Proceedings of Meetings on Acoustics

Volume 18, 2012 http://acousticalsociety.org/

164th Meeting of the Acoustical Society of America
Kansas City, Missouri
22 - 26 October 2012
Session 4aSC: Speech Communication

4aSC1. The role of phonological alternation in speech production: evidence from Mandarin tone sandhi

Stephen Politzer-Ahles* and Jie Zhang

***Corresponding author's address: Department of Linguistics, University of Kansas, 1541 Lilac Lane, Lawrence, Kansas 66044-3177, sjpa@ku.edu**

We investigate the role of phonological alternation during speech production in Mandarin using implicit priming, a paradigm in which participants respond faster to words in sets that are phonologically homogeneous 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 a phonological alternation which changes the tone. Sets that are heterogeneous at the surface level (in which the heterogeneity is due to a phonological operation) failed to elicit priming, as did sets that are heterogeneous at the underlying and surface levels (in which the heterogeneity is due to the lexical representations). This finding suggests that the phonological alternation was computed before the initiation of articulation, offering evidence that the progression from underlying phonological representations to articulatory execution may be mediated online by phonological input-to-output mapping. Furthermore, sets of words that are heterogeneous only at the surface level showed a different trend than sets of words that are heterogeneous at both levels, suggesting that both the surface and underlying levels of representation play a role during speech production.

Published by the Acoustical Society of America through the American Institute of Physics

Received 3 Oct 2012; published 5 Dec 2012

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. For instance, while the word handbag presumably has an underlying form /hændbæg/, context-based phonological constraints cause it to be pronounced [hæmbæg]. This is intended to be an account of speakers' competence, however, rather than a performance model. There is little direct evidence that speakers must "do phonology" as they speak and listen (Ladefoged, 1980), and current models of speech production have little to say about when and how phonological alternation happens. For instance, Levelt and colleagues' (1999) model of speech production assumes that coarticulatory variation results from overlap in motor gestures—in other words, the model treats most alternation as an unconscious reflex of the physical process 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. The current study presents evidence that both underlying representations and underlying-to-surface form mapping operations play a role during online speech production.

To test whether morpho-phonological alternation is computed before articulatory preparation, we adopted the *implicit priming* (also known as *form preparation*) paradigm, which allows the experimenter to infer what units are active during phonological encoding prior to the initiation of articulation (Levelt, 1999). In implicit priming, participants memorize small sets of words (e.g. {loner, local, lotus} or {loner, beacon, major}) that are each paired with a cue. Participants are then 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 in a set are phonologically homogeneous in terms of some portion of the word (e.g., in the first example set above, where all words begin with [loʊ]). This facilitation occurs because when the items have homogeneous onsets the participants are able to prepare at least part of the response word 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. Properties of Mandarin third-tone sandhi allow us to test sets of words which are homogeneous at the words' underlying levels but heterogeneous at the words' surface levels, in order to see whether how this phonological alternation influences reaction times in the implicit priming task.

Below we briefly describe Mandarin tone and its role in processing (including relevance to implicit priming), and the phonological and psychological characteristics of third-tone sandhi. We then report the results of two experiments that use the implicit priming paradigm to test the role third-tone sandhi plays during speech production by Mandarin speakers.

1.1. Mandarin lexical tone

Mandarin has four lexical contour tones, which are phonologically distinctive: for example, *shou*^{1} 收 means "collect", *shou*^{2} 熟 means "ripe", *shou*^{3} 手 "hand", and *shou*^{4} 受 "receive" (Zhang, 2010).[1] Chen and colleagues (2002; see also O'Seaghdha et al., 2010; Zhang, 2008) have shown that tone is relevant for implicit priming. They obtained a facilitation effect for sets in which the segmentals and tones of all the target words' first syllables are the same (for instance, the set *fei*^{3}*cui*^{4} 翡翠, *fei*^{3}*die*^{2} 匪谍, *fei*^{3}*ce*^{4} 悱恻, *fei*^{3}*bang*^{4} 诽谤). On the other hand, for sets in which the target words' first syllables were segmentally homogeneous but differed in tone (for instance, the set *fei*^{1}*ji*^{1} 飞机, *fei*^{2}*pang*^{4} 肥胖, *fei*^{3}*cui*^{4} 翡翠, *fei*^{4}*yan*^{2} 肺炎), the facilitation effect was still present but was substantially reduced. In other words, tone is part of the linguistic representation that must be phonologically encoded; tonal heterogeneity in a set reduces implicit priming.

Findings from traditional explicit priming converge with those from implicit priming to suggest that tone, while not fully independent from segmentals, nevertheless plays a role in the activation of Chinese words.[2] Lee (2007) found that targets preceded by segmentally and tonally identical primes showed significant facilitation, whereas targets preceded by primes that were only segmentally identical did not. This result is consistent with the suggestion of Chen and colleagues (2002; see also O'Seaghdha et al., 2010) that the primary unit of retrieval and activation for Chinese words is the complete syllable, with both segments and tone specified. Thus, tonal identity or heterogeneity is necessary but not sufficient to yield substantial priming.

1.2. Tone sandhi

Mandarin has a tone sandhi pattern whereby a third tone (T3) followed by another third tone changes into a second tone (T2):

1) T3 → T2 / __.T3

For instance, while *shui*^{3} 水 "water" normally has third tone, in the compound *shui*^{3}*guo*^{3} 水果 "fruit" it is pronounced with second tone (as [*shui*^{2}*guo*^{3}]). Third-tone sandhi is phonological in nature (i.e., it does not have strong phonetic motivation) and exceptionless (Zhang & Lai, 2010; Kuo et al., 2007; Peng, 1996). Acoustically, third-tone sandhi is incompletely neutralizing: T2 derived via tone sandhi has a lower fundamental frequency (i.e., more similar to the fundamental frequency of T3) than lexically underlying T2 (Kuo et al., 2007; Xu, 1997; Peng, 1996; Zee, 1980). The neutralization is even less complete in wug words than existing words (Zhang & Lai, 2010). On the other hand, there is

[1] In the notation used here, superscript 1 corresponds to a **H**igh tone (55 in Chao numbers), 2 to a **R**ising tone (35), 3 to a **L**ow tone (213), and 4 to a **F**alling tone (51).

[2] The results summarized below concern the role of tone and segmentals in priming of monosyllables. In disyllables, the direction of priming effects differs depending on which syllable the phonological match or mismatch occurs in (see Zhou & Marslen-Wilson, 1997, regarding Mandarin, and Cutler & Chen, 1995, regarding Cantonese).

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 (Peng, 1996; Wang & Li, 1967), but recent evidence from visual world eye-tracking suggests that the perceptual difference does influence listeners' implicit processing (Speer & Xu, 2008).

Of greatest relevance for the present research are studies that 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 and retrieve the underlying representation of the syllable). Since third-tone sandhi is at least partially neutralizing, a [T2.T3] sequence could have the underlying form /T2.T3/ (which corresponds to the surface form) or /T3.T3/ (which yields the surface form after the application of tone sandhi). Speer and Xu (2008) report priming studies that potentially provide evidence that hearing a T2 syllable activates the syllable's T3 counterpart. In a concept formation task, Peng (1996) 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 (1997) present a pair of auditory-auditory priming experiments which suggest that words with T2 derived from tone sandhi may have a different mental representation than words with lexical T2. Xu (1991) demonstrates evidence that speakers generate the surface forms of words subject to tone sandhi in a short-term memory recall task.

Overall, 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 examine the role of tone sandhi during production.

1.3. The present study

The present study examines the role tone sandhi plays during implicit priming by building upon Chen and colleagues' (2002) finding that sets of words in which the underlying tones differ do not show a substantial implicit priming effect. Specifically, we examine whether implicit priming is spoiled for sets that are tonally heterogeneous not because of different underlying tones, but because of different surface tones derived by tone sandhi (see section 2.2 for a detailed description of the stimuli). In Experiment 1, participants produced 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. The disruption of implicit priming effect in these sets was compared to the disruption caused when a member of the set began with second tone at the underlying level. In Experiment 2, we used pseudowords that undergo tone sandhi instead of real words that undergo tone sandhi, to rule out the possibility that the compound words are lexically stored in their surface forms and can thus be accessed directly during production without undergoing tone sandhi.

The aim of the present study is to test how the words undergoing phonological alternations are represented during speech production. If it is their underlying forms that are the representations that are encoded for articulation, then such sets of words should show facilitation like homogeneous sets; if it is their surface forms that are encoded for articulation, then the facilitation for such sets should be reduced. Pilot data reported by Politzer-Ahles and Zhang (2012) suggested that sets that are homogeneous at the underlying level but heterogeneous at the surface level do not show priming—in other words, that the surface forms of the words in these sets were derived via morpho-phonological mapping before the initiation of articulation. However, that study did not address the possibility that compound words are simply listed in the lexicon according to their surface forms. If lexical listing of surface forms were the case, the critical sets may have patterned like heterogeneous sets not because of phonological mapping but because these sets are in fact heterogeneous at the underlying level as well as the surface level. Therefore, in the present study we report two experiments: one in which the words undergoing tone sandhi are real words, and one in which they are novel compounds. If the loss of priming for sets including a sandhi-undergoing word is due to lexical storage of the surface forms, then priming should be observed for those sets when the sandhi-undergoing word is a novel compound. On the other hand, if the loss of priming is due to the phonologically-driven derivation of a surface form, then it should be observed with both real and novel compounds.

The present study used a version of the odd-man-out implicit priming paradigm (Cholin et al., 2004). In this paradigm, each set of words shown in the experiment has either three or four items. The three-item sets are always homogeneous but the four-item sets may be homogeneous (if the additional item has the same phonological makeup as the first three) or heterogeneous (if the additional item has a different phonological makeup). We showed each set of words to the same participant twice: once with three items, and later with all four items. We hypothesized that, when the fourth item maintained the homogeneity of the set, reaction times to the four-item set would be faster than reaction times to the three-item set (participants would be more familiar with the items when seeing the set for the second time, and adding a fourth item to the set would not substantially increase the difficulty of the task when the fourth item maintained the homogeneity of the set). On the other hand, when the fourth item spoiled the homogeneity of the set, reaction times to the four-item set would be similar to or slower than reaction times to the three-item set (any benefit gained by familiarity with the items would be offset by the heterogeneous items' spoiling the homogeneity of the set and the priming that would have been associated with that homogeneity). The critical test was whether the four-item set with a sandhi-undergoing item would show facilitation comparable to that for a homogeneous four-item set, or whether it would show a lack of facilitation comparable to that for an underlyingly heterogeneous four-item set.

2. Methods

2.1. Participants

All participants were native speakers of Mandarin who grew up in Beijing. Experiment 1 had 20 participants (4 males, mean age 23.1, range 19-30), and Experiment 2 had 16 participants (3 males, mean age 22.6, range 19-26).[3] All participants provided

[3] An additional 13 participants (4 from Experiment 1 and 9 from Experiment 2) were excluded from the analysis. Four completed the task with below 75% accuracy, and the other nine were removed to ensure the same number of participants on each stimulus list in each experiment. However, the data pattern

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

Homogeneous	Heterogeneous-Derived	Heterogeneous-Lexical	Heterogeneous-Unrelated
(市场) 企业 qi^3ye^4	(市场) 企业 qi^3ye^4	(市场) 企业 qi^3ye^4	(市场) 企业 qi^3ye^4
(关机) 启动 qi^3dong^4	(关机) 启动 qi^3dong^4	(关机) 启动 qi^3dong^4	(关机) 启动 qi^3dong^4
(街头) 乞丐 qi^3gai^4	(街头) 乞丐 qi^3gai^4	(街头) 乞丐 qi^3gai^4	(街头) 乞丐 qi^3gai^4
(出发) 起身 qi^3shen^1	(开始) 起点 qi^3dian^3	(仪式) 旗手 qi^2shou^3	(制冷) 空调 $kong^1tiao^2$

Table 2. *A sample stimulus set for Experiment 2. Cue words corresponding to each target are shown in parentheses.*

Homogeneous	Heterogeneous-Derived	Heterogeneous-Lexical	Heterogeneous-Unrelated
(市场) 企业 qi^3ye^4	(市场) 企业 qi^3ye^4	(市场) 企业 qi^3ye^4	(市场) 企业 qi^3ye^4
(关机) 启动 qi^3dong^4	(关机) 启动 qi^3dong^4	(关机) 启动 qi^3dong^4	(关机) 启动 qi^3dong^4
(街头) 乞丐 qi^3gai^4	(街头) 乞丐 qi^3gai^4	(街头) 乞丐 qi^3gai^4	(街头) 乞丐 qi^3gai^4
(学术) 起究 qi^3jiu^1	(配料) 起韭 qi^3jiu^3	(配料) 旗韭 qi^2jiu^3	(奇幻) 空谜 $kong^1mi^2$

their informed consent and received payment. All methods were approved by the Human Subjects Committee of Lawrence, University of Kansas, and by the Academic Affairs Committee of Peking University.

2.2. Materials

Four critical sets of word pairs were prepared for each experiment (for a sample set, see Table 1). Each pair was made of two bisyllabic words, with the first word serving as a cue and the second as a target. The two words in the pair always had a clear semantic or associative relationship. Each set had three critical word pairs and several possible "odd-man" pairs, which differed depending on the condition (see below). The target words in these pairs met the following criteria. The first phone of each word was a stop or an affricate. The first syllable was always third tone, and the second syllable was always first, second, or fourth tone; any differences on the first syllable's tone caused by coarticulation with the second would be minimal, as Mandarin third tone is not very susceptible to anticipatory coarticulation (Xu, 1997). The initial syllables of all three critical words in a set were phonologically identical but written with different characters in mainland Chinese orthography.

Each experiment had a 2 (NUMBER OF ITEMS) × 4 (CONDITION) design. As described in section 1.3, each set of words was presented once with three pairs in the set, and again with four pairs. When the set was presented with four pairs, the fourth pair determined the condition of that set (see Table 1 for examples).[4] In the Homogeneous condition, the fourth target had the same properties as the three critical targets: 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-Derived condition (the condition of interest), the first syllable of the odd-man-out target was segmentally and tonally identical to the other items at the underlying level, but undergoes sandhi which changes its tone, such that the set of four words was homogeneous at the input (underlying) level but heterogeneous in tone at the output (surface) level. In the Heterogeneous-Lexical 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 Experiment 1, across conditions, an effort was made to make sure the "fourth items" had similar lexical frequency, as measured in the SUBTLEX-CH word form corpus (Cai & Brysbaert, 2010).

In addition, to distract participants from the third-tone manipulation, four filler sets were prepared. Unlike the critical sets, none of the filler sets was made up entirely of third-tone targets. One filler set was homogeneous; two were heterogeneous; and one was homogeneous when presented with only three-items, but the fourth item was an odd-man out that made the set segmentally heterogeneous when presented with all four items.

Experiment 2 used the same critical word pairs in each set (including the filler sets). The fourth pairs in each set, however, were *AO-AO novel compounds (see e.g. Zhang & Lai, 2010). These words consisted of two real morphemes that together do not form an existing compound word. The phonological makeup of the novel compounds followed the same design as those in Experiment 1. A sample set for Experiment 2 is shown in Table 2.

2.3. Design and procedure

For each experiment, the four critical sets were organized into four lists in a Latin square design. The presentation and timing of stimuli was controlled by the Presentation software package (http://www.neurobs.com). The experiment was divided into two blocks: in the first block, the three-item versions of each set were presented, and in the second block, the four-item versions of each set were presented in the same order. In the first block, the order of the four critical and four filler sets was randomized (with the condition that the first two sets of the block were fillers); the same random order was then used for the second

reported here was also observed when analyzing the data with the latter nine participants included.

[4] Note that the three-item sets are always homogeneous, regardless of condition, since the condition (Homogeneous, Heterogeneous-Derived, Heterogeneous-Lexical, or Unrelated) is determined by the fourth item. Nevertheless, we will still refer to these sets using the name of their associated four-item condition (e.g., "Unrelated three-item sets") when comparing them with the corresponding four-item sets.

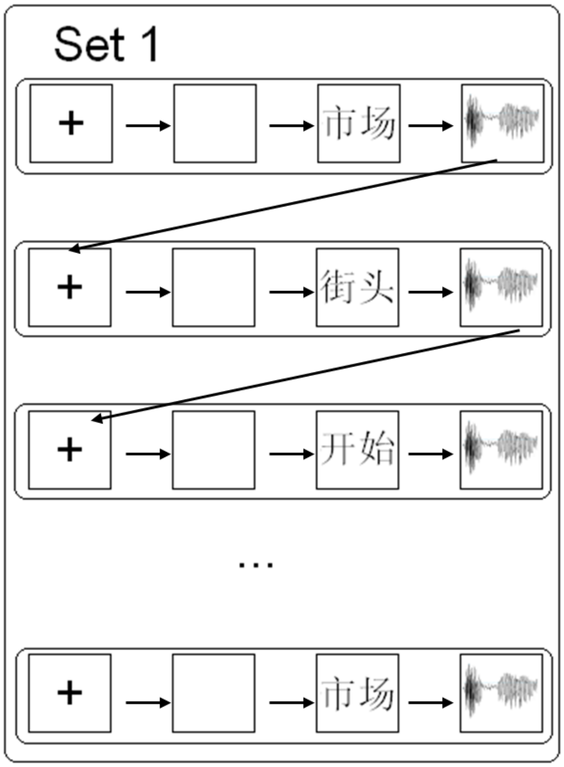

Figure 1. *Schematic trial flow for the training and test phases of a set. Each set consists of multiple trials, each of which consists of a fixation point, a blank screen, and the presentation of the cue word until the participant's vocal response, which initiates the next trial. After all trials in a given set are completed, the memorization phase for next set begins.*

block. 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. Participants were tested one at a time in a quiet room at Peking University.

For each set, participants completed a memorization phase, a training phase, and a test phase. During the memorization phase, the three or four critical pairs (for instance, the four word 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 headphones. The auditory stimuli were produced by a female native speaker from Beijing with broadcast training; the speaker 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 training phase. During the training phase, the participant responded to two repetitions of each cue (6-8 trials in total) in a random order. If the participant did not respond accurately to all cues, as determined by the experimenter, then she repeated the practice phase until she was able to respond accurately. The test phase for that set then commenced.

During the test phase (see Figure 1 for an example), the cue words for the set were presented in a random order and the participant responded by saying the associated target word as quickly as possible into a free-standing Steinberg C12 microphone. Each cue word was repeated six times in the test phase (for a total of 18 or 24 trials, depending on whether the set included three or four cue words). 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 became blank for 350, 600, 850, 1300, or 1600 ms (the duration for each trial was determined randomly at runtime), after which time one of the cue words (e.g., 市场) was presented at the center of the screen. When the participant's vocal response exceeded a pre-defined sound threshold (e.g., when she spoke $qi^{3}ye^{4}$ in response to the cue), the word disappeared from the screen. Audio recording began at the same time the cue was presented, and continued for 2500 ms. 750 ms after the offset of the participant's verbal response, the next trial began. Pseudorandomization of the trials was executed by the experimental control program at runtime. The three or four items in the set were presented once in a random order, and once they were each presented then all three or four items were again presented in a random order, and this process was repeated until all items had been repeated six times; this procedure ensured that the same item was not presented more than two times consecutively. The whole experiment took approximately forty 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 within the 2500 ms recording. Response onset latencies were measured manually using Praat (http://praat.org). Trials in which the response was incorrect were not included in the analysis of reaction times, nor trials in which the log reaction time was an outlier (determined on a subject-by-subject basis). The fourth words were not included in any analysis—adding the fourth item to the set creates heterogeneity and spoils the priming effect for the other items in the set, even if the odd-man-out item itself is not included in the measurements (Cholin et al., 2004).[5] For each experiment, log-transformed reaction times were analyzed using linear mixed effects models (Baayen et al., 2008) with random intercepts for participants and items. To account for variability in reaction time based on a given trial's order within the set or within a pseudorandomization order, a baseline model was first built with

[5] The question might be raised, then, how the study can inform our understanding of tone sandhi without directly measuring words that undergo tone sandhi. The goal of the present study, however, is not to measure the speed or effort with which tone sandhi words themselves are produced; rather, it is to investigate how such words are represented during speech production. To address that question, the present study adopted an experimental paradigm (odd-man-out implicit priming) in which the way these words are represented was expected to influence the production of *other* words, by way of maintaining or spoiling the heterogeneity of a set.

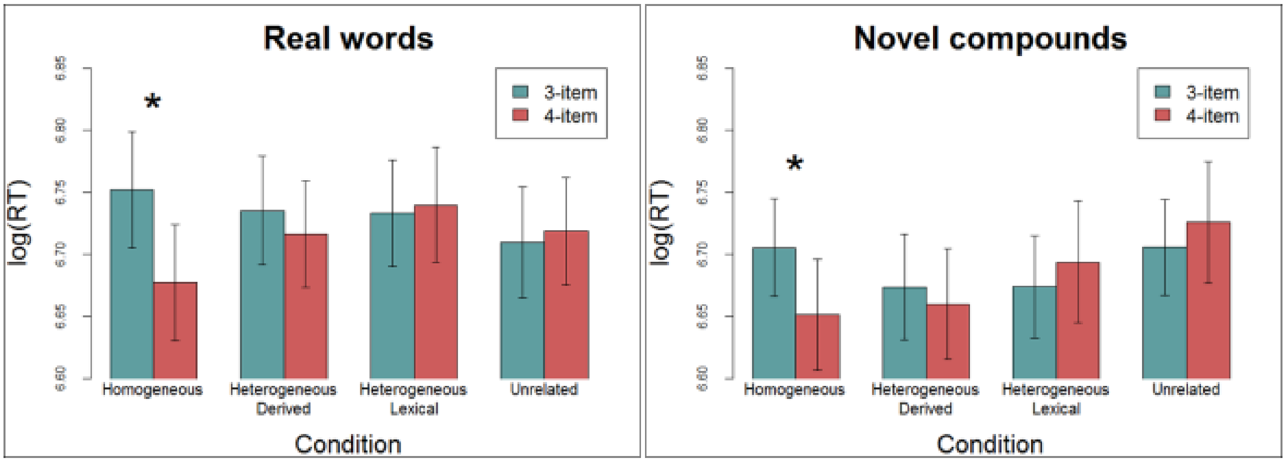

Figure 2. *Mean log reaction times for each condition in Experiment 1(left) and Experiment 2 (right). Error bars represent ±1 standard error of the by-subject mean. Asterisks denote conditions where the difference in reaction times between four-item and three-item sets was significant in the linear mixed model.*

these predictors. The experimental effects of interest (number of items in the set, and condition—Homogeneous, Heterogeneous-Derived, Heterogeneous-Lexical, or Unrelated) were compared with the baseline model via log-likelihood tests. *p*-values for the model coefficients were calculated using Markov chain Monte Carlo sampling, implemented in the `pvals.fnc()` function of the library `languageR` (Baayen et al., 2008).

3. Results

3.1. Accuracy

Accuracy was 97.0% across all conditions and participants. Accuracy was analyzed using a generalized linear mixed model (GLMM) for each experiment. In Experiment 1, the GLMM showed main effects of Number of Items and Condition. Accuracy was marginally better in 4-item sets than 3-item sets (96.7% vs. 95.6%, z = 1.66, p = .096), and marginally better in Unrelated than Homogeneous sets (96.8% vs. 94.9%, z = -1.68, p = .093). In Experiment 2, the GLMM showed only a main effect of Condition. Accuracy was highest on the Unrelated and Heterogeneous-Derived conditions (99.0% and 98.8%, respectively), whereas performance in the Homogeneous and Heterogeneous-Lexical conditions was significantly less accurate than in Unrelated (Homogeneous: 97.2%, z = -2.64, p = .008; Heterogeneous-Lexical: 97.0%, z = -2.84, p = .004).

3.2. Reaction times

Figure 2 shows the mean reaction times in each experiment. It is clear that Homogeneous four-item sets were produced substantially faster than corresponding three-item sets, whereas four-item sets in other conditions did not noticeably differ from their three-item counterparts. Furthermore, the Heterogeneous-Derived and Heterogeneous-Lexical conditions appear to have different trends: reaction times for Heterogeneous-Derived four-item sets show a trend towards being somewhat faster than their three-item counterparts, whereas reaction times for Heterogeneous-Lexical four-item sets show a trend towards being somewhat slower. These patterns are noticeable in both experiments (i.e., both with real words and novel compounds). Statistical analysis confirmed these impressions.

After removal of incorrect responses and outliers (see section 2.4), 4899 observations remained for statistical analysis. The experimental effects (reaction time differences between the four- and three-item sets for each condition) did not differ significantly between Experiments 1 and 2. There was a significant Experiment × Condition interaction ($X^2(4) = 17.17$, p = .002) due to the fact that all conditions other than Unrelated were produced more quickly in Experiment 2 (with novel compounds) than Experiment 1 (with real words). However, the Experiment × Condition × Number of Items interaction, which would indicate different experimental effects between the real and novel word experiments, did not approach significance ($X^2(4) = 0.75$, p = .945). Therefore, we report the experimental effects pooled across experiments.

Across experiments, the Condition × Number of Items interaction was significant ($X^2(3) = 25.27$, $p < .001$). The model coefficients revealed that reaction times for four-item sets were significantly faster than those to three-item sets in the Homogeneous condition (b = -0.06, t = -4.89, $p < .001$), but not in any other condition ($|b|$s < 0.018, $|t|$s < 1.49, ps > .140). In other words, facilitation was only observed for the Homogeneous condition, and was not observed for Unrelated, Heterogeneous-Lexical, or Heterogeneous-Derived. Furthermore, although neither the Heterogeneous-Derived four-item set nor the Heterogeneous-Lexical four-item set differed significantly from its corresponding three-item set, they showed opposite patterns (Heterogeneous-Derived tended towards faster responses in the four-item than the three-item set, while Heterogeneous-Lexical showed the opposite trend). Accordingly, there was a marginally significant difference between the effect of Number of Items for Heterogeneous-Derived and the effect for Heterogeneous-Lexical (b = -0.03, t = -1.84, p = .064).

4. Discussion

Building upon previous research in implicit priming and the production of Mandarin third-tone sandhi, the present study investigated whether tone sandhi is computed prior to articulation. We found that neither sets of words which were heterogeneous at only the surface level (because of tone sandhi) nor sets which were heterogeneous at the underlying and levels (because of lexical tone) showed an implicit priming effect. Furthermore, the Heterogeneous-Derived and Heterogeneous-Lexical sets were found to pattern differently: even though neither set showed significant priming, the Heterogeneous-Derived four-item set showed a trend towards facilitation relative to its corresponding three-item set, whereas the Heterogeneous-Lexical four-item set showed a trend towards inhibition. Finally, the pattern did not differ between sets in which the phonological alternation was instantiated on a real word and those in which it was instantiated on a novel compound. The significance of these results is discussed below.

Implicit priming effects are argued to stem from the planning of linguistic units before production (Afonso & Álvarez, 2011; O'Seaghdha et al., 2010; Chen et al., 2002; Levelt, 1999; but see Kawamoto, 1999). Thus, the lack of implicit priming effects for sets with a sandhi-derived odd-man-out out indicates that what participants were trying to prepare was a heterogeneous set—a set in which tone sandhi had already applied to one of the words. In other words, the word forms sent to the articulatory system were not underlying forms (which would make the set homogeneous), but surface forms (which make the set heterogeneous). Importantly, the lack of priming for these sets was replicated in Experiment 2 with novel compounds. This finding suggests that the lack of priming is not due to lexical storage of the sandhi-undergoing disyllables in their surface form. If such lexical storage were the case, we would expect a facilitation effect to appear when the sandhi words are novel compounds, which could not have been stored. Overall, the results are not consistent with a model in which speakers prepare speech based only on underlying forms and then allow alternation to happen heuristically at the level of articulation; rather, the results suggest that an additional level of phonological alternation intercedes between word-form retrieval and articulation, at least in the case of third-tone sandhi. This is in line with Chen's (1999) proposal, based on speech-error data, that tone sandhi applies before phonetic spell-out and articulation.

We note that the output of the phonological tone sandhi is not necessarily a fully-specified surface form or articulatory program. Tone sandhi alters the syllable's tonal target, but other phonetic and articulatory factors may further affect the implementation of those targets (Xu & Wang, 2001; Xu, 2001, 2004). 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.

It is of particular interest that the Heterogeneous-Derived sets did not pattern exactly like the Heterogeneous-Lexical sets in these experiments. At the surface form, both sets are approximately identical (three words with third tone and one with second tone). Therefore, if the output form of the sandhi word was fully derived before the initiation of articulation and this output form was what was supplied to the articulatory system, similar effects would be expected for both Heterogeneous-Derived and Heterogeneous-Lexical sets. The fact that Heterogeneous-Derived sets behaved differently, showing a trend towards facilitation rather than a trend towards inhibition, suggests that the underlying form still plays a role in the production of words subject to tone sandhi, even though articulatory preparation relies on the surface form rather than the underlying form. This finding is consistent with a proposal (originally put forth for Taiwanese tone sandhi) that the lexicon includes a listing of stored positional allomorphs for syllables that undergo tone sandhi (Zhang et al., 2011; Tsay & Myers, 1999). Production of the Heterogeneous-Derived sets may have been influenced by residual activation of stored third-tone allomorphs of the odd-man-out syllable, whereas there presumably are not stored third-tone allomorphs of the odd-man-out second-tone syllable in the Heterogeneous-Lexical condition (because Mandarin has no tone sandhi pattern that changes second-tone to third-tone syllables). The possibility that the results of the present experiments could be accounted for by an allomorph-listing model does not contradict our interpretation of the results as evidence for online phonological operations; indeed, allomorph selection in this context is a phonological operation, since the choice of allomorph is constrained by the phonological grammar and thus requires considering the phonological context. Further research, however, is needed to elucidate the role that may be played by activated but non-pronounced allomorphs in implicit priming tasks.

5. Conclusions

The present study examined 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 (both in real words and in novel compounds) behaves like underlying second tone in blocking the implicit priming effect, suggesting that tone sandhi is computed before articulation. 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. Furthermore, the results suggest that the underlying form of a syllable that has undergone sandhi still has a subtle influence on the speech production process; this result suggests that storage and selection of positional allomorphs may play a role in the online deployment of phonological knowledge during production. 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 (Zhang & Lai, 2010; Chen, 2000); these topics could shed new light on the relationship between phonetic and phonological encoding during speech production.

6. Acknowledgements

The authors would like to thank the members of the University of Kansas Phonetics & Psycholinguistics Laboratory (KUPPL) and participants in the 2012 Symposium on Tonal Aspects of Languages for feedback on this research. We would also like to thank Xiaolin Zhou for providing facilities at the Center for Brain and Cognitive Science (Peking University) for data collection.

7. References

Afonso, O., & Álvarez, C. (2011). Phonological effects in handwriting production: evidence from the implicit priming paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1474-1483.

Baayen, H., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE, 5*, e10729.

Chen, J. (1999). The representation and processing of tone in Mandarin Chinese: evidence from slips of the tongue. *Applied Psycholinguistics, 20*, 289-301.

Chen, J., Chen, T., & Dell, G. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language, 46*, 751-781.

Chen, M. (2000). *Tone sandhi: patterns across Chinese dialects*. Cambridge: Cambridge University Press.

Cholin, J., Schiller, N., and Levelt, W. (2004). The preparation of syllables in speech production. *Journal of Memory and Language, 50*, 47-61.

Cutler, A., & Chen, H. (1995). Phonological similarity effects in Cantonese word recognition, *Proceedings of the 13th International Congress of the Phonetic sciences,* 106-109.

Kawamoto, A. (1999). Incremental encoding and incremental articulation in speech production: evidence based on response latency and initial segment duration. *Behavioral and Brain Sciences, 22*, 48-49.

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: "Experimental Studies in Word and Sentence Prosody", Mouton de Gruyter, 212-237.

Ladefoged, P. (1980). What are linguistic sounds made of? *Language, 56*, 485-502.

Lee, C. (2007). Does horse activate mother? Processing lexical tone in form priming. *Language and Speech, 50*, 101-123.

Levelt, W. (1999). Models of word production. *Trends in Cognitive Sciences, 3* 223-232.

Levelt, W., Roelofs, A., & Meyer, A. (1999) A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1-75.

O'Seaghdha, P., Chen, J., & Chen, T. (2010). Proximate units in word production: phonological encoding begins with syllables in Mandarin but with segments in English. *Cognition, 115*, 282-302.

Peng, S. (1996). *Phonetic implementation and perception of place coarticulation and tone sandhi,* Ph.D. dissertation, The Ohio State University.

Politzer-Ahles, S., & Zhang, J. (2012). The role of tone sandhi in speech production: evidence for phonological parsing. Talk presented at *3rd Conference on Tonal Aspects of Languages*.

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

Tsay, J., & Myers, J. (1999). Taiwanese tone sandhi as allomorph selection. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society, 22*, 394-405.

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

Xu, Y. (1991). Depth of phonological recoding in short-term memory. *Memory and Cognition, 19*, 263-273.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics, 25*, 61-83.

Xu, Y. (2001). Sources of tonal variations in connected speech. *Journal of Chinese Linguistics, 17*, 1-31.

Xu, Y. (2004). Understanding tone from the perspective of production and perception. *Language and Linguistics, 5*, 757-797.

Xu, Y., & Wang, Q. (2001). Pitch targets and their realization: evidence from Mandarin Chinese. *Speech Communication, 33*, 319-337.

Zee, E. (1980). A spectrographic investigation of Mandarin tone sandhi. *UCLA Working Papers in Linguistics,* 49, 98-116.

Zhang, J. (2010). Issues in the analysis of Chinese tone. *Language and Linguistics Compass, 4*, 1137-1153.

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

Zhang, J., Lai, Y., & Sailor, C. (2011). Modeling Taiwanese speakers' knowledge of tone sandhi in reduplication. *Lingua, 121*, 181-206.

Zhang, Q. (2008). Phonological encoding in monosyllabic and bisyllabic Mandarin word production: implicit priming paradigm study. *Acta Psychologica Sinica, 40*, 253-262.

Zhou, X., & Marslen-Wilson, W. (1997). The abstractness of phonological representation in the Chinese mental lexicon. Ed. Chen, H., *Cognitive Processing of Chinese and Related Asian Languages*, 3-27.. The Chinese University Press.