EVIDENCE FOR THE ROLE OF TONE SANDHI IN MANDARIN SPEECH PRODUCTION

Stephen Politzer-Ahles[1], Jie Zhang[2]

[1]*New York University Abu Dhabi, United Arab Emirates;* [2]*University of Kansas, United States*

ABSTRACT

The present study investigates the representation of sandhi-undergoing words 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 homogeneous at the underlying level but 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 surface and underlying levels (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. Other recent independent studies on this phenomenon are also discussed.

SUBJECT KEYWORDS

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. 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 what speakers know about their language (competence), however, rather than a model of what they do during language use (performance). 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 et al.'s (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 speak directly to a separate cognitive level for the computation of phonological alternation.

There are clearly, however, many kinds of alternations that cannot be handled by known articulatory mechanisms alone: alternations that are not coarticulatory or phonetically natural, 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 surface representations play a role during online speech production before the initiation of overt articulation.

To test whether morpho-phonological alternation is computed before articulatory preparation, we adopt 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 (Meyer, 1990, 1991). 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 initial 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 framework of Levelt and colleagues (1999), this facilitation occurs during the stage of phonological encoding (the association of segments into slots in a prosodic frame, called *prosodification*) and before a word is phonetically “spelled out” and encoded into articulatory commands. 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. Words that undergo third-tone sandhi are presumed to be articulated with different forms than their lexical representations, and implicit priming allows us to test when the alternation occurs (i.e., before or after phonological encoding), if at all, during speech production. This can be done by testing sets of words that are homogeneous at the underlying level but heterogeneous at the surface level, and vice versa.

Below we briefly describe Mandarin tone and its role in processing (including relevance to implicit priming), the phonological and psychological characteristics of third-tone sandhi, and previous studies addressing this research question using similar methods. 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 TONES

Mandarin has four phonologically distinctive tones: 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 et al. (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 segments 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 is phonologically encoded; tonal heterogeneity in a set reduces implicit priming.

1.2. THIRD 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]]). Mandarin third-tone sandhi is phonological in nature (i.e., it does not have strong phonetic motivation) and is applied without exception (Zhang & Lai, 2010; Kuo et al., 2007; Peng, 2000). 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, 2000; Zee, 1980). The neutralization is even less complete in novel words than existing words (Zhang & Lai, 2010). 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 (Peng, 1996; Wang & Li, 1967), but recent evidence from visual world eye-tracking—which uses an implicit behavioral measure rather than measuring respondent's overt choices—suggests that the perceptual difference does influence lexical access (Speer & Xu, 2008).

Of greater 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 (2000) 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.

A few previous studies (Chen & Chen, this volume; Chen, 2012; Chen, Shen, & Schiller, 2011) have investigated the production of third-tone sandhi in production using implicit priming, with similar designs as that of the present study (see below). Chen, Shen, & Schiller (2011) found that sets of words containing a mixture of sandhi-undergoing T3 (surfacing as T2) and non-sandhi-undergoing T3 (surfacing as T3) showed implicit priming effects, just as homogeneous sets of T3 words do, whereas sets containing both sandhi-undergoing T3 (surfacing as sandhi-derived T2) and underlying T2 did not show as much priming as homogeneous sets; they argued that both the T2 and T3 allophones of a given underlying T3 syllable may be stored in the mental lexicon. Chen and Chen (this volume) found priming in both types of sets, and argued (based on other experiments in their study) that it was due to the phonetic similarity of the initial portions of T2 and T3 (both of which begin with low, dipping F0 contours). They suggest that the articulatory system may have been able partially prepare the articulatory gestures for the initial part of the tonal contour. On the other hand, Chen (2012) found no significant priming in sets which shared underlying tone but not surface tone. Importantly, these studies used almost all T2- or T3-initial words in the relevant experiments, which may have unduly drawn attention to the tone sandhi and introduced response biases. That is to say, participants who were aware that the experiment was about T2 and T3, or even that the experiment was about third-tone sandhi in particular, may not have produced the words as naturally as naive participants would. Furthermore, previous studies have only tested implicit priming in real words, which makes it difficult to disentangle effects of the online computation of tone sandhi from frequency- or experience-based effects (e.g., storage of surface forms); novel compounds, which cannot be directly retrieved from the mental lexicon, are a valuable comparison case which has not yet been tested.

1.3. THE PRESENT STUDY

The present study examines the role tone sandhi plays during implicit priming by building upon Chen et al.'s (2002) finding that sets of words in which the underlying tones differ do not show a substantial implicit priming effect. Specifically, we examine whether participants preparing to articulate a sandhi-undergoing compound are engaged with the underlying representation or the surface representation of the sandhi-undergoing syllable. We used a version of the *odd-man-out* implicit priming paradigm comparable to that used by Afonso & Álvarez (2011), where heterogeneous sets did not include all four different tones as in Chen et al. (2002), but included three items with the same initial tone and one with a same or different tone. In this paradigm, the three items with the same tones should be produced faster when the fourth item also has the same tone (maintaining the homogeneity of the set), and slower when it has a different tone (spoiling the homogeneity of the set). Thus, even without testing reaction times to the sandhi-undergoing words, the method allows us to probe the representation of these words by observing how the inclusion of sandhi-undergoing words affects the reaction times to other words in the set. Importantly, unlike previous implicit priming studies of tone sandhi, the present study included fillers with various tone combinations to distract participants from the sandhi manipulation.

The aim of the present study is to test how the words undergoing phonological alternations are represented during speech production, and how input forms are mapped to output forms. If it is their underlying forms that are the representations that are encoded for articulation (i.e., if tone sandhi does not occur until after phonological encoding), sets of words that are homogeneous at the underlying level only should show facilitation just like fully homogeneous sets; if it is their surface forms that are encoded for articulation, then the facilitation for such sets should be reduced. Note that the mapping of input-to-output forms does not necessarily mean "alternation" in the generative phonology sense—it could involve the selection of an appropriate allomorph for the first syllable of the compound, which is still a phonological operation (it relies on phonological knowledge) and could cause the effects predicted above (c.f. Chen et al., 2011).

To this end, we conducted two experiments. 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 (see section 2.2 for a detailed description of the stimuli). The disruption of the 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. If the sandhi-undergoing T3 items spoil the implicit priming for non-sandhi-undergoing T3 items, that could be evidence that it is the surface forms of such words (rather than the underlying forms) that are encoded for articulation, and thus that the phonological alternation had been computed during or before phonological encoding; another possibility, however, would be that compound words are simply listed in the lexicon according to their post-sandhi forms (Zhou & Marslen-Wilson, 1997). 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, we conducted a second experiment in which the words undergoing tone sandhi were novel compounds. If the loss of priming for sets including a sandhi-undergoing word is due to lexical storage of the surface forms in the first experiment, 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.

Finally, the second experiment also tests the role of tone sandhi in a context that is the converse of that described above: sets of words that are heterogeneous at the underlying level but homogeneous at the surface level (i.e., with three critical T2 items, and a fourth item that is underlyingly T3 but undergoes sandhi to become T2). Again, testing whether these sets behave like heterogeneous or homogeneous sets allows us to infer whether articulatory encoding was driven by the underlying or the surface representations of the sandhi-undergoing words.

2. EXPERIMENT ONE

2.1. METHODS

2.1.1. 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, University of Kansas.

2.1.2. MATERIALS

Five critical sets of word pairs were prepared for the experiment. 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-out" 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 a stop or an affricate. The first syllable was always third-tone, and the second syllable was always a different

Table 1. *A sample stimulus set. Cue words corresponding to each target are shown in parentheses. The first three rows show the critical items, and the last row shows the odd-man-out items. The Three-Item set (which includes just the three critical items and no odd-man item) is not shown.*

Unrelated	**Homogeneous**	**Heterogeneous**	**Sandhi**
(市场) 企业 *qi*^{3}*ye*^{4}	(市场) 企业 *qi*^{3}*ye*^{4}	(市场) 企业 *qi*^{3}*ye*^{4}	(市场) 企业 *qi*^{3}*ye*^{4}
(关机) 启动 *qi*^{3}*dong*^{4}	(关机) 启动 *qi*^{3}*dong*^{4}	(关机) 启动 *qi*^{3}*dong*^{4}	(关机) 启动 *qi*^{3}*dong*^{4}
(街头) 乞丐 *qi*^{3}*gai*^{4}	(街头) 乞丐 *qi*^{3}*gai*^{4}	(街头) 乞丐 *qi*^{3}*gai*^{4}	(街头) 乞丐 *qi*^{3}*gai*^{4}
(制冷) 空调 *kong*^{1}*tiao*^{2}	(出发) 起身 *qi*^{3}*shen*^{1}	(爱国) 旗帜 *qi*^{2}*zhi*^{4}	(开始) 起点 *qi*^{3}*dian*^{3}

tone (but never the light tone, *qing sheng*, which appears on unstressed 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 (Xu, 1997). The first syllables of all three critical words were phonologically identical but written with different characters. The entire set of stimuli is shown in Appendix A.

Sets in all conditions other than *Three-Item* were presented with a fourth pair of words, which was an "odd-man-out" item if it differed phonologically from the other items in such a way that it spoiled the homogeneity of the set. In the *Unrelated* condition, the first syllable of the odd-man target shared neither segments nor tone with the other three targets. 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, and thus we predicted that the critical items in this condition would be produced faster than those in the Unrelated condition (a facilitation effect). In the *Heterogeneous* condition, the first syllable of the odd-man-out target was segmentally identical to the other items but had lexical second tone, making the set heterogeneous at both the input and output levels; we predicted that the facilitation for this condition, compared to the Unrelated condition, would be smaller than that for the Homogeneous condition compared to the Unrelated condition. In the *Sandhi* condition (the condition of interest), the first syllable of the odd-man target 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 this condition, if articulatory preparation uses the phonological output forms the facilitation effect should be comparable to that for the Heterogeneous condition, whereas if articulatory preparation uses the input forms the effect should be comparable to that for the Homogeneous condition. 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 (Cai & Brysbaert, 2010). For a sample set of target words, see Table 1.

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 word pairs), two were heterogeneous in terms of both segmentals and tones (with three word pairs), and one was homogeneous except for a heterogeneous unrelated "odd-man-out" (with four pairs).

2.1.3. 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 memorization phase and a test phase. During the memorization phase, the three or four critical pairs 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, 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 (see Figure 1 for an illustration of the procedure). Within each set, each cue word was repeated four times, yielding 12-16 trials (yielding a total of 60 data points per participant—odd-man out trials were not used as data points); 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 was presented at the center of the screen. The recording began at the moment the cue word appeared, and continued for 2 seconds.

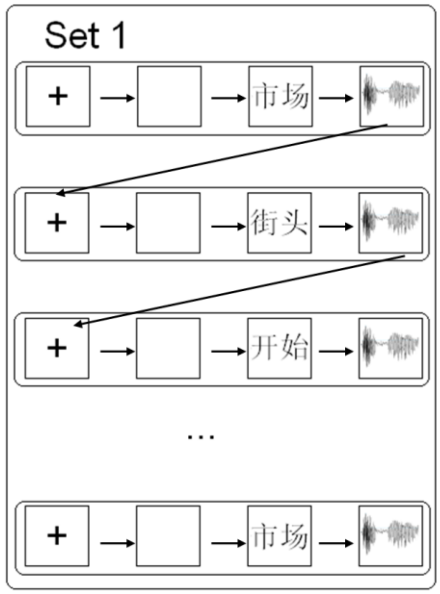


Figure 1. *Schematic trial flow for the test phase of a set. Each set consists of multiple trials, each of which consists of a fixation point, a jittered blank screen, and presentation of the cue word until the participant's vocal response, which initiates the next trial. After all trials in a given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

When the participant's vocal response exceeded a pre-defined sound threshold, 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.

2.1.4. 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 (Cholin et al., 2004). Reaction times were log-transformed to approximate a normal distribution (the log transformation resulted in the skewness values closest to zero), and observations with reaction times deviating from that subject's mean by more than three standard deviations were excluded from the analysis. Statistical analysis was conducted in the R statistical computing environment using linear mixed effects models for reaction times, and generalized linear mixed effects models for accuracy. The mixed model included effects of Set Number (representing how far into the overall experiment a given observation's set occurred), Trial In Set (representing how many trials into its own set a given observation occurred), and Repeated (representing whether or not a given trial was the same item as the previous trial)[2] as nuisance covariates (which were not of interest for the hypotheses under question but were included to account for variance not related to the manipulation of interest; the pattern of results reported below still holds if these are not included), as well as an effect of CONDITION and crossed random intercepts for subjects, items, and lists (models with more complex random effects structures did not converge). The baseline condition was Unrelated. Model evaluation (to test main effects and interactions) was performed using log-likelihood tests, and where effects were significant then 95% confidence intervals for individual coefficients in the best model were calculated via bootstrapping with 1000 simulations in the `bootMer` `{lme4}` package; coefficients were considered significant at $p < .05$ iff the 95% confidence interval did not include zero.

2.2. RESULTS

2.2.1. ACCURACY

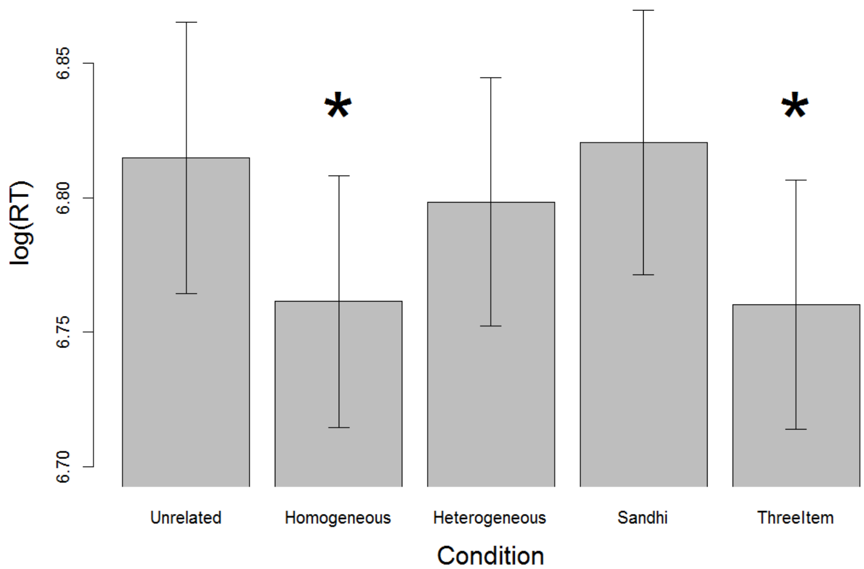
Participants responded correctly on 85.5% of trials in the Unrelated condition, 88.6% in Homogeneous, 91.1% in Heterogeneous, 91.9% in Sandhi, and 89.1% in Three-Item. There was not a significant difference in accuracy between conditions ($\chi^2(4) = 6.85$, $p = .144$).

2.2.2. REACTION TIMES

Table 2. *Model summary for Experiment 1. While statistics were calculated on log-transformed data, the second through fourth columns report model coefficients and 95% confidence intervals transformed back to milliseconds for ease of interpretation. "Intercept" corresponds to reaction times for the Unrelated set, and coefficients in bold correspond to priming effects relative to that set. Asterisks indicate coefficients that are significant according to the confidence intervals. The final column indicates, for ease of exposition, the standard deviation in milliseconds of the RTs for the corresponding condition (which was not part of the regression model).*

Coefficient	*b* (log)	*b* (ms)	Lower bound	Upper bound	*t*	SD (ms)
(Intercept)	6.888	980.41	919.53	1045.32	209.44*	263.77
Set Number	0.002	1.75	-1.98	5.59	0.91	n/a
Number in Set	-0.006	-5.76	-8.10	-3.49	-4.84*	n/a
Repeated	-0.111	-103.37	-127.32	-79.65	-8.12*	n/a
Homogeneous	-0.058	-55.29	-86.68	-24.09	-3.39*	231.42
Heterogeneous	-0.024	-23.27	-55.64	9.57	-1.40	240.72
Sandhi	-0.008	-8.09	-41.65	25.81	-0.49	259.18
Three-Item	-0.073	-69.35	-100.43	-38.88	-4.20*	230.96

Figure 2. *Mean log reaction times in Experiment 1. Error bars represent ±2 standard errors. Asterisks denote conditions that showed significant priming (i.e., conditions that were significantly faster than the corresponding Unrelated condition).*



After the removal of outliers and incorrect responses (see section 2.1.4), 1604 observations remained for analysis. The effect of CONDITION was significant ($\chi^2(4) = 27.50$, $p < .001$). Table 2 shows the model coefficients; the last four rows of the second column represent the priming effects, in milliseconds, for each condition. The model indicated that critical words in Homogeneous and Three-Item sets, but not those in Heterogeneous and Sandhi sets, were produced significantly faster than critical words in Unrelated sets. For ease of comparison with Experiment 2, mean reaction times are shown in Figure 2.

To directly compare the reaction times for the Sandhi set to those for the Homogeneous and Heterogeneous set, we re-computed the model using Sandhi as the baseline. In this analysis, reaction times for Sandhi sets were significantly slower than those for Homogeneous (b = 0.049, CI = 13.52 – 77.72, t = 2.79) and Three-Item (b = 0.076, CI = 38.32 – 97.87, t = 4.32) sets.

Finally, because effects in Chinese odd-man-out priming have been argued to be weaker than those in the traditional priming paradigm (Chen & Chen, 2006), we performed a second analysis in which only trials following odd-man out trials were used. Such trials may be considered to incur a switch cost when the preceding odd-man out is phonologically different than the trial in question (as is the case for Unrelated and Heterogeneous trials) but not when it is the same (as in Homogeneous trials). The Three-Item condition was not included (since it has no odd-man-out items), nor was the Repeated covariate (since the items were, by definition, not repetitions of the previous trial). The analysis included 324 observations and again showed a main effect of Condition ($\chi^2(3) = 12.94$, $p = .005$); the results are summarized in Appendix C. Once again, Homogeneous sets showed priming and Sandhi sets did not. In this analysis, the priming effect for Heterogeneous sets also reached significance. Sandhi significantly differed from Homogeneous ($b = 71.32$, CI = 14.96 – 125.28, $t = 2.40$), but not from Unrelated ($b = -29.44$, CI = -96.95 – 33.12, $t = -0.90$) or Heterogeneous ($b = 46.61$, CI = -13.59 – 102.75, $t = 1.52$).

2.3. DISCUSSION

Experiment 1 found that a sandhi-undergoing T3 item behaved like an underlyingly T2 item for the purposes of articulatory preparation: that is to say, sandhi-undergoing T3 items spoiled the implicit priming effect for non-sandhi-undergoing T3 items in the same set, just as underlyingly T2 items did. This result is contrary to the findings of Chen and Chen (this volume) and Chen et al. (2011), in whose experiments sandhi-undergoing T3 did not spoil priming for surface T3 items. Methodological differences between these experiments are discussed in the General Discussion.

A surprising result from the present study is the failure for priming in segmentally homogeneous but tonally heterogeneous conditions (e.g., Heterogeneous and Sandhi) to reach significance in the overall analysis. In the priming studies by Chen et al. (2002) and Zhang (2008), the priming for such sets was small (12 ms and 16 ms, respectively), but significant; in the present study, Heterogeneous elicited a larger numerical effect (24 ms), 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 increased the variance in the data (compared to the traditional implicit priming paradigm) and made it more difficult to identify weak effects. Nevertheless, the crucial effect—the lack of priming for Sandhi sets—was very far from significance and thus cannot be attributed to any lack of power. Another possible explanation for the lack of priming in the Heterogeneous 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 (Afonso & Álvarez, 2011; Roelofs, 2006), Chen et al. have argued that in Chinese, an unrelated odd-man-out target does not completely spoil the priming (Chen & Chen, 2006). In other words, differences between Heterogeneous and Unrelated sets in traditional implicit priming are likely to be more robust than in odd-man-out implicit priming. Thus, it is possible that implicit priming failed to appear in the tonally heterogeneous sets here because they were being compared to a baseline that also exhibited some priming (albeit less priming than that in the Homogeneous set). Indeed, the facilitation effect for this condition was significant when we analyzed only trials immediately following odd-man out items—an analysis that should highlight the difference between Unrelated and other conditions. The lack of priming for Heterogeneous sets in our initial analysis does not influence the main finding of the experiment, which is that Sandhi sets behaved differently from Homogeneous sets.

Another surprising finding was that the Sandhi condition was numerically slower than the Heterogeneous condition. While this difference was not significant, it is a trend in the opposite direction of what would be expected if the underlying form facilitated production. This trend cannot be directly due to any processing costs in realizing tone sandhi, since the reaction time measurements are based only on the items that do *not* instantiate sandhi (sandhi only occurs on the odd-man out item, which is not analyzed). Regardless, the predictions for the study regarded whether the sandhi set would behave more like a Homogeneous set or more like a Heterogeneous set, and thus the results are closer to the latter prediction. Because there was no *a priori* prediction regarding a slowdown unique to Sandhi sets, and because this trend is not replicated in the real-word condition of Experiment 2 (see below), we do not discuss it further here.

As described in the Introduction, it is impossible to tell on the basis of Experiment 1 whether the effect of tone sandhi arose during phonological encoding or in the lexicon. That is to say, it is possible that tone sandhi was not an operation that happened on-line during speech preparation, but rather that the compound words used in the experiment had simply been stored in the lexicon as T2-T3 words already. Thus, Experiment 2 was conducted to investigate whether the pattern of results observed above would also appear when the odd-man items in the experiment were novel words and thus not stored in the mental lexicon. Furthermore, since the results of Experiment 1 suggest that implicit priming was driven by the surface form rather than the underlying form of the sandhi-undergoing targets (because sets which were homogeneous at the underlying level and heterogeneous at the surface level behaved as if they were heterogeneous), it is informative to test the converse situation: sets which are heterogeneous at the underlying level but homogeneous at the surface level thanks to tone sandhi.

3. EXPERIMENT TWO

Experiment 2 had three goals: to test the converse of the manipulation tested in the previous experiments (sets with T2 critical items and a sandhi odd-man out, such that the set would appear heterogeneous at the underlying level and homogeneous at the surface level); to test whether the effect observed in Experiment 1 extends to novel-word stimuli; and to replicate Experiment 1. Therefore, Experiment 2 included sets of words in which the critical targets all began with T2. We also manipulated the lexicality of the odd-man-out items between participants: for half of the participants, the odd-man-out targets were novel compounds (see below). Thus the experiment had a 2 (SET TONE: T3, T2) × 4 (CONDITION: Unrelated, Homogeneous, Heterogenous, Sandhi) × 2 (LEXICALITY: real words, novel words) mixed design. In both T3 and T2 sets, Homogeneous sets are expected to show priming compared to Unrelated sets, and Heterogeneous sets are not (or the priming is expected to be smaller than that for Homogeneous sets). If the surface forms of sandhi-undergoing compounds are computed during or before phonological encoding, then Sandhi sets would be predicted to show priming relative to Unrelated sets in T2 sets, but not in T3 sets.

One methodological change was made: because the rate of errors and/or timed-out responses (14.5% to 8.1%) and the mean reaction time in Experiment 1 were both higher than the error rates and mean reaction times observed in many other implicit priming experiments, the amount of practice participants were given on each set and the number of repetitions for each item were increased.

3.1. METHODS

3.1.1. PARTICIPANTS

All participants were native speakers of Mandarin who grew up in Beijing. The real-word condition had 25 participants (9 males, mean age 22.8, range 18-31), and the novel-word condition had 22 participants (9 males, mean age 21.4, range 18-29). All participants provided their informed consent and received payment. All methods were approved by the Academic Affairs Committee of Peking University.

3.1.2. MATERIALS

Experiment 2 used similar materials as Experiment 1. The Heterogeneous fourth items in T3 sets were modified such that they were always T2-T3 compounds (whereas in Experiment 1, the second syllable of these items was not always T3); this ensured that these items were phonetically comparable to Sandhi items, and differed only in the fact that they had T3 underlying forms. In addition, four T2 sets were created following the same constraints at the T3 sets, except that in the T2 sets the critical words always began with a T2 syllable rather than a T3 (see Table 3). Thus, *Homogeneous* odd-man-out targets were T2-T3 compounds (whereas in the T3 sets they were T3-Tx compounds, where Tx was any tone other than T3), and *Heterogeneous* odd-man-out targets were compounds with T3 on the first syllable and a non-T3 second syllable. Four additional filler sets were created according to the same criteria as in the previous experiments, yielding eight critical and eight filler sets. Unlike in Experiment 1, all sets in this experiment had four items.

Table 3. *A sample stimulus set. Cue words corresponding to each target are shown in parentheses. In each cell, the first three rows show the critical items, and the last row shows the odd-man-out items. The upper row of cells shows the T3 sets, and the lower row shows the T2 sets. The novel-word counterparts of these sets are not shown here.*

	Unrelated	**Homogeneous**	**Heterogeneous**	**Sandhi**
T3	(市场) 企业 *qi*^{3}*ye*^{4} (关机) 启动 *qi*^{3}*dong*^{4} (街头) 乞丐 *qi*^{3}*gai*^{4} (制冷) 空调 *kong*^{1}*tiao*^{2}	(市场) 企业 *qi*^{3}*ye*^{4} (关机) 启动 *qi*^{3}*dong*^{4} (街头) 乞丐 *qi*^{3}*gai*^{4} (出发) 起身 *qi*^{3}*shen*^{1}	(市场) 企业 *qi*^{3}*ye*^{4} (关机) 启动 *qi*^{3}*dong*^{4} (街头) 乞丐 *qi*^{3}*gai*^{4} (仪式) 旗手 *qi*^{2}*shou*^{3}	(市场) 企业 *qi*^{3}*ye*^{4} (关机) 启动 *qi*^{3}*dong*^{4} (街头) 乞丐 *qi*^{3}*gai*^{4} (开始) 起点 *qi*^{3}*dian*^{3}
T2	(培训) 徒弟 *tu*^{2}*di*^{4} (设计) 图案 *tu*^{2}*an*^{4} (军队) 屠杀 *tu*^{2}*sha*^{1} (抽象) 具体 *ju*^{4}*ti*^{3}	(培训) 徒弟 *tu*^{2}*di*^{4} (设计) 图案 *tu*^{2}*an*^{4} (军队) 屠杀 *tu*^{2}*sha*^{1} (水泥) 涂抹 *tu*^{2}*mo*^{3}	(培训) 徒弟 *tu*^{2}*di*^{4} (设计) 图案 *tu*^{2}*an*^{4} (军队) 屠杀 *tu*^{2}*sha*^{1} (红薯) 土豆 *tu*^{3}*dou*^{4}	(培训) 徒弟 *tu*^{2}*di*^{4} (设计) 图案 *tu*^{2}*an*^{4} (军队) 屠杀 *tu*^{2}*sha*^{1} (农业) 土壤 *tu*^{3}*rang*^{3}

Novel-compound counterparts of each fourth item in each set (including fillers) were constructed. The items were novel compounds consisting of two real morphemes that together do not form existing compound words. The phonological makeup of the novel compounds followed the same design as the real words. Note that the *critical* words in the experiment were always real words. The full set of stimuli is shown in Appendix B.

3.1.3. DESIGN AND PROCEDURE

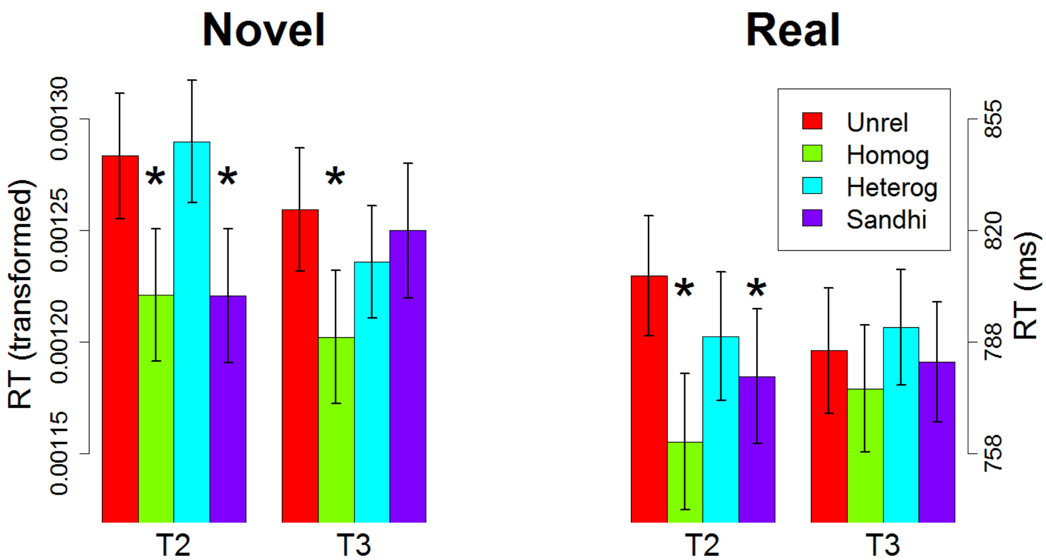
Experiment 2 used roughly the same procedure as Experiment 1. In this experiment, however, between the memorization phase and test phase for each set (see section 2.1.3) there was also a training phase. Once a participant finished memorizing the items for a given set, she responded to two repetitions of each cue (eight 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. This procedure was implemented to ensure better accuracy in the aggregate data, to avoid removing participants because of incorrectly-memorized sets, and to give participants time at the beginning of each set to acclimate to the words and remove the effect of ORDERINSET that was observed in Experiment 1 (exploratory analyses suggested that the effect of ORDERINSET was curvilinear, happening mainly in the beginning of each set). After the participant completed the training phase accurately, the test phase for that set then commenced.

The other differences in procedure between this experiment and Experiment 1 were that, in Experiment 2, acoustic recording continued for 2500 ms after the presentation of the cue, and that the items in a set were randomized differently. Randomization of items within a test phase as accomplished by collecting two repetitions of each of the four items and randomly presenting these eight trials, then repeating that process two more times; this ensured that, within the first eight trials of a block, every target (including the odd-man out) would be produced twice. The whole experiment took approximately forty minutes.

3.1.4. DATA ANALYSIS

Data preprocessing was conducted in the same way as in Experiment 1. Statistics were conducted on reaction times transformed to reflected reciprocals, which yielded the most normal distribution (raw and log-transformed data had skewness values above 1). The mixed model included the same nuisance covariates as in Experiment 1, and then the predictors of interest (LEXICALITY, SET TONE, and CONDITION) and their interactions. In the interaction terms, CONDITION was nested under SET TONE (because direct comparisons across T2 and T3 sets were not of interest) and these were nested under LEXICALITY. In all other respects the statistical analysis was the same as in Experiment 1.

Figure 3. *Mean reaction times in Experiment 2. Error bars represent ±2 standard errors. Asterisks denote conditions that showed significant priming (i.e., conditions that were significantly faster than the corresponding Unrelated condition). Reaction times were transformed to reflected reciprocals; the y-axis on the right of the figure gives the equivalents in raw milliseconds.*



3.2. RESULTS

3.2.1. ACCURACY

Accuracy was 96.3% across all conditions and participants, and no participant had below 86% accuracy. The general linear mixed model showed a significant LEXICALITY × SET TONE × CONDITION three-way interaction ($\chi2(3) = 9.89$, $p = .020$). In real-word T3 sets, accuracy was significantly or marginally higher for Sandhi (97.1%) and Heterogeneous (96.7%) than for Unrelated (94%) sets ($z = 1.97$, $p = .049$; $z = 1.78$, $p = .073$), whereas in novel-word T3 sets, accuracy was significantly or marginally *lower* for Sandhi (94.7%) and Heterogeneous (97.2%) than for Unrelated (98.7%) sets ($z = -3.17$, $p = .002$; $z = -1.93$, $p = .054$).

3.2.2. REACTION TIMES

Figure 3 shows the mean reaction times for each group. Averaging across LEXICALITY the numerical pattern of the data corresponds to our prediction: specifically, Sandhi sets show facilitation in T2 contexts but not T3 contexts. However, the pattern differed between real-word and novel-word sets, as indicated by a significant three-way interaction ($\chi 2(3) = 10.34$, $p = .016$). Table 4 shows the model coefficients in the maximal regression model.

For T3 sets, significant facilitation was only observed in the Homogeneous condition with a novel-word odd-man-out. Unlike in the previous experiments, facilitation for the Homogeneous condition with real words did not reach significance, although this may be due to insufficient power, as the numerical trend is towards facilitation. Importantly, neither the Heterogeneous nor the Sandhi sets differed from Unrelated. Sandhi sets were significantly slower than Homogeneous sets in the novel-word condition ($b = 37.06$, CI = 17.15 – 55.83, $t = 3.51$), although this was not the case in the real-word condition ($b = 2.19$, CI = -18.25 – 20.31, $t = 0.21$)—unlike in Experiment 1, where this difference was significant with real words. In short, the results for novel words closely mirror the results observed for real words in Experiment 1, although the results for real words in the present experiment surprisingly did not reach significance.

For T2 sets, the expected facilitation effect for Homogeneous sets, compared to Unrelated sets, was significant for both real and novel words, whereas Heterogeneous sets did not show significant facilitation. Most importantly, Sandhi sets showed significant facilitation for both real and novel words, consistent with our prediction. Comparing the Sandhi sets directly against the Homogeneous and Heterogeneous sets, however, yields some interesting patterns. With novel words, Sandhi sets were faster than Heterogeneous sets ($b = -36.05$, CI = -59.92 – -14.43, $t = -3.19$) but did not differ from Homogeneous sets ($b = 6.00$, CI = -15.78 – 25.82, $t = 0.55$). On the other hand, with real words, even though Sandhi sets were significantly faster than Unrelated sets they were not as fast as Homogeneous sets ($b = 28.31$, CI = 8.72 – 46.65, $t = 2.86$) and did not significantly differ from Heterogeneous sets ($b = -5.82$, CI = -27.28 – 13.32, $t = -0.57$). In short, in novel-word T2 sets Sandhi words behaved like Homogeneous words, whereas in real-word T2 sets they showed priming but behaved differently than Homogeneous words.

Table 4. *Model summary for Experiment 2. While statistics were calculated on reflected reciprocal data (rr), the second through fourth columns report model coefficients and 95% confidence intervals transformed back to milliseconds for ease of interpretation. Coefficients in bold represent the coefficients of interest, which correspond to priming effects (comparisons against Unrelated sets). Asterisks indicate coefficients that are significant according to the confidence intervals. The final column indicates, for ease of exposition, the standard deviation in milliseconds of the RTs for the corresponding condition (which was not part of the regression model).*

Coefficient	*b* (rr)	*b* (ms)	Lower bound	Upper bound	*t*	SD (ms)
(Intercept)	1.30e-03	855.09	792.85	924.92	28.00*	246.97
Set Number	1.48e-06	1.08	0.11	2.10	2.06*	n/a
Repeated	-1.36e-04	-89.20	-97.33	-80.33	-17.78*	n/a
Number in Set	-9.33e-07	-0.68	-1.24	-0.14	-2.33*	n/a
Real	-5.23e-05	-36.59	-94.21	32.78	-1.08	245.88
Novel, T3	-2.46e-05	-17.60	-52.77	21.27	-0.91	241.13
Real, T3	-2.75e-05	-19.68	-55.06	18.97	-1.03	236.82
Novel, T2, Homogeneous	-5.93e-05	-41.28	-61.18	-20.24	-3.65*	231.68
Real, T2, Homogeneous	-7.72e-05	-52.94	-72.44	-33.20	-5.10*	246.87
Novel, T3, Homogeneous	-6.12e-05	-42.50	-62.36	-20.69	-3.79*	220.99
Real, T3, Homogeneous	-2.17e-05	-15.55	-36.15	6.79	-1.42	210.85
Novel, T2, Heterogeneous	1.04e-06	0.76	-20.88	23.75	0.06	263.33
Real, T2, Heterogeneous	-2.49e-05	-17.86	-37.19	2.60	-1.66	250.51
Novel, T3, Heterogeneous	-2.45e-05	-17.53	-38.69	5.45	-1.54	212.41
Real, T3, Heterogeneous	-4.52e-06	-3.29	-24.39	19.27	-0.30	235.07
Novel, T2, Sandhi	-5.03e-05	-35.29	-55.49	-14.87	-3.11*	253.42
Real, T2, Sandhi	-3.36e-05	-23.85	-44.37	-3.47	-2.19*	253.15
Novel, T3, Sandhi	-3.41e-06	-2.48	-24.17	21.16	-0.21	273.06
Real, T3, Sandhi	-1.84e-05	-13.24	-34.06	7.77	-1.20	230.26

As in Experiment 1, we re-analyzed the data using only trials that immediately followed odd-man out items. The three-way interaction between LEXICALITY, SET TONE, and CONDITION was marginal ($\chi^2(3) = 6.96$, $p = .073$), and the model is summarized in Appendix D. Unlike in the previous analysis, here the priming effect for Homogeneous sets was significant in all conditions. In T3 sets, neither Heterogeneous nor Sandhi sets showed priming. In T2 sets the effects differed between real- and novel-word sets. In real-word T2 sets, both Heterogeneous and

Sandhi sets showed significant priming and did not significantly differ from one another (b = 0.36, CI = -33.64 – 39.62, t = 0.02), whereas Sandhi sets were slower than Homogeneous (b = 43.94, CI = 10.54 – 75.23, t = 2.54). In novel-word T2 sets, on the other hand, Sandhi sets were significantly faster than Heterogeneous (b = -53.81, CI = -97.93 – -11.84, t = -2.55) and did not differ from Homogeneous (b = 15.51, CI = -20.62 – 50.30, t = 0.80). In short, this analysis provided further evidence that Sandhi sets in the T2 conditions behaved like Homogeneous sets in the novel-word context but like Heterogeneous sets in the real-word context; and both the real-word and novel-word T3 contexts replicated the main finding of Experiment 1, which was that post-sandhi heterogeneity spoiled priming.

3.3. DISCUSSION

Because Experiment 1 suggested that the surface forms of sandhi words were computed during or before phonological encoding, and were the forms that drive implicit priming, the present experiment tested this same hypothesis using T2 sets, where sandhi would be expected to cause facilitation rather than spoil facilitation. This prediction was borne out in the novel-word sets: Sandhi sets yielded facilitation (indistinguishable from that in Homogeneous sets) in T2 environments, whereas they did not do so in T3 environments in the previous experiments. This suggests that priming was being driven by the surface form (recall that, in T2 sets, sandhi causes the odd-man-out target to have the same surface form as the other items). With real words, on the other hand, the facilitation in Sandhi sets, while significant, was smaller in size than that for Homogeneous sets; in fact, reaction times for Sandhi sets were in between those for Homogeneous and Heterogeneous sets, and statistically were more comparable to Heterogeneous. (While Sandhi sets showed significant priming and Heterogeneous sets did not, the two conditions nevertheless failed to differ from one another because the priming effects were both near the border of significance.) The fact that the priming effect, while significant, was smaller than the priming effect for Homogeneous sets suggests that the underlying forms played a role during speech production in the real-word condition—recall that in previous studies tonally heterogeneous sets still showed small priming effects, but these effects were smaller than effects for homogeneous sets (Chen et al., 2002; Zhang, 2008). Thus, the underlying form of sandhi-undergoing items seems to have reduced the priming effect. It is worth noting that the priming effect for real-word Sandhi sets, in both T2 and T3 contexts, seem to be numerically in between that for Homogeneous and Heterogeneous sets (see Figure 3), which would be consistent with another implicit priming study on tone sandhi that used a slightly different paradigm (Politzer-Ahles & Zhang, 2012), although the current study did not find statistical support for a difference between Sandhi and Heterogeneous in either Experiment 1 or 2. It remains to be seen whether such a difference will prove more robust in future studies; if priming effects for Sandhi sets reliably fall in between those for Homogeneous and Heterogeneous sets, that could constitute evidence for the activation of both surface and underlying forms during production.

The direction of the interaction with LEXICALITY in the present experiment is opposite what was predicted—recall that the motivation for including novel words which was to see if real words were stored in terms of their pre-compiled *surface* forms (c.f. Zhou & Marslen-Wilson, 1997), since those are the forms that are frequently articulated and heard, whereas novel words might be show more activation of the underlying form since the surface form of the compound is not lexicalized. Why, then, would there be activation of the underlying form in real words but not novel words? This activation of the underlying form may be activation that has spread from *semantic* activation of the real word, which in turn could activate the underlying T3 representation of the first morpheme in the compound; with novel words, on the other hand, participants may have just been repeating meaningless articulatory programs (which, even if they were not lexicalized, would remain fresh in their memory from having just practiced the words in a given set) without semantically activating the underlying morpheme that corresponds to the surface form. These interpretations are tentative, however, since they are not the pattern that was originally predicted. It should also be noted that a similar experiment which also manipulated lexicality, in a somewhat different implicit priming paradigm, did not find differences between real- and novel-word sets (Politzer-Ahles & Zhang, 2012).

The other goal of Experiment 2 was to include a direct replication of Experiment 1. The results are equivocal on this point, because the real-word T3 conditions in Experiment 2 (the only set of conditions that entirely replicates the conditions included in Experiment 1) did not show any significant effects in the main analysis, even for Homogeneous sets (which are expected to show facilitation under any theory). Without a significant priming effect for Homogeneous sets, the lack of priming for Sandhi sets in this condition is not interpretable in isolation. Nevertheless, the novel-word conditions in the present experiment showed the patterns predicted on the basis of Experiment 1 (and the novel-word T3 condition showed the same pattern as what was observed in Experiment 1). Furthermore, the analysis of trials immediately following odd-man-out items, which are the items most likely to show the relevant effects, replicated the main finding of Experiment 1. Thus, the results provide evidence that, at least under some conditions, post-sandhi heterogeneity may spoil implicit priming effects—unlike what Chen et al. (2011) and Chen (this volume) found—and that implicit priming may in some cases be driven by the surface form of a sandhi-undergoing compound.

Finally, accuracy rates for Experiment 2 were higher than for Experiment 1 and were on par with other implicit priming experiments, suggesting that the lower accuracy in Experiment 1 was due to low-level methodological factors (number of repetitions, etc.) rather than the odd-man-out paradigm itself. On the other hand, while overall reaction times across all conditions were faster than in Experiment 1, they remained slower than the average reaction times in many other implicit priming experiments (such as Chen & Chen, this volume). This is probably due to the higher number of repetitions per item (and thus greater familiarity with the items) in traditional implicit priming studies compared to the present study; differences in lexical properties (such as frequency) of the stimuli used, and differences in the strength of the associative relationship between cues and targets, may also have contributed to the slower reaction times.

5. GENERAL 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 during or before phonological encoding (i.e., the fitting of abstract segmental and tonal representations into a prosodic frame). We found that sets of words which were heterogeneous at the surface level but not the underlying level (because of tone sandhi) behaved similarly to sets of words which were unambiguously heterogeneous (because of lexically specified tone): both types of sets failed to yield an implicit priming effect. On the converse, the behavior of sets of words that were *homogeneous* at the surface level but not the underlying level was moderated by lexicality: when the sandhi-undergoing word was a novel compound, the set behaved like a homogeneous set, but when the sandhi-undergoing word was a real compound the priming effect was somewhat spoiled. These results suggest that the surface form is often what matters when participants are preparing speech, but that there may be situations where the underlying form exerts an influence as well.

Implicit priming effects are argued to stem from the planning of linguistic units before production, specifically during the association of phonological segments or tones into slots in a prosodic frame (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 sandhi-derived heterogeneity 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 be determined by articulatory mechanisms at a later stage of production (as suggested by, e.g., Chen et al., 2011, who propose that phonological encoding uses the abstract underlying forms of sandhi compounds and that the surface forms are derived after phonological encoding). 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 generally not underlying forms, but surface forms.[3] This is in line with Chen's (1999) proposal, based on speech-error data, that tone sandhi applies before phonetic spellout and articulation. These results should not come as a surprise, given that there are many theoretical reasons to assume that not all alternation can result from known articulatory mechanisms (see the Introduction).

The real-word T2 data, however, are difficult to reconcile with such an account. While the priming observed in the novel-word sandhi condition is consistent with this account (after sandhi is realized during phonological encoding, the output forms fed to the articulatory system would be homogeneous and thus yield priming), there is no explanation for why priming was reduced in the real-word sandhi condition if articulatory encoding is based only on the surface forms. This pattern suggests that, even if sandhi is realized early, the underlying form still plays a role at some point in production. It is not clear, though, why the underlying form did now exert a similar influence in the the novel-word T2 sets, or in Experiment 1. It would be valuable to test the issue of storage and computation in other ways beyond novel words, such as by investigating the production of sandhi outside of compound words (e.g., sandhi occurring across word boundaries).

The results of the present study may also speak to the question of storage vs. composition of words that undergo phonological alternation. As mentioned in the Introduction, it is not known whether sandhi-undergoing compounds are actually stored as underlying T3-T3 strings that must later be translated into T2-T3 articulatory commands, or if they are simply stored in their surface forms (see, e.g., Zhou & Marslen-Wilson, 1997, for discussion). While the present experiment was not specifically designed to test this question, the results are less compatible with the latter account than the former. In experiment 2, sets with novel sandhi-undergoing compounds behaved even more like their surface forms (e.g., novel words that made their sets underlyingly homogeneous but surface heterogeneous behaved just like novel words that made their sets heterogeneous at both the underlying and surface levels) than those with real sandhi-undergoing compounds. If the effects observed were due just to the representation of stored underlying forms, the opposite pattern would be expected. Furthermore, in Experiment 2 the real sandhi-undergoing compounds in both T3 and T2 sets showed effects that were numerically in between those for heterogeneous and homogeneous sets; this is consistent with the T3-T3 forms of the words exerting some influence during the course of speech production, even though those forms are not actually uttered.

5.1. OTHER IMPLICIT PRIMING STUDIES ON TONE SANDHI

At this point it is worthwhile to make an in-depth comparison of the results of the present study to those of other recent studies on this phenomenon. Chen and Chen's experiments (this volume) observed significant and comparable priming effects for conditions corresponding to our T2-Sandhi, T3-Sandhi, T2-Heterogeneous, and T3-Heterogeneous sets.[4] They argued that this occurred because Mandarin T3 and T2 are phonetically similar and therefore the initial portion of the tone contour could have been prepared early, regardless of when sandhi was realized. These results are inconsistent with those of the present study; there are, however, numerous methodological differences between the studies. Whereas the present study used odd-man-out implicit priming, Chen and Chen's (this volume) study used the traditional implicit priming paradigm (Meyer, 1990, 1991); therefore, the studies used different types of Unrelated baselines. Furthermore, the present experiment included fillers to distract participants from the goal of the study, whereas Chen and Chen's (this volume) did not. Chen and Chen's (this volume) experiment included several sets in which all the words undergo third-tone sandhi (half of the Unrelated sets in Experiment 2), whereas the present study did not. In the present study, during the learning/memorization phase for each set, participants were presented with both written characters and auditory tokens in order to avoid biasing them towards either the underlying or surface forms; participants in Chen and Chen’s (this volume) experiment, on the other hand, were only presented with written characters. Furthermore, Chen and Chen (this volume) tested participants from Taiwan, whereas the present study tested participants from mainland China; it is possible that these different groups of speakers have different representations of tones.[4] Finally, Chen and Chen’s (this volume) experiment did not test Homogeneous conditions, making it impossible to make some of the crucial comparisons made in the present study (note that, because tonal heterogeneity does not always completely spoil priming, it is not sufficient to just test whether sandhi-induced heterogeneity makes reaction times faster compared to Unrelated sets; it is also necessary to compare reaction times to Homogeneous sets); in fact, some of our results look similar to Chen and Chen’s if the Homogeneous condition is ignored (e.g., in the real-word T2 conditions of Experiment 1, Sandhi sets show priming and do not differ significantly from Heterogeneous sets; it is only by comparing these to Homogeneous sets that we notice the priming has been reduced).[5]

Chen and Chen's (this volume) finding does not necessarily preclude our account of the results of the present study. According to their account, it is possible that the output forms of sandhi-undergoing words are computed prior to articulatory preparation, and thus that T3-Sandhi sets in their study were comparable to lexically Heterogeneous sets, but that the phonetic similarity of T2 and T3 reduced the effect of the heterogeneity. Under this interpretation, the sandhi was computed implicitly prior to articulatory preparation in both studies, and the differences between their results and ours results may be due to differences in the extent to which the experimental designs made participants aware of these phonetic similarities. Thus, their account may still be consistent with our conclusion about how the sandhi was realized. Further study will be necessary to test this account and determine what aspects of experimental design may causes these differences in the effect of phonetic similarity.

Our study is much more similar in design to another recent study independently conducted by Chen, Shen, and Schiller (2011), which also used mainland Chinese participants. They found comparable priming effects for both T3-Homogeneous and T3-Sandhi sets, relative to Unrelated sets (which they refer to as Heterogeneous sets in their report), and they found a reduction of priming for T2-Sandhi sets compared to T2-homogeneous sets. The former result is inconsistent with ours, in which the priming for T3-Sandhi sets was comparable to T3-Heterogeneous sets rather than T3-Homogeneous sets; the latter result is consistent with our finding for real words but not with our finding for novel words. It is not yet clear what factors may have contributed to the difference between their results and ours. While their study used a design that was similar to ours, their study is again different than ours in that ours included fillers and the previous studies did not. Therefore, we reiterate the possibility that differences in the salience of the T3 manipulation between our study and the previous study may have caused participants to adopt different strategies and/or to differ in their awareness of small phonetic similarities or differences between T3 and T2. Again, since this speculation relies on several untested assumptions (that the salience of the manipulation in the experimental context influences implicit priming effects, and that the phonetic similarity between T2 and T3 can be made more or less salient for speakers by the experimental context), further study will be necessary to better understand the differences between these experiments. A potential challenge for this account is that an pair of experiments by Chen (2012) found similar results as ours for T3-Sandhi and T3-Heterogeneous sets, but also did not include fillers. Overall, the difference between the results of the present study and those of previous studies suggests that the basis of implicit priming effects in tone sandhi is more nuanced than previously assumed, and may be affected by lexical factors and by contextual factors such as participants' attentional biases.

Further study will be needed to develop a more complete account of how these factors influence the derivation of tone sandhi during speech production.

6. CONCLUSION

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—may be precompiled at some point before articulatory encoding. This suggests that models of speech production may need to 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. However, differences between the present study's results and results of similar recent studies, as well as differences between real words and novel words in the present study, illustrate the necessity for replications and further studies to test this account and to better understand the ways in which experimental design can influence implicit priming effects observed in this type of study, and the ways in which the underlying forms of phonologically altered words may exert an influence on speech production.

NOTES

1. In the notation used here, superscript 1 corresponds to a High tone (55 in Chao numbers), 2 to a Rising tone (35), 3 to a Low tone (213), and 4 to a Falling tone (51).
2. While any item could be preceded by anywhere from 0 to 3 repetitions of the same item, repetitions of 2 or 3 were so rare that coding this variable as an ordinal predictor did not improve the model and yielded dummy-coded predictors that were too highly correlated to give meaningful coefficients. Therefore, we simply treated items preceded by 0 repetitions of the same item as "not repeated", and items preceded by 1, 2, or 3 repetitions as "repeated".
3. An anonymous reviewer suggested that, if tone sandhi were applied online rather than being stored lexically, Sandhi sets would show slower reaction times than Heterogeneous sets, since participants completing a Sandhi set must decide for every word whether or not to apply sandhi, whereas they do not need to make this decision

in Heterogeneous sets (since such sets do not include any sandhi-undergoing words). A numerical trend in this direction was observed in Experiment 1, but it did not reach statistical significance and was not replicated in Experiment 2. We note that the prediction of slower reaction times for Sandhi sets compared to Heterogeneous sets relies on an assumption—which has not been empirically tested—that applying sandhi is an operation that is different than just selecting the appropriate phonological form of a word (and thus in the case of Sandhi sets participants must choose whether to do this operation, but in the case of Heterogeneous sets there is no choice to be made). On the other hand, if the basic operation of tone sandhi is allomorph selection, then it may be no more costly than normal phonological encoding—that is to say, in both Sandhi and Heterogeneous sets, regardless of whether the tone to be articulated is a T2 or T3 allomorph, participants must nevertheless wait until the visual prompt and then decide which tone to produce. Under such an understanding of how tone sandhi could be applied, it is possible that sandhi application could be an online operation but also not cause Sandhi sets to have a longer reaction time than Heterogeneous sets. Thus, on the basis of the present data it is not possible to adjudicate between these accounts.

4. T3 is assumed to have different underlying forms in the varieties of Mandarin spoken in mainland China and in Taiwan (Zhang & Lai, 2010). In mainland China the underlying form is /213/ (the tone has a rise when it appears in phrase-final position), whereas in Taiwan it is /21/ (the tone has no rise even in final position). Thus, one might predict the similarity of T3 and T2 to be even greater in mainland than in Taiwanese Mandarin. As noted above, however, design differences between our study and Chen and Chen's (this volume) make it difficult to compare the Heterogeneous effect sizes directly. Furthermore, as the experiments involve T3 in non-final position, no sets involve a comparison between T2 and final-rise T3 ([213]).

5. We note, however, that the priming effect sizes observed by Chen and Chen (this volume) were numerically similar to the effect sizes elicited by Homogeneous conditions in their previous research (e.g., Chen et al., 2002) using the same paradigm with different participants.

ACKNOWLEDGEMENTS

The experiments reported here were previously presented at the Third International Symposium on Tonal Aspects of Languages and at the 164th Meeting of the Acoustical Society of America.

REFERENCES

AFONSO, Olivia, and Carlos ÁLVAREZ. 2011. "Phonological effects in handwriting production: evidence from the implicit priming paradigm." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37: 1474-1483.

CAI, Qing, and Marc BRYSBAERT. 2010. "SUBTLEX-CH: Chinese word and character frequencies based on film subtitles." *PLoS ONE* 5: e10729.

CHEN, Jenn-Yeu. 1999. "The representation and processing of tone in Mandarin Chinese: evidence from slips of the tongue." *Applied Psycholinguistics* 20: 289-301.

_____, and Train-Min CHEN. This volume. " Mechanism and locus of tone 3 sandhi during Mandarin Chinese spoken Word production." *Journal of Chinese Linguistics*.

_____, ______, and Gary Dell. 2002. "Word-form encoding in Mandarin Chinese as assessed by the implicit priming task." *Journal of Memory and Language* 46: 751-781.

CHEN, Train-Min, and Jenn-Yeu CHEN. 2006. "Morphological encoding in the production of compound words in Mandarin Chinese." *Journal of Memory and Language* 54: 491-514.

CHEN, Yiya, Rachel SHEN, and Niels SCHILLER. 2011. "Representation of allophonic tone sandhi variants." *Proceedings of Psycholinguistics Representation of Tone* (satellite workshop to *International Congress of Phonetic Sciences*).

CHEN, Xiaocong. 2012. " On the role of the underlying tonal representation in the phonological encoding stage of Tone sandhi production of the Mandarin Third tone". M.A. thesis, Guangdong University of Foreign Studies.

CHOLIN, Joana, Niels SCHILLER, and Willem LEVELT. 2004. "The preparation of syllables in speech production." *Journal of Memory and Language* 50: 47-61.

KAWAMOTO, Alan. 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, Yu-ching, Yi XU, and Moira YIP. 2007. "The phonetics and phonology of apparent cases of iterative tone change in Standard Chinese. In *Phonology and Phonetics: Tones and Tunes, Volume 2*, 212-237. Berlin: Mouton de Gruyter.

LADEFOGED, Peter. 1980. "What are linguistic sounds made of?" *Language* 56: 485-502.

LEVELT, Willem, Ardi ROELOFS, and Antje MEYER. 1999. "A theory of lexical access in speech production." *Behavioral and Brain Sciences* 22: 1-75.

MEYER, Antje. 1990. "The time course of phonological encoding in language production: the encoding of successive syllables of a word." *Journal of Memory and Language* 29: 534-545.

______. 1991. "The time course of phonological encoding in language production: phonological encoding inside a syllable". *Journal of Memory and Language* 30: 69-89.

O'SEAGHDHA, Padraig, Jenn-Yeu CHEN, and Train-Min CHEN. 2010. "Proximate units in word production: phonological encoding begins with syllables in Mandarin but with segments in English." *Cognition* 115: 282-302.

PENG, Shu-hui. 2000. "Lexical versus 'phonological' representations of Mandarin sandhi tones." In *Language acquisition and the lexicon: Papers in laboratory phonology V*, edited by Michael Broe and Janet Pierrehumbert, 152-167. Cambridge, UK: Cambridge University Press.

ROELOFS, Ardi. 2006. "The influence of spelling on phonological encoding in word reading, object naming, and word generation." *Psychonomic Bulletin & Review* 13: 33-37.

SPEER, Shari, and Lei XU. 2008. "Processing lexical tone in third-tone sandhi." Talk presented at *Laboratory Phonology* 11.

WANG, William, and Kung-Pu LI. 1967. "Tone 3 in Pekingese." *Journal of Speech and Hearing Research* 10: 629-636.

XU, Yi. 1991. "Depth of phonological recoding in short-term memory." *Memory and Cognition* 19: 263-273.

______. 1997. "Contextual tonal variations in Mandarin." *Journal of Phonetics* 25: 61-83.

ZEE, Eric. 1980. "A spectrographic investigation of Mandarin tone sandhi." *UCLA Working Papers in Linguistics* 49: 98-116.

ZHANG, Jie. 2010. "Issues in the analysis of Chinese tone." *Language and Linguistics Compass* 4: 1137-1153.

______, and Yuwen LAI. 2010. "Testing the role of phonetic

ZHANG, Qingfang. 2008. "Phonological encoding in monosyllabic and bisyllabic Mandarin word production: implicit priming paradigm study." *Acta Psychologica Sinica* 40: 253-262. [in Chinese]

ZHOU, Xiaolin, and William MARSLEN-WILSON. 1997. "The abstractness of phonological representation in the Chinese mental lexicon." In *Cognitive Processing of Chinese and Related Asian Languages*, edited by Hsuan-Chih Chen, 3-27. Hong Kong: The Chinese University Press.

APPENDIX A

The stimuli used for Experiment 1 are shown below. In each table, the first three lines show the critical items and the rest show the odd-man items, which varied depending on condition (all participants saw the same critical items, but the odd-man items they saw differed). The numbers below the condition names give the mean log frequency for the odd-man out words in that condition.

Critical sets

	bi^3	qi^3	ju^3	zhi^3	zhu^3
Critical	(不齿) 鄙视 bi^3shi^4 (冠军) 比赛 bi^3sai^4 (河堤) 彼岸 bi^3an^4	(市场) 企业 qi^3ye^4 (关机) 启动 qi^3dong^4 (街头) 乞丐 qi^3gai^4	(难过) 沮丧 ju^3sang^4 (口腔) 咀嚼 ju^3jue^2 (编程) 矩阵 ju^3zhen^4	(现金) 纸币 zhi^3bi^4 (闲人) 止步 zhi^3bu^4 (皇上) 旨意 zhi^3yi^4	(盯着) 瞩目 zhu^3mu^4 (烹调) 煮饭 zhu^3fan^4 (关心) 嘱咐 zhu^3fu^4
Unrelated (1.825)	(几何) 图形 tu^2xing^2	(制冷) 空调 $kong^1tiao^2$	(艺术) 画家 hua^4jia^1	(保暖) 毛皮 mao^2pi^2	(咳嗽) 病症 $bing^4zheng^4$
Homogeneous (1.895)	(冰心) 笔名 bi^3ming^2	(出发) 起身 qi^3shen^1	(匿名) 举报 ju^3bao^4	(到达) 指标 zhi^3biao^1	(臆测) 主观 zhu^3guan^1
Heterogeneous (2.151)	(眼镜) 鼻梁 bi^2liang^2	(爱国) 旗帜 qi^2zhi^4	(情况) 局势 ju^2shi^4	(航班) 直达 zhi^2da^2	(慢慢) 逐渐 zhu^2jian^4
Sandhi (1.955)	(西装) 笔挺 bi^3ting^3	(开始) 起点 qi^3dian^3	(行为) 举止 ju^3zhi^3	(跟随) 指引 zhi^3yin^3	(地位) 主导 zhu^3dao^3

Filler sets

2

Unrelated 1: (张骞) 西域 *xi*^{1}*yu*^{4}, (负荆) 请罪 *qing*^{3}*zui*^{4}, (生意) 客户 *ke*^{4}*hu*^{4}

Unrelated 2: (学问) 知识 *zhi*^{1}*shi*, (劈柴) 斧头 *fu*^{3}*tou*^{2}, (抗议) 罢工 *ba*^{4}*gong*^{1}

Odd-man: (沙发) 客厅 *ke*^{4}*ting*^{1}, (教材) 课本 *ke*^{4}*ben*^{3}, (面对) 客服 *ke*^{4}*fu*^{2}, (刺探) 间谍 *jian*^{4}*die*^{2}

Homogeneous 1: (座谈) 议论 *yi*^{4}*lun*^{4}, (奇怪) 异常 *yi*^{4}*chang*^{2}, (看法) 意见 *yi*^{4}*jian*^{4}

Homogeneous 2: (喇嘛) 西藏 *xi*^{1}*zang*^{4}, (水果) 西瓜 *xi*^{1}*gua*^{1}, (面试) 西装 *xi*^{1}*zhuang*^{1}

APPENDIX B

The stimuli used for Experiment 2 are shown below. In each table, the first three lines show the critical items and the rest show the odd-man items, which varied depending on condition (all participants saw the same critical items, but the odd-man items they saw differed). For each condition, the upper row shows the real-word odd-man-out and the lower row shows the novel-word odd-man-out. The numbers below the condition names give the mean log frequency for the real odd-man out words in that condition. For the fillers, the real- and novel-word items for a given set are given in braces {}.

T3 sets

	bi^{3}	qi^{3}	zhi^{3}	zhu^{3}
Critical	(不齿) 鄙视 $bi^{3}shi^{4}$ (冠军) 比赛 $bi^{3}sai^{4}$ (河堤) 彼岸 $bi^{3}an^{4}$	(市场) 企业 $qi^{3}ye^{4}$ (关机) 启动 $qi^{3}dong^{4}$ (街头) 乞丐 $qi^{3}gai^{4}$	(现金) 纸币 $zhi^{3}bi^{4}$ (闲人) 止步 $zhi^{3}bu^{4}$ (皇上) 旨意 $zhi^{3}yi^{4}$	(盯着) 瞩目 $zhu^{3}mu^{4}$ (烹调) 煮饭 $zhu^{3}fan^{4}$ (关心) 嘱咐 $zhu^{3}fu^{4}$
Unrelated (2.417)	(艺术) 画家 $hua^{4}jia^{1}$ (眼泪) 画哭 $hua^{4}ku^{1}$	(制冷) 空调 $kong^{1}tiao^{2}$ (奇幻) 空谜 $kong^{1}mi^{2}$	(咳嗽) 病症 $bing^{4}zheng^{4}$ (甘甜) 病蜜 $bing^{4}mi^{4}$	(面试) 西装 $xi^{1}zhuang^{1}$ (杂志) 西刊 $xi^{1}kan^{1}$
Homogeneous (1.783)	(冰心) 笔名 $bi^{3}ming^{2}$ (寂寞) 笔孤 $bi^{3}gu^{1}$	(出发) 起身 $qi^{3}shen^{1}$ (学术) 起究 $qi^{3}jiu^{1}$	(到达) 指标 $zhi^{3}biao^{1}$ (甘甜) 指蜜 $zhi^{3}mi^{4}$	(臆测) 主观 $zhu^{3}guan^{1}$ (精彩) 主妙 $zhu^{3}miao^{4}$
Heterogeneous (1.378)	(流血) 鼻孔 $bi^{2}kong^{3}$ (历史) 鼻古 $bi^{2}gu^{3}$	(仪式) 旗手 $qi^{2}shou^{3}$ (配料) 旗韭 $qi^{2}jiu^{3}$	(公正) 执法 $zhi^{2}fa^{3}$ (白饭) 执米 $zhi^{2}mi^{3}$	(野炊) 竹筒 $zhu^{2}tong^{3}$ (时刻) 竹秒 $zhu^{2}miao^{3}$
Sandhi (2.044)	(西装) 笔挺 $bi^{3}ting^{3}$ (历史) 笔古 $bi^{3}gu^{3}$	(开始) 起点 $qi^{3}dian^{3}$ (配料) 起韭 $qi^{3}jiu^{3}$	(跟随) 指引 $zhi^{3}yin^{3}$ (白饭) 指米 $zhi^{3}mi^{3}$	(地位) 主导 $zhu^{3}dao^{3}$ (时刻) 主秒 $zhu^{3}miao^{3}$

T2 sets

	$chang^{2}$	di^{2}	$tang^{2}$	tu^{2}
Critical	(肚子) 肠胃 $chang^{2}wei^{4}$ (普遍) 常用 $chang^{2}yong^{4}$ (味道) 尝试 $chang^{2}shi^{4}$	(音乐) 笛声 $di^{2}sheng^{1}$ (传人) 嫡系 $di^{2}xi^{4}$ (实在) 的确 $di^{2}que^{4}$	(容器) 搪瓷 $tang^{2}ci^{2}$ (昆虫) 螳螂 $tang^{2}lang^{2}$ (亲戚) 堂哥 $tang^{2}ge^{1}$	(培训) 徒弟 $tu^{2}di^{4}$ (设计) 图案 $tu^{2}an^{4}$ (军队) 屠杀 $tu^{2}sha^{1}$
Unrelated (2.517)	(首饰) 珠宝 $zhu^{1}bao^{3}$ (酱油) 珠醋 $zhu^{1}cu^{4}$	(优酷) 视频 $shi^{4}pin^{2}$ (爱着) 视靠 $shi^{4}kao^{4}$	(蟾蜍) 青蛙 $qing^{1}wa^{1}$ (完整) 青满 $qing^{1}man^{3}$	(抽象) 具体 $ju^{4}ti^{3}$ (面粉) 具撒 $ju^{4}sa^{3}$

Homogeneous (1.716)	(奥运) 长跑 *chang*^{2}*pao*^{3} (久远) 长古 *chang*^{2}*gu*^{3}	(竞争) 敌手 *di*^{2}*shou*^{3} (住宅) 敌府 *di*^{2}*fu*^{3}	(甘甜) 糖果 *tang*^{2}*guo*^{3} (潮湿) 糖雨 *tang*^{2}*yu*^{3}	(水泥) 涂抹 *tu*^{2}*mo*^{3} (方向) 涂北 *tu*^{2}*bei*^{3}
Heterogeneous (2.410)	(盛大) 场面 *chang*^{3}*mian*^{4} (用人) 场雇 *chang*^{3}*gu*^{4}	(建筑) 底层 *di*^{3}*ceng*^{2} (钱财) 底富 *di*^{3}*fu*^{4}	(休息) 躺卧 *tang*^{3}*wo*^{4} (雕琢) 躺玉 *tang*^{3}*yu*^{4}	(红薯) 土豆 *tu*^{3}*dou*^{4} (书包) 土背 *tu*^{3}*bei*^{4}
Sandhi (2.094)	(画面) 场景 *chang*^{3}*jing*^{3} (久远) 场古 *chang*^{3}*gu*^{3}	(面对) 抵挡 *di*^{3}*dang*^{3} (住宅) 底府 *di*^{3}*fu*^{3}	(家具) 躺椅 *tang*^{3}*yi*^{3} (潮湿) 躺雨 *tang*^{3}*yu*^{3}	(农业) 土壤 *tu*^{3}*rang*^{3} (方向) 土北 *tu*^{3}*bei*^{3}

Filler sets:

Unrelated 1: (事业) 贫穷 *pin*^{2}*qiong*^{2}, (保暖) 毛衣 *mao*^{2}*yi*^{1}, (电脑) 科技 *ke*^{1}*ji*^{4},
{ (傍晚) 夕阳 *xi*^{1}*yang*^{2} / (金子) 夕淘 *xi*^{1}*tao*^{3} }

Unrelated 2: (食品) 养分 *yang*^{3}*fen*^{4}, (旅游) 宾馆 *bin*^{1}*guan*^{3}, (传统) 习俗 *xi*^{2}*su*^{2},
{ (风扇) 凉快 *liang*^{2}*kuai*^{4} / (床单) 凉铺 *liang*^{2}*pu*^{4} }

Unrelated 3: (学问) 知识 *zhi*^{1}*shi*, (橙子) 橘子 *ju*^{2}*zi*, (砍柴) 斧头 *fu*^{3}*tou*^{2},
{ (抗议) 罢工 *ba*^{4}*gong*^{1} / (裤子) 罢条 *ba*^{4}*tiao*^{2} }

Unrelated 4: (张骞) 西域 *xi*^{1}*yu*^{4}, (洗澡) 肥皂 *fei*^{2}*zao*^{4}, (负荆) 请罪 *qing*^{3}*zui*^{4},
{ (生意) 客户 *ke*^{4}*hu*^{4} / (大量) 客批 *ke*^{4}*pi*^{1} }

Odd-man 1: (简单) 复杂 *fu*^{4}*za*^{2}, (老爸) 父亲 *fu*^{4}*qin*^{1}, (不好) 负面 *fu*^{4}*mian*^{4},
{ (情况) 局势 *ju*^{2}*shi*^{4} / (温和) 局淡 *ju*^{2}*dan*^{4} }

Odd-man 2: (沙发) 客厅 *ke*^{4}*ting*^{1}, (困难) 客服 *ke*^{4}*fu*^{2}, (教材) 课本 *ke*^{4}*ben*^{3},
{ (刺探) 间谍 *jian*^{4}*die*^{2} / (麻烦) 件扰 *jian*^{4}*rao*^{3} }

Homogeneous 1: (水果) 西瓜 *xi*^{1}*gua*^{1}, (越南) 西贡 *xi*^{1}*gong*^{4}, (喇嘛) 西藏 *xi*^{1}*zang*^{4},
{ (批萨) 西餐 *xi*^{1}*can*^{1} / (复习) 西考 *xi*^{1}*kao*^{3} }

Homogeneous 2: (奇怪) 异常 *yi*^{4}*chang*^{2}, (看法) 意见 *yi*^{4}*jian*^{4}, (座谈) 议论 *yi*^{4}*lun*^{4},
{ (绘画) 艺术 *yi*^{4}*shu*^{4} / (向前) 艺步 *yi*^{4}*bu*^{4} }

APPENDIX C

Model summary for Experiment 1, using only trials that followed odd-man out trials.

Coefficient	*b* (log)	*b* (ms)	Lower bound	Upper bound	*t*	SD (ms)
(Intercept)	6.891	983.51	886.32	1085.99	128.78*	258.13
Set Number	0.007	7.31	-1.02	15.85	1.67	n/a
Number in Set	-0.005	-5.05	-10.54	0.62	-1.90	n/a
Homogeneous	-0.108	-100.77	-156.87	-42.58	-3.24*	223.15
Heterogeneous	-0.080	-76.06	-134.41	-13.96	-2.37*	273.26
Sandhi	-0.030	-29.44	-90.72	35.37	-0.90	290.88

APPENDIX D

Model summary for Experiment 2, using only trials that followed odd-man out trials.

Coefficient	*b* (rr)	*b* (ms)	Lower bound	Upper bound	*t*	SD (ms)
(Intercept)	1.18e-03	841.99	774.54	918.69	22.75*	246.97
Set Number	6.06e-07	0.43	-1.33	2.26	0.46	n/a
Number in Set	-2.32e-07	-0.16	-1.18	0.86	-0.32	n/a
Real	-2.36e-05	17.09	-54.78	108.12	0.43	245.88
Novel, T3	-4.88e-06	-3.45	-54.45	54.17	-0.13	241.13
Real, T3	-7.45e-05	-49.69	-93.60	-2.42	-1.96*	236.82
Novel, T2, Homogeneous	-6.50e-05	-43.71	-78.21	-4.81	-2.21*	231.68
Real, T2, Homogeneous	-1.48e-04	-93.30	-123.70	-62.67	-5.37*	246.87
Novel, T3, Homogeneous	-7.08e-05	-47.35	-81.41	-10.59	-2.41*	220.99
Real, T3, Homogeneous	-5.99e-05	-40.40	-73.51	-3.25	-2.13*	210.85
Novel, T2, Heterogeneous	3.51e-05	25.61	-14.25	70.87	1.19	263.33
Real, T2, Heterogeneous	-7.73e-05	-51.48	-83.95	-16.12	-2.83*	250.51
Novel, T3, Heterogeneous	-2.97e-05	-20.56	-56.30	18.10	-1.03	212.41
Real, T3, Heterogeneous	-1.16e-06	-0.82	-38.56	41.04	-0.04	235.07
Novel, T2, Sandhi	-4.12e-05	-28.20	-63.63	13.86	-1.39	253.42
Real, T2, Sandhi	-7.79e-05	-51.81	-85.94	-16.52	-2.77*	253.15
Novel, T3, Sandhi	-1.81e-05	-12.67	-49.40	29.12	-0.61	273.06
Real, T3, Sandhi	-8.93e-06	-6.28	-43.29	33.84	-0.33	230.26

连读变调在汉语言语产生中的作用

Stephen Politzer-Ahles[1], 张杰 [2]
[1]阿拉伯联合酋长国，纽约大学阿布扎比分校； [2]美国，堪萨斯大学

题要
本研究采用odd-man-out内隐启动范式来研究汉语词汇中音系变体的产生机制。在内隐启动实验中，受试者对发音相同的词组的反应时间要比对发音不同的词组的反应时间快。本文通过两个实验来考察当一组双音节词的首音节音调只在底层和表层结构之一的表征相同，而在另一层级的表征由于上声变调而不同时是否能观察到启动效应。结果表明这类刺激只在首音节表层音调相同时才能产生启动效应。这一结果说明在言语产生的过程中，上声变调执行于发音动作开始之前。这为介于底层结构和发音动作之间的抽象音系变体过程的存在提供了依据。本文还对新近发表的类似研究进行了概括与讨论。

关键词
变调　词汇产生　内隐启动范式　普通话