction times tend to be faster when the targets are phonologically homogeneous in terms of some portion of the word (e.g., in the first example set above, where all words begin with [lou]). This occurs because when the items have homogeneous onsets the participants are able to prepare at least part of the response word even before they see the cue. In the present study we focus on Mandarin Chinese, which has a phonological alternation—third-tone sandhi—that lends itself well to testing via implicit priming. Below we briefly summarize relevant implicit priming findings in Chinese, and the alternation tested in this study.

Chen and colleagues [5] (see also [6,7]) showed that implicit priming is obtained for sets of bisyllabic Chinese compounds when the segmentals and tones of all the target words' first syllables are the same, even if the characters are different (for instance, the set *fei³cui⁴* 翡翠, *fei³die²* 匪谍, *fei³ce⁴* 悱恻, *fei³bang⁴* 诽谤).¹ On the other hand, when the target words' first syllables were segmentally homogeneous but differed in tone (for instance, the set *fei¹ji¹* 飞机, *fei²pang⁴* 肥胖, *fei³cui⁴* 翡翠, *fei⁴yan²* 肺炎), the priming effect was still present but was much smaller. Likewise, in explicit priming, targets preceded by segmentally and tonally identical primes show significant facilitation, whereas targets preceded by primes that were only segmentally identical but that differ in tone show no [8] or reduced [9] priming. In short, tone is part of the linguistic representation that must be phonologically encoded and that matters for priming: tonal heterogeneity in a set spoils implicit priming, and tonal mismatch between a prime and target spoils explicit priming.

1.1. Third-tone sandhi and the present study

Mandarin has a tone sandhi rule whereby a third tone (T3, written as 213 in Chao numbers) followed by another third tone changes into a second tone (T2, written as 35):

$$213 \rightarrow 35 / _ .213 \quad (1)$$

Third-tone sandhi is phonological in nature (i.e., it does not have strong phonetic motivation) and exceptionless [10,11]. Third-tone sandhi is incompletely neutralizing in the acoustic domain: a sandhi-derived T2 has a lower F0 (i.e., is more like T3) than a lexical T2 [11-14]. The neutralization is even less complete in wug words than existing words [10]. On the other hand, there is conflicting evidence as to whether the neutralization is complete perceptually. Early studies showed that listeners were unable to consciously discriminate sandhi-derived T2 from lexical T2 [13,15], but recent evidence from visual world eye-tracking suggests that the perceptual difference does influence listeners' processing [16].

Several studies have examined whether listeners hearing a syllable with surface T2 in a sandhi context activate its T3 counterpart (i.e., whether they undo the tone sandhi to retrieve the underlying representation). In a priming experiment, Speer

Table 1. A sample stimulus set. Cue words corresponding to each target are shown in parentheses.

Three-Item	Homogeneous	Heterogeneous-T3	Heterogeneous-T2	Heterogeneous-Unrelated
(市场) 企业 qi^3ye^4				
(关机) 启动 qi^3dong^4				
(街头) 乞丐 qi^3gai^4				
	(出发) 起身 qi^3shen^1	(开始) 起点 qi^3dian^3	(爱国) 旗帜 qi^2zhi^4	(制冷) 空调 $kong^1tiao^2$

and Xu [16] found comparable priming effects for T2 and T3 targets following T2 primes, albeit not across all tasks they tested. They also found that T2 words that have T3 counterparts in the language were responded to more slowly than T2 words that did not, suggesting that listeners hearing T2 activate both T2 and T3 representations only if their lexicon includes a valid T3 counterpart to that syllable. In a concept formation task, Peng [13] found that participants were less accurate when trained to categorize surface T2 in sandhi-appropriate contexts as part of the same category as lexical T2, and more accurate when trained to categorize it with lexical T3; these results also suggest that listeners automatically undo third-tone sandhi. Zhou and Marslen-Wilson [17] performed a series of auditory-auditory priming experiments with sandhi T2, and found some evidence that may suggest sandhi T2 and lexical T2 are stored differently. In their experiments, T3 primes facilitated access of sandhi-derived T2 targets (Experiment 1) but inhibited access of lexical T2 targets (Experiment 2). Furthermore, words with sandhi-derived T2 took longer to recognize than words with lexical T2, although there are several possible causes for the slowdown. Also, Xu [18], using a short-term memory recall task, found that word lists including both lexical T2 and sandhi-derived T2 tones affected memory recall like homogeneous lists would, suggesting that tone sandhi was computed subconsciously and obligatorily during the task.

In sum, empirical studies on the online use of phonological knowledge during perception of tone sandhi are scarce, and the results are not unequivocal. In the present study we turn to production, examining the role tone sandhi plays during implicit priming by building upon Chen and colleagues' [5] finding that tonal heterogeneity reduces the implicit priming effect. Specifically, we examine whether implicit priming is spoiled for sets that are heterogeneous because of words undergoing tone sandhi, rather than sets that are heterogeneous because of words with underlyingly different tone. To do this, we had participants produce sets of words that all began with third tone at the underlying level, but which included one word whose first syllable changes to second tone because of tone sandhi. We compared the disruption of implicit priming effect in these sets to the disruption caused when a member of the set began with second tone at the underlying level.

2. Methods

2.1. Participants

Thirty native speakers of Mandarin (16 females; age 18-42, mean 23.3) from the University of Kansas community participated in the study. An additional seven participants were excluded from the analysis because they produced incorrect words or nonstandard pronunciations that influenced the intended hetero/homogeneity of one or more sets. All participants provided their informed consent and received payment. All methods were approved by the Human Subjects Committee of Lawrence.

2.2. Materials

Five critical sets of word pairs were prepared for the experiment. Each pair was made of two two-character words, with the first word serving as a cue and the second as a target (see Table 1 for a sample set of cues and targets). The two words in the pair always had a clear semantic or associative relationship. Each set had three critical word pairs (e.g., the pairs 市场~企业, 关机~启动, and 街头~乞丐 in Table 1) and several possible "odd-man" pairs, which differed depending on the condition (see below). The three critical pairs were always present regardless of condition, and the target words in these pairs met the following criteria. The first phone of each word was either a stop or an affricate. The first syllable was always third-tone, and the second syllable was always a different tone (but never the light tone, *qing sheng*, which appears on stressless syllables); any differences on the first syllable's tone caused by coarticulation with the second should be minimal, as Mandarin third tone is not very susceptible to anticipatory coarticulation [12]. The first syllables of all three critical words were phonologically identical but written with different characters.

Depending on the condition, a set of three critical pairs could also be presented with a fourth, "odd-man-out" pair (see Table 1). In the Homogeneous condition, the fourth target had the same properties as the three critical items: its first syllable was identical to the other targets' in terms of segmentals and tone. In the three heterogeneous conditions, the fourth item was an "odd-man-out"—an item that differed phonologically from the other items in such a way that it spoiled the homogeneity of the set. In the Heterogeneous-T3 condition (the condition of interest), the first syllable of the odd-man target (起 qi^3 in Table 1) was segmentally identical to the other items, but undergoes sandhi changing it to second tone, such that the set of four pairs was homogeneous at the input (underlying) level but heterogeneous in tone at the output (surface) level. In the Heterogeneous-T2 condition, the first syllable of the odd-man target was segmentally identical to the other items but had lexical second tone, making the set heterogeneous at both the input and output levels. Finally, in the Heterogeneous-Unrelated condition, the first syllable of the odd-man target shared neither segmentals nor tone with the other three targets. (In a fifth condition, Three-Item, only the three critical pairs were presented.) Across conditions, an effort was made to make sure the "fourth items" had similar lexical frequency, as measured by the SUBTLEX-CH word form corpus [19].

In addition, to distract participants from the third-tone manipulation, five filler sets were prepared. Unlike the critical sets, none of the filler sets included all third-tone targets. Two filler sets were homogeneous (with three cue-target pairs), two were heterogeneous in terms of both segmentals and tones (with three pairs), and one was homogeneous except for a heterogeneous unrelated odd-man-out pair.

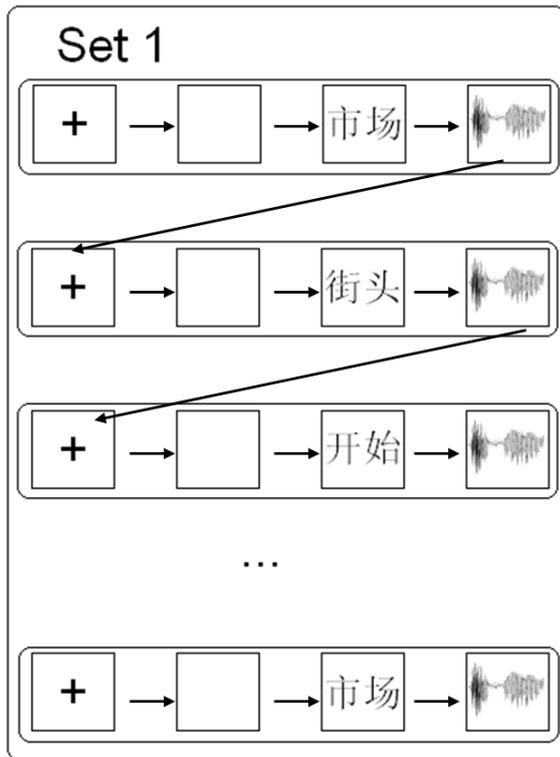


Figure 1. Schematic trial flow for the test phase of a set. Each set consists of multiple trials, each of which consists of a fixation point, a jittered blank screen, and presentation of the cue word until the participant's vocal response, which initiates the next trial. After all trials in a given set are completed, the training phase for next set begins; following the training phase, the test phase for the next set follows the same procedure as that shown above, using the new set of target words.

2.3. Design and procedure

The five critical sets were organized into five lists in a Latin square design. Sets were not repeated within lists. The presentation and timing of stimuli was controlled by Presentation software (<http://www.neurobs.com>). During the experiment, the presentation order of the five critical sets was randomized, and each critical set was preceded by one of the filler sets. The main experiment was preceded by a practice block, using one heterogeneous set of three pairs (none of which were used in the formal experiment), which followed the same procedure as the main experiment.

Each set included a training phase and a test phase. During the training phase, the three or four critical pairs (for instance, the four pairs under the Homogeneous column in Table 1: 市场~企业, 关机~启动, 街头~乞丐, and 出发~起身) were presented simultaneously in Chinese characters at the center of the screen. The cue-target pairing was always maintained, but the order of the pairs was randomized. While the written words remained on the screen, auditory tokens of both the cues and targets were played once to the participant over speakers; the auditory stimuli were produced by a female native speaker from Beijing who was naïve to the purpose of the study and did not participate in the experiment. Participants were allowed to view the words for as long as they needed to memorize the cue-target pairings before pressing a button to move on to the test phase.

During the test phase (see Figure 1 for an example), the cue words for each set were presented in a random order and the participants responded by saying the associated target words as quickly as possible into a head-mounted microphone. Within each set, each cue word was repeated four times, yielding 12-16 trials (depending on whether the set included three or four cue words); the order of the 12-16 trials was fully randomized. Each trial began with a "+" presented at the center of the screen for 500 ms, which participants were instructed to fixate on. After the fixation point, the screen remained blank for 350, 600, 850, or 1300 ms (the duration was selected randomly at runtime for each trial). Next, one of the cue words (e.g., 市场) was presented at the center of the screen. The recording began at the moment the cue word appeared, and continued for 2 seconds. When the participant's vocal response exceeded a pre-defined sound threshold (e.g., when she spoke *qi³ye⁴*), the word disappeared from the screen. The screen remained blank for 1100 ms after the initiation of the participant's vocal response, after which time the fixation point for the next trial was presented. The whole experiment took approximately fifteen minutes.

2.4. Data analysis

Each participant's recorded responses were listened to by the first author and coded as either correct; incorrect; beginning with a nonspeech sound; beginning with a filled pause, hesitation, or self-correction; or no response. Response onset latencies were measured manually using Praat (<http://praat.org>). Only correct responses to cue words in the critical conditions were included in the analysis of response times. Response times to the "odd-man-out" words were not included—a basic tenet of the odd-man-out design is that adding the fourth item to the set creates heterogeneity and spoils the priming effect for all items in the set, even if the odd-man-out item itself is not included in the measurements.

3. Results and discussion

3.1. Accuracy

Across participants, 88% of trials were responded to correctly; 2.4% were incorrect; 0.5% began with nonspeech sounds; 1.9% began with a hesitation, filled pause, or self-correction; and 7.2% had no response within the allotted time. Participants did not differ across conditions in terms of number of correct responses, $F(3.87, 112.19) < 1$.

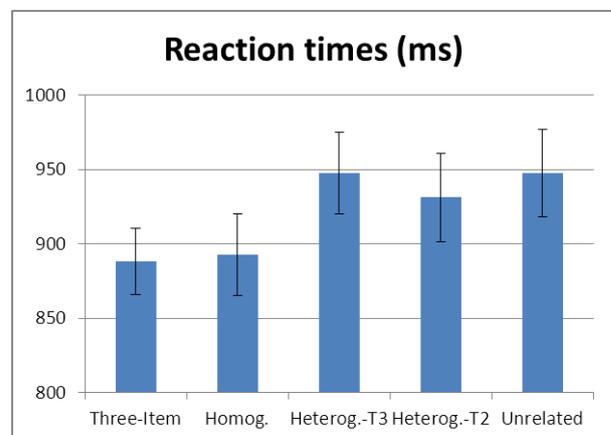


Figure 2. Mean reaction times to each condition. Error bars represent standard error.

3.2. Reaction times

Figure 2 shows the mean reaction times (by-participants, averaging across items and across repetitions within each set) to each condition. It seems that participants tended to respond fastest in the Three-Item and Homogeneous conditions, and responded more slowly in all Heterogeneous conditions (Heterogeneous-T3, Heterogeneous-T2, and Heterogeneous-Unrelated).

Participants were also slower to respond to the first several trials in a set and faster to respond to later trials, reflecting an effect of familiarity with the items; statistical analyses indeed revealed a small but significant negative correlation between a trial's reaction time and its order within a set, $r = -.103$, $p < .001$. Thus, to investigate the effects of set hetero/homogeneity while accounting for variability due to the order of presentation, we used linear regression, sequentially regressing reaction times on Order-in-Set, the dummy-coded Condition variable, and the interaction term between these variables.² The results are summarized in Tables 2 and 3.

The analysis revealed a significant effect of Condition (i.e., the quality of the regression model improved when different means were allowed for different conditions), indicating that the mean reaction times significantly differed across sets even after accounting for variability due to order (Table 2). On the other hand, the interaction of Condition and Order was not significant. Table 3 shows the regression coefficients for the final regression model (that including the Order and Condition factors but not the interaction). Particularly, the unstandardized b coefficients represent the priming effects for each condition. Both the Three-Item and Homogeneous conditions elicited significant priming effects (55 ms and 48 ms, respectively), whereas the 23-ms numerical effect for Heterogeneous-T2 did not reach significance, and Heterogeneous-T3 did not yield priming at all. Re-coding the Condition variable to use Heterogeneous-T3, rather than Heterogeneous-Unrelated, as the baseline for comparison showed that the mean response time to Heterogeneous-T3 was significantly slower than for Three-Item and Homogeneous

Table 2. Summary of the sequential regression. Each row shows the factors added at each step and the significance of the change in R^2 resulting from that step. Significant effects are indicated with an asterisk.

Factors added	R^2	df	F_{change}	p
Order	.009	1, 1614	15.45	<.001*
Condition	.020	4, 1610	4.13	.002*
Order*Condition	.022	4, 1606	1.07	.371

Table 3. Regression coefficients. The unstandardized b coefficient for each condition (Three-Item, Homogeneous, Heterogeneous-T3, and Heterogeneous-T2) represents the difference in seconds between that condition's mean response time and the mean response time for Heterogeneous-Unrelated. Significant coefficients are indicated with an asterisk.

Coefficient	b	β	t	p
(Constant)	988.4			
Order	-6.0	-.109	-4.31	<.001*
Three-Item	-54.7	-.089	-2.78	.005*
Homogeneous	-48.4	-.078	-2.48	.013*
Heterogeneous-T3	7.5	-.012	-0.39	.698
Heterogeneous-T2	-22.6	-.037	-1.17	.244

($p = .001$ and $p = .004$, respectively), and did not differ from Heterogeneous-T2 or Heterogeneous-Unrelated ($p = .115$ and $p = .698$).

3.3. Discussion

Building upon previous research in implicit priming and the production of Mandarin third-tone sandhi, the present study investigated whether tone sandhi is abstractly computed prior to articulation. We found that sets of words which were heterogeneous at the surface level (because of tone sandhi) behaved similarly to sets of words which were heterogeneous at the underlying level (because of lexical tone): both types of sets failed to yield an implicit priming effect. The comparable lack of priming for sandhi-derived and underlyingly heterogeneous sets (Heterogeneous-T3 and Heterogeneous-T2, respectively) suggests that it was the surface form, not the underlying form, that mattered when participants were preparing speech.

Implicit priming effects are argued to stem from the planning of linguistic units before production [4-6,20] (but see [21]). Thus, the lack of implicit priming effects for sets with a sandhi-derived odd-man out indicates that what participants were trying to prepare was a heterogeneous set, in other words, a set in which tone sandhi had already applied to one of the words. Thus, the results are not consistent with a model in which speakers only prepare speech based on underlying forms and then allow alternation to happen heuristically at the level of articulation; rather, they suggest that an additional abstract level of phonological alternation intercedes between word-form retrieval and the articulation, at least in the case of this sandhi. The word forms sent to the articulatory system were not underlying forms, but forms derived as output of a phonological operation.³ This is in line with Chen's proposal, based on speech-error data, that tone sandhi applies before phonetic spellout and articulation [25]. These results should not come as a surprise, given that there are many theoretical reasons to assume that not all alternation can result from articulatory heuristics (see the Introduction).

A surprising result from the present study is the failure for priming in segmentally homogeneous but tonally heterogeneous conditions (e.g., Heterogeneous-T2 and Heterogeneous-T3) to reach significance. In previous implicit studies investigating tone [5,7], the priming for such sets was small (12 ms and 16 ms, respectively), but significant; in the present study, Heterogeneous-T2 elicited a larger numerical effect (23 ms in the regression analysis, 19 ms in by-participants raw means, or 27 ms in by-items raw means), but did not reach significance in any analyses. This is likely due to our different experimental design: in the present study, sets were not repeated within participants, and thus comparisons within participants were made across sets. These factors were not confounded, since the Latin square design allowed sets to be counterbalanced across participants, but they may have caused greater variability in the data and made it more difficult to identify weak effects. Nevertheless, the crucial effect—the lack of priming for Heterogeneous-T3 sets—was very far from significance ($p = .698$) and thus cannot be attributed to any lack of power.

Another possible explanation for the lack of priming in the Heterogeneous-T2 and Heterogeneous-T3 conditions is the use of the odd-man-out version of the implicit priming task. While this task has successfully yielded robust results in several studies [20,26], in Chen and Chen's Chinese odd-man-out experiment [27] (Experiment 1B) the unrelated odd-man-out did not completely spoil priming. RTs were slightly faster for odd-man-out sets than for completely heterogeneous sets;

Chen and Chen speculate that this may be due to participants' use of response strategies. Thus, it is possible that implicit priming failed to appear in the tonally heterogeneous sets because they were being compared to a baseline that also exhibited some priming. This possibility does not influence the main finding of the study—that Heterogeneous-T3 and Heterogeneous-T2 behaved similarly with respect to implicit priming. Nevertheless, it will be worthwhile in future studies to see if this finding is replicated when using a design and analysis method that tests spoil effects rather than priming effects [28].

Another unexpected finding is the direction of the difference between Heterogeneous-T3 and Heterogeneous-T2 response times. While the comparison between Heterogeneous-T3 and Heterogeneous-T2 did not reach significance, Heterogeneous-T3 did show a surprising trend towards being slower than Heterogeneous-T2. Such an effect could not be due to computational resources used at the moment the speaker is producing the sandhi-derived word, since these words (the "odd-man-out" words) were not included in the analysis. It will be worthwhile in future studies to examine whether this trend is robust, and to consider what it might tell us about the online production of third-tone sandhi.

The results of the present study provide evidence that some alternations—at the very least, those that are very phonological in nature, like third-tone sandhi—must be precompiled at some point before the initiation of articulation. This suggests that models of speech production must also include a level of phonological input-output operations, before discrete phonological outputs (surface representations) are sent to the articulatory system to be converted into continuous articulatory programs. An alternative explanation for the present results, however, is that the words undergoing third-tone sandhi, being compound words, are simply stored in the lexicon with the sandhi already computed. In other words, no phonological operation is necessary because, counter to the assumption presented in this paper, these compounds begin with second tone at the underlying level, even if their constituent morphemes do not. Indeed, competence models based on selection of stored allomorphs have been proposed for Taiwanese tone sandhi [29,30]; Zhou & Marslen-Wilson [17] also test such a model for Mandarin, although their results are not entirely consistent with full storage. Furthermore, Zhang and Lai [10] show that third-tone sandhi applies less accurately to wug words than to real words, demonstrating that lexical storage probably matters for the realization of sandhi. To address this alternative account, then, the present study will have to be extended to test novel compounds that undergo tone sandhi. If a similar pattern of results is obtained, we will have strong evidence for the online performance of phonological alternation during speech production. Given that Xu's [18] short-term memory experiment showed that tone sandhi was applied online to novel compounds, our hypothesis is that the pattern of effects in the present study will also be observed in an experiment with novel compounds; this remains an empirical question, however, which needs to be tested. Another interesting extension of the present work would be to test tone sandhi in phrases, occurring across word boundaries, as this is another instance of tone sandhi that lexical storage of surface forms likely could not account for [17].

4. Conclusions

The present study compared modulations of the implicit priming effect due to tonal heterogeneity introduced by either underlying or sandhi-derived second tone in Mandarin

Chinese. The results showed that sandhi-derived second tone behaves like underlying second tone in blocking the implicit priming effect, suggesting that tone sandhi is computed at an abstract level before articulation—although an alternative hypothesis that compounds are lexically listed with sandhi pre-compiled has yet to be tested. The implication is that speech production requires a mechanism for phonologically-based alternation in addition to gradient alternation that arises at the level of articulatory encoding. The method presented here may prove useful for investigating phonological alternations in other languages, and for investigating coarticulatory alternations or tone sandhis with varying degrees of phonetic motivation [10,31]—these topics could shed new light on the relationship between phonetic and phonological encoding during speech production.

5. Acknowledgements

The authors would like to thank the members of the University of Kansas Phonetics & Psycholinguistics Laboratory (KUPPL) for feedback on this research.

6. Notes

1. Mandarin has four lexical tones; here I indicate them with 1 (corresponding to a High tone, 55 in Chao numbers), 2 (Rising, 35), 3 (Low, 213), and 4 (Falling, 51).
2. The data were also analyzed using repeated measures ANOVAs, both by participants and by items, on the condition means; the overall pattern of results was the same as that shown in the regression analysis.
3. Note that the output of the phonological operation is not necessarily a fully-specified surface form or articulatory program; rather, the phonological tone sandhi alters the syllable's tonal target, but other phonetic and articulatory factors may further affect the implementation of those targets [22-24]. That is to say, it is possible that the representations being used in the implicit priming task are abstract, intermediate phonological representations mediating between the lexical underlying forms and the surface articulatory programs; this hypothesis needs to be tested in further studies.

7. References

- [1] Ladefoged, P., "What are linguistic sounds made of?", *Language*, 56:485-502, 1980.
- [2] Xu, Y., "Principles of tone research", *Proceedings of the International Symposium on Tonal Aspects of Language*, 3-13, 2006.
- [3] Levelt, W., Roelofs, A., and Meyer, A., "A theory of lexical access in speech production", *Behavioral and Brain Sciences*, 22:1-75, 1999.
- [4] Levelt, W., "Models of word production", *Trends in Cognitive Sciences*, 3:223-232, 1999.
- [5] Chen, J., Chen, T., and Dell, G., "Word-form encoding in Mandarin Chinese as assessed by the implicit priming task", *Journal of Memory and Language*, 46:751-781, 2002.
- [6] O'Seaghdha, P., Chen, J., and Chen, T., "Proximate units in word production: phonological encoding begins with syllables in Mandarin but with segments in English", *Cognition*, 115:282-302, 2010.
- [7] Zhang, Q., "Phonological encoding in monosyllabic and bisyllabic Mandarin word production: implicit priming paradigm study", *Acta Psychologica Sinica*, 40: 253-262, 2008.
- [8] Lee, C., "Does horse activate mother? Processing lexical tone in form priming", *Language and Speech*, 50:101-123, 2007.
- [9] Lee, Y., *The role of lexical tone in spoken word recognition of Chinese*, M.A. thesis, University of Kansas, 2008.
- [10] Zhang, J., and Lai, Y., "Testing the role of phonetic knowledge in Mandarin tone sandhi", *Phonology*, 27: 153-201, 2010.

- [11] Kuo, Y., Xu, Y., and Yip, M., "The phonetics and phonology of apparent cases of iterative tone change in Standard Chinese", in *Phonology and Phonetics: Tones and Tunes, Volume 2: Experimental Studies in Word and Sentence Prosody*, 212-237, Mouton de Gruyter, 2007.
- [12] Xu, Y., "Contextual tonal variations in Mandarin", *Journal of Phonetics*, 25:61-83, 1997.
- [13] Peng, S., *Phonetic implementation and perception of place coarticulation and tone sandhi*, Ph.D. dissertation, The Ohio State University, 1996.
- [14] Zee, E., "A spectrographic investigation of Mandarin tone sandhi", *UCLA Working Papers in Linguistics*, 49:98-116, 1980.
- [15] Wang, W., & Li, K., "Tone 3 in Pekingese", *Journal of Speech and Hearing Research*, 10:629-636, 1967.
- [16] Speer, S., & Xu, L., "Processing lexical tone in third-tone sandhi", Talk presented at Laboratory Phonology 11, 2008.
- [17] Zhou, X., & Marslen-Wilson, W., "The abstractness of phonological representation in the Chinese mental lexicon", in *Cognitive Processing of Chinese and Related Asian Languages*, ed. Chen, H., 3-27, The Chinese University Press, 1997.
- [18] Xu, Y., "Depth of phonological recoding in short-term memory", *Memory and Cognition*, 19:263-273, 1991.
- [19] Cai, Q., and Brysbaert, M., "SUBTLEX-CH: Chinese word and character frequencies based on film subtitles", *PLoS ONE*, 5(6), e10729, 2010.
- [20] Afonso, O., and Álvarez, C., "Phonological effects in handwriting production: evidence from the implicit priming paradigm", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37:1474-1483, 2011.
- [21] Kawamoto, A., "Incremental encoding and incremental articulation in speech production: evidence based on response latency and initial segment duration", *Behavioral and Brain Sciences*, 22:48-49, 1999.
- [22] Xu, Y., & Wang, Q., "Pitch targets and their realization: evidence from Mandarin Chinese", *Speech Communication*, 33:319-337, 2001.
- [23] Xu, Y., "Sources of tonal variations in connected speech", *Journal of Chinese Linguistics*, 17:1-31, 2001.
- [24] Xu, Y., "Understanding tone from the perspective of production and perception", *Language and Linguistics*, 5:757-797, 2004.
- [25] Chen, J., "The representation and processing of tone in Mandarin Chinese: evidence from slips of the tongue", *Applied Psycholinguistics*, 20:289-301, 1999.
- [26] Roelofs, A., "The influence of spelling on phonological encoding in word reading, object naming, and word generation", *Psychonomic Bulletin and Review*, 13:33-37, 2006.
- [27] Chen, T., and Chen, J., "Morphological encoding in the production of compound words in Mandarin Chinese", *Journal of Memory and Language*, 54:491-514, 2006.
- [28] Cholin, J., Schiller, N., and Levelt, W., "The preparation of syllables in speech production", *Journal of Memory and Language*, 50:47-61, 2004.
- [29] Zhang, J., Lai, Y., and Sailor, C., "Modeling Taiwanese speakers' knowledge of tone sandhi in reduplication", *Lingua*, 121:181-206, 2011.
- [30] Tsay, J., and Myers, J., "Taiwanese tone sandhi as allomorph selection", *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, 22:394-405, 1999.
- [31] Chen, M., *Tone sandhi: patterns across Chinese dialects*. Cambridge: Cambridge University Press, 2000.