Journal of Memory and Language 145 (2025) 104690

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Lexical tone is different and special: evidence from a speeded repeated production task

Chuchu Li [a,*], Sin Hang Lau [b], Victor S. Ferreira [b]

[a] Department of Psychiatry, University of California, San Diego, United States
[b] Department of Psychology, University of California, San Diego, United States

ARTICLE INFO

Keywords:
Tone
Mandarin speech production
Phonological encoding

ABSTRACT

Priming experiments and speech error studies have found cross-linguistic differences in phonological encoding. Notably, the first selectable unit (the proximate unit) differs between English and Mandarin Chinese, with the former selecting segmental units like consonants (Cs) and vowels (Vs) first, while the latter selects syllables as a whole. Further, Mandarin Chinese is tonal, meaning the same syllable is a different word depending on the tone it is spoken with. However, it remains unclear how tone is represented and processed during phonological encoding in speech production – attached to the vowel or CV, or processed independently. Across three experiments, we investigated these questions by measuring how quickly speakers produced sequences of tone-bearing CV syllables. Unlike English, speed of production was not directly linked to plan reuse (see Sevald & Dell, 1994). Instead, speech rate was robustly faster when each CV was produced with only one tone (i.e., about equal speech rate for ba2 di1 da1 bi2 and ba1 ba1 ba1 ba1), compared to when a particular CV was produced with more than one tone (i.e., slower speech rate for ba1 ba2 ba1 ba2). We suggest that Mandarin speakers represent CVs as syllable "chunks," integrating tone—a part of the structural frame with the CV (rather than a vowel), and producing the same CV with more than one tone in a sequence is difficult as a result of needing to reassign different tones to the same CV chunk.

The consensus model of language production specifies that to produce a word, speakers formulate a message, then select a representation of a word called a *lemma*, which activates the corresponding abstract sound representations and phonotactics, engages articulatory planning and finally articulation (e.g., Levelt, 1989; Levelt et al., 1999; Roelofs, 1997) – a general architecture we term here the "Levelt-type production model". Though the broad stages in this model are not claimed to vary across languages, evidence from the literature suggests that the processing units at the stage of phonological encoding differ. Thus, an inclusive production model must be able to outline production mechanisms in a way that accommodates different units during phonological encoding.

There is considerable evidence of cross-linguistic differences in terms of what phonological unit gets selected immediately after lemma selection. To unify models of phonological encoding and accommodate cross-linguistic differences at this stage, O'Seaghdha (2015) refer to the first selectable phonological units below the word as *proximate units.* In Indo-European languages (e.g., English), studies of various paradigms (e.g., speeded repeated production, form preparation, picture-word, masked priming, speech errors, see Alderete and O'Séaghdha, 2022 for review) suggest that the proximate units are segments (i.e., individual consonants and vowels). To simplify the nuanced discussion and to be consistent with the study designs in the current work, we focus our review on the production of monosyllabic words.

Segments as the proximate unit in Indo-European languages

Sevald and Dell (1994) examined how English speakers plan phonological units in the production of monosyllabic consonant–vowel-consonant (CVC) words using a speeded repeated production task, in which participants repeated four-word sequences as many times as possible in eight seconds. The experiments manipulated the repetition pattern of the onset C, V, and final C, such that each independently followed an AAAA, ABBA, or ABAB pattern (i.e., three different levels of repetition frequency, in descending order). The general logic of the paradigm is that repetition benefits (faster speech) are expected when a phonological unit is reused more frequently, reflecting easier processing compared to alternating between different units. On the other hand,

* Corresponding author at: Department of Psychiatry, University of California, 9500 Gilman Drive, La Jolla, San Diego, CA 92093, United States.
E-mail address: chl441@health.ucsd.edu (C. Li).

https://doi.org/10.1016/j.jml.2025.104690
Received 23 March 2024; Received in revised form 12 August 2025; Accepted 13 August 2025
Available online 16 August 2025

repetition may also pose a cost due to spreading activation of similar but nonidentical planning sequences, which results in selection difficulty among these candidates (see the phonological competition model; O'Seaghdha, et al., 1992; Peterson, 1991; Peterson et al., 1989). For example, repeating the onset /p/ in PICK PUN PICK PUN may benefit speech production; however, when producing PUN, the repeated /p/ cues /i/ because /i/ had just been produced after /p/ in PICK, leading to a competition with the intended /u/. This inhibitory effect may outweigh the benefits from the repetition of /p/, an effect sometimes termed *sequentially cued competition*. Critically, because it is the repetition patterns of segments that determine these benefits and costs, and this paradigm identifies functional units that are directly selected for and assembled during phonological encoding, we infer that segments serve as proximate units in English (see also O'Seaghdha & Marin, 2000).

Similarly, results from the form preparation task (also known as the implicit priming paradigm) also provides evidence that segments are the proximate units in Indo-European languages such as English and Dutch (Meyer, 1990, 1991; Roelofs & Meyer, 1998). In this paradigm, participants studied pairs of prompt-target words in blocks, with the target words in a block being either homogeneous or heterogeneous with respect to the proposed proximate unit. Borrowing examples from the English experiment in O'Seaghdha et al. (2010), when the onset segment was the proposed proximate unit, the target words in the homogeneous block all shared the same onset (*night-**d**ay; tint-**d**ye; bread-**d**ough; wet-**d**ew*), whereas the target words in the heterogeneous block had different onsets (*night-**d**ay; snap-**p**ea; grain-**r**ye; pig-**s**ow*). Participants were asked to memorize prompt-target pairs in study blocks. At test, participants were presented with each prompt word and asked to produce the corresponding target. Participants' reaction times for the target words were faster when the onset segments were homogenous rather than heterogeneous, as speakers engaged in form preparation using onset segments. This form preparation benefit implies that onset segments are the first phonological units selected and buffered in production —- that segments are the proximate units in English. Compared to the speeded repeated production task, the form preparation task only measures reaction time and does not require repeated production of phonological competitor pairs, focuses more on planning and elicits less activation of competitors, and thus generally obtains facilitatory priming from repeating segments.

Syllables as the proximate unit in Mandarin Chinese

Also using the form preparation task, repeated onset segments did not facilitate the production of Chinese words (e.g., Chen et al., 2002). Instead, benefits only appeared when the whole syllable repeated, with or without tone repetition (e.g., ***qing***1-liang2; ***qing***1-sheng1; ***qing***1-ting2; ***qing***1-jiao1, or ***fei1***-ji1; ***fei2***-pang4; ***fei3***-cui4; ***fei4***-yan2), suggesting that atonal syllables are proximate units in Mandarin Chinese. In contrast, syllables are represented as abstract schemas in English (Sevald et al, 1995). This cross-language difference in the role of segment versus syllable might be due to a number of different linguistic characteristics of these languages. For example, the structure of a Chinese syllable is simpler (e.g., no consonant clusters like sk-, spl-), the number of Chinese syllables is much smaller than in English, syllable boundaries are unambiguous, and re-syllabification occurs only in English (e.g., *read a book* is re-syllabified as *rea-da-book*). In particular, according to the WEAVER++ model (Levelt et al., 1999), only after all the phonemes are retrieved can speakers determine the arrangement of syllables in a word. In Mandarin Chinese, the lack of re-syllabification may allow speakers to retrieve the syllable as an integral unit at the beginning of phonological encoding (also see discussion in Li, Wang & Idsardi, 2015).

Studies using a masked priming paradigm shows consistent results. In word naming, the onset priming effect was robust in English (Forster & Davis, 1991) but not in Mandarin (Verdonschot et al., 2013). Instead, a syllabic priming effect was found in Mandarin masked priming, where atonal syllable overlap between the prime and the first syllable of a disyllabic target led to faster naming times (i.e., *ba4* primes *ba2* but not *bai2*; *bai4* primes *bai2* but not *ba2*), even when orthography or tone did not overlap (Chen et al., 2003). Chen et al. (2016) also showed syllabic priming effects and suggested that sharing initial consonants elicits interference, if anything, perhaps indicative of competition among similar coactivated words or syllables. Once again, masked priming suggests that proximate units are segments in English and atonal syllables in Mandarin.

The representation and processing of tone in Mandarin Chinese

In addition to the syllable versus segmental contrast, another key difference between Mandarin and the languages that most production theories have been developed for is that Mandarin is a tonal language, in which tones are lexical and affect meanings. For example, in Mandarin, the CV sequence *ma* can mean "mother", "hemp", "horse", or "to scold" depending on which of the four tones it is produced with. Despite the prevalence of tone in African (Odden, 1995) and East Asian languages (Yip, 1995), its representation in phonological encoding and the cognitive mechanisms involved in selecting or integrating tone with other phonological elements has not been well-established.

Earlier linguistic theories diverge on how tone is represented and processed in speech production. Halle and Stevens (1971) posited that tones are encoded as phonological features of vowels, whereas Leben (1978) argued that tones are representationally independent of vowels but realized as phonetic features of vowels (i.e., suprasegmental features).

A sizable portion of the existing literature on tone representation has relied on analyzing speech errors. One view is that tone functions similarly as segments and is as susceptible to selection errors just as consonants and vowels are (e.g., Alderete et al., 2019; Wan & Jaeger 1998; Moser, 1991; Shen, 1993). Some studies have reported various movement errors of tone in context (i.e., perseverations, anticipations, and exchanges), which are also commonly found with consonants and vowels in English (Fromkin, 1971). For example, in Mandarin naturalistic conversations, Wan and Jaeger (1998) found that lexical substitutions and blends have the same tones twice as often as expected by chance. In tone errors, such as the perseveration error in which *tui1 jian4 **han2*** (推荐函, *meaning letter of support*) was incorrectly produced as *tui1 jian4 **han4*** (*推荐汗*, ungrammatical and meaning recommending some sweat), tone slipped while the segmental content remained intact (i.e., Tone 4 from "jian" perseverated and substituted Tone 2 while "han" remained as intended). This suggests that tone is sensitive to context like segments are.

Another view is that tone is part of the metrical frame that segments slot into right before articulation and is thus relatively immune to selection errors (e.g., Chen, 1999; Kember et al., 2015; Roelofs, 2015). For example, Chen (1999) reported that tone errors were rarer than segment errors, and that the majority of tone errors were perseverations, whereas segmental errors were primarily anticipations. That is, tones were found to be both quantitatively and qualitatively different from segments, which is incompatible with the idea of tone engaging in segment-like early encoding. Chen proposed that tone is initially represented as part of the phonological frame similar to lexical stress in English (i.e., tone is suprasegmental) and later translated into the phonetic configuration of the vowel during phonetic encoding.

Investigations using controlled experimental methods shows that these patterns of spontaneous speech error collection are not due to perceptual biases (Chen, 1999; Alderete & Davies, 2019). One such method is a tongue twister task that carefully manipulates the patterns of segments and tones. Using a tongue twister paradigm, Kember et al. (2015) designed 120 four-character Mandarin sequences that rotated pairs of initial segments or tones, or both, across ABAB or ABBA format, across positions (e.g., "突苦哭土" (*tu1 ku3 ku1 tu3*) comprises ABBA initial segments and ABAB tones). Participants produced almost three

times as many segment errors as tone errors, which validates Chen's (1999) claim that segments are more prone to error than tones, even with a task that supposedly affords both elements the same opportunities for error due to competition between recently used components in the same tongue twister sequence (e.g., k vs. t for segments and tone 1 vs. tone 3 for tones in the above example). (It is fair to note that part of this disparity may be because Mandarin possesses a much larger inventory of consonants and vowels than tones.).

Evidence from the form preparation paradigm which measures reaction time rather than speech errors, described above, also indicates that tone is a part of metrical frame in phonological encoding, as well as that the smallest phonological unit that benefits Mandarin production through repetition is the atonal syllables (rather than tone-bearing syllables; Chen et al., 2002; O'Seaghdha et al., 2010). Critically, although syllable + tone elicited larger form preparation effects than atonal syllables, tone alone did not lead to form preparation effects (Chen et al., 2002). Accordingly, Chen and Chen (2013) proposed a modified Levelt-type production model that includes the proximate principle, suggesting that atonal syllable representations are the first to be selected at the stage of word form encoding in Chinese, and tone serves as a part of the metrical frame, similar to stress in Indo-European languages. Tonal syllables are formed later through unit-to-frame associations. This modified Levelt-type production model is also supported by ERP evidence in a Cantonese picture-word interference paradigm – interference from tonal versus atonal syllables had similar effects in a stimulus-locked analysis, which is more sensitive to early-stage stimulus processing; but the two conditions produced different effects in response-locked analysis, which is more sensitive to later processing closer to overt naming (Wong et al., 2023). Using a primed picture naming task, a recent ERP study on Mandarin word production also suggested that tone-to-syllable integration happens in a later ERP time window (Chen & Zhang, 2025).

Note that these differences between tonal and atonal languages do not require that Chinese production arise from a fundamentally different architecture. Indeed, Roelofs (2015) accounted for form preparation effects in different languages using WEAVER++ simulations, including English, Mandarin Chinese, and Japanese with three different proximate units, supporting an inclusive model that accommodates cross-language difference in phonological encoding. Though the WEAVER++ model did not involve the concept of proximate unit, it proposed that phonological contents (e.g., segments or syllables) and frames (e.g., tones) are activated in a parallel fashion, followed by serial content-to-frame association (Levelt et al., 1999). At the phonological encoding stage, the cross-language difference was revealed in what phonological content was retrieved and what served as the frames, similar to the proximate unit principle and was incorporated in Roelofs's (2015) Mandarin model directly. The proximate principle also explained the role of other phonological units. For example, in Chinese production, phonemic segments are included as non-proximate units that are selected later than the proximate unit — the atonal syllable. An attempt to prepare for the segment in advance would violate the standard operating procedure of phonological encoding in Chinese and may result in aborting the procedure and restarting (this does not happen in English, which does not have a phonological unit smaller than the proximate unit), sometimes leading to inhibitory effects when the onset repeats (Chen & Chen, 2013, Experiment 2) or when the rime repeats (Li, Wang, & Idsardi, 2015).

The present study

Given that speech error studies have revealed different patterns of tone and segment errors, and that form preparation studies only measure onset latency, the current work aims to investigate Mandarin phonological processing using a production time measure. In particular, we used a method that further lends insights into the representation and online processing of tone, focusing on how tone is integrated with atonal syllables when producing multiple monosyllables continuously, which may improve the modified Levelt-type production model with the proximate principle (Chen & Chen, 2013). Specifically, we adapted Sevald and Dell's (1994) paradigm of speeded repeated production of monosyllabic words and asked Mandarin speakers to produce C + V + tone sequences with different repetition patterns. Since Mandarin proximate units and tone representations (or phonological encoding during speech production) have not been studied using production times, we first used a design that entertained all possibilities. Experiment 1 was thus exploratory in nature, exploring possibilities that included but were not limited to C, V, and tone functioning as independent units in phonological encoding, CV functioning together as one integrated syllabic unit, or tone functioning as a phonological feature of V or CV syllable. Although not all possibilities are plausible (e.g., all segments in a syllable function together as a whole unit seemed more possible than each segment functions separately), we included all combinations, given that the experimental design provides the opportunity to do so. Experiments 2 and 3 are confirmatory tests of results in Experiment 1 that further examines how different phonological units are encoded in speech production.

The first possibility is that tone is represented and processed with segments. Due to pitch changes having a larger influence on how a vowel is articulated relative to how a consonant is articulated (Lin et al., 2013), it is more likely that (if it is represented with segments) tone would be attached to vowels rather than to consonants. If tone is represented together with vowels, that means there are as many copies of vowel representations as there are tones (e.g., /a/1, /a/2, /a/3, /a/4, /i/1, /i/2, /i/3, /i/4, etc.). Assuming the pinyin-based analysis that there are five monophthongs in Mandarin, that would mean there are 20 vowel representations in total. To fit this possibility in with the form preparation evidence showing that atonal syllables can be initially selected, it may be that the atonal syllable is selected as a whole unit at first, then non-proximate units were processed, including the onset (consonant) and vowel with tone.

The second possibility is that tone is represented together with the whole syllable, so that there are as many copies of syllable representations as there are tones (e.g., /ba/1, /ba/2, /ba/3, /ba/4, /bi/1, /bi/2, /bi/3, /bi/4). This possibility might be less likely given strong evidence of form preparation benefits from repeated syllables alone regardless of tone.

These two possibilities benefit from not needing an additional integration process, but they do not seem computationally efficient, considering the increase in representational load. They assume that speakers of tonal languages potentially have to hold many times more representations in their syllabary or library of segments than strictly necessary, based on the segmental inventory of the language. That leads to the third possibility, that tone is represented independently of syllables and segments but needs to be integrated with them instead. This possibility necessitates specifying the timing and mechanism through which tone is combined with the rest of the phonological units, but it achieves computational efficiency by avoiding duplicates of units with similar or identical features (other than tone).

In the context of a speeded repeated production task, tone being attached to vowels or syllables would predict that production would be slower if more distinct tones are in a sequence, as a result of increased competition between vowels or syllables with different tones. A specific example is that if tone is attached to vowels, V + tone repeating in AAAA should elicit a faster speech rate than other V + tone patterns regardless of how the initial C repeats; if tone is attached to the entire CV syllable, only the condition where the entire CV syllable with tone repeats should yield benefits. In contrast, if tone is represented independently of syllables or vowels, the repetition pattern or reuse of tone may not necessarily have an impact on production times (i.e., tone repeating as AAAA may elicit similar production speed compared to tone repeating in ABAB or ABBA when the syllable sequence stays the same), if speakers have to integrate a tone every time they produce a syllable, regardless of the

pattern of the sequences.

Experiment 1: Exploration of consonant, vowel, and tone encoding

In Experiment 1, we adopted a full factorial design to explore how Mandarin speakers encode phonological information in speech production. Following Sevald and Dell (1994), we used monosyllabic words and asked participants to produce C + V + tone sequences (resembling the English C + V + C sequences in Sevald & Dell, 1994) of different repetition frequency patterns.

Method

Participants

Fifty-eight undergraduates from the University of California San Diego participated in the experiment in exchange for course credit. Fourteen participants were excluded from our analyses because they either did not produce more than 50 % usable data (exclusion criteria explained below in Coding and Data Analysis), or had extensive prior knowledge of tonal languages other than Mandarin (including Cantonese). All 44 remaining participants indicated that they were native Mandarin Chinese speakers with little to no exposure to other tonal languages, and that they moved to the US after the age of 15.

Materials and design

We used a word sequence repetition task (adapted from Sevald & Dell, 1994) to assess the effects of the repetition frequency of each position (consonant, vowel, or tone) and their interaction on speech rate. Materials in this task were four-syllable sequences. All four-syllable sequences comprised semantically-unrelated monosyllabic words (i.e., all sequences were nonwords) with no alternate pronunciation and were easy to read, so that semantic processing and orthographic difficulty would not affect performance. Additionally, we avoided Tone 3 (due to tone sandhi) and polyphones. The study used a within-subjects, 3 (C Pattern) x 3 (V Pattern) x 3 (Tone Pattern) factorial design, such that each of the three Positions of each monosyllabic word (C, V, and tone; generated from three different phoneme sets) repeated independently in one of three repetition patterns (AAAA, ABBA, or ABAB). This created 27 possible combinations for each phoneme set. The experiment used three phoneme sets, so that there were 81 trials in the experiment. Each phoneme set consisted of two sounds in each position, and three counterbalancing lists were constructed to randomize which sound in each position was arbitrarily assigned as Sound A or Sound B in each repetition pattern across participants. Table 1 shows the pinyin of the sounds in the three phoneme sets. Taking Phoneme Set 1 as an example, if the C repeated in a AAAA pattern (*b_b_b_b_*), V in a ABBA pattern (*_a_i_i_a_*), and tone in a ABAB pattern (*_2_1_2_1*), the resulting four-word sequence would be "拔逼鼻巴" (*ba2 bi1 bi2 ba1*). Table 2 includes 27 example trials from one Phoneme Set.

Procedure

Participants first completed a language history questionnaire, in which they indicated whether they identified as native speakers of Mandarin Chinese, whether they spoke any other tonal languages, and the age at which they moved to the US.

Then, an experimenter explained the instructions of the experimental task (modeled closely after Sevald & Dell, 1994) and went through five practice trials with each participant. The beginning of each trial was signaled by a fixation cross (presented for 200 ms) and a high-pitched tone. Then, participants saw a four-syllable sequence written in Simplified Chinese characters centered on the screen, with the word "准备" (prepare) printed underneath for eight seconds. They were instructed to silently rehearse the sequence during that period in preparation for the upcoming production phase. At the end of the eight seconds, participants heard three low-pitched tones and one high-pitched tone as a countdown signal for the production phase. The high-pitched tone and the disappearance of the word "准备" (prepare) signaled the beginning of the eight second long production phase, during which participants were instructed to produce the four-word sequence aloud repeatedly as quickly and as accurately as possible. At the end of the production phase, a high-pitched tone signaled the end of the trial. Participants pressed the

Table 1
Sounds in each phoneme set.

Phoneme Set 1			Phoneme Set 2			Phoneme Set 3		
C	V	Tone	C	V	Tone	C	V	Tone
b	a	2	n	i	2	l	a	1
d	i	1	t	u	4	p	u	4

Table 2
Example trials of the 27 conditions from phoneme set 1.

Pinyin	Characters	C pattern	V pattern	Tone pattern	CV pattern
ba2 ba2 ba2 ba2	拔拔拔拔	AAAA	AAAA	AAAA	AAAA
ba2 ba1 ba1 ba2	拔巴巴拔	AAAA	AAAA	ABBA	AAAA
ba2 ba1 ba2 ba1	拔巴拔巴	AAAA	AAAA	ABAB	AAAA
ba2 bi2 bi2 ba2	拔鼻鼻拔	AAAA	ABBA	AAAA	ABBA
ba2 bi1 bi1 ba2	拔逼逼拔	AAAA	ABBA	ABBA	ABBA
ba2 bi1 bi2 ba1	拔逼鼻巴	AAAA	ABBA	ABAB	ABBA
ba2 bi2 ba2 bi2	拔鼻拔鼻	AAAA	ABAB	AAAA	ABAB
ba2 bi1 ba1 bi2	拔逼巴鼻	AAAA	ABAB	ABBA	ABAB
ba2 bi1 ba2 bi1	拔逼拔逼	AAAA	ABAB	ABAB	ABAB
ba2 da2 da2 ba2	拔达达拔	ABBA	AAAA	AAAA	ABBA
ba2 da1 da1 ba2	拔搭搭拔	ABBA	AAAA	ABBA	ABBA
ba2 da1 da2 ba1	拔搭达巴	ABBA	AAAA	ABAB	ABBA
ba2 di2 di2 ba2	拔笛笛拔	ABBA	ABBA	AAAA	ABBA
ba2 di1 di1 ba2	拔低低拔	ABBA	ABBA	ABBA	ABBA
ba2 di1 di2 ba1	拔低笛巴	ABBA	ABBA	ABAB	ABBA
ba2 di2 da2 bi2	拔笛达鼻	ABBA	ABAB	AAAA	ABCD
ba2 di1 da1 bi2	拔低搭鼻	ABBA	ABAB	ABBA	ABCD
ba2 di1 da2 bi1	拔低达逼	ABBA	ABAB	ABAB	ABCD
ba2 da2 ba2 da2	拔达拔达	ABAB	AAAA	AAAA	ABAB
ba2 da1 ba1 da2	拔搭巴达	ABAB	AAAA	ABBA	ABAB
ba2 da1 ba2 da1	拔搭拔搭	ABAB	AAAA	ABAB	ABAB
ba2 di2 bi2 da2	拔笛鼻达	ABAB	ABBA	AAAA	ABCD
ba2 di1 bi1 da2	拔低逼达	ABAB	ABBA	ABBA	ABCD
ba2 di1 bi2 da1	拔低鼻搭	ABAB	ABBA	ABAB	ABCD
ba2 di2 ba2 di2	拔笛拔笛	ABAB	ABAB	AAAA	ABAB
ba2 di1 ba1 di2	拔低巴笛	ABAB	ABAB	ABBA	ABAB
ba2 di1 ba2 di1	拔低拔低	ABAB	ABAB	ABAB	ABAB

spacebar to advance to the next trial whenever they were ready. (Note that the removal of the visual stimulus ensures that initial orthographic processing is completed before articulation begins, so that the production stage more purely reflects phonological rather than orthographic encoding.) The experimenter provided feedback to the participants for five practice trials before the participants completed the 81 experimental trials independently.

Coding and data analysis

The spoken responses for each trial were audio-recorded and later coded into the dependent variable (speech rate) with the aid of Audacity, an audio processing program. Specifically, we manually counted the number of syllables produced within the eight second production period, excluding any production that overlapped with the beginning or terminal beep signal of each trial. Based on our goal to indirectly infer the proximate units of speech using speech rate when a sequence was correctly produced, we excluded 523 (14.67 %) trials which included speech errors, significant disfluencies (other than breathing, such as the inclusion of fillers or any other non-speech sounds or restarting a sequence), abnormal reaction times (below 150 ms or above 800 ms in response to the beginning signal) or other technical errors, resulting in 3,041 (85.33 %) analyzable trials. Two counters (native Mandarin speakers) were each responsible for counting half of the trials and cross-checking the other half that the other counter initially counted. Of the analyzable trials, only 29 trials led to inter-counter disagreements, which were all resolved with a third round of counting. The syllable counts were then transformed into average production times per syllable by dividing 8,000 ms by the syllable count.

The statistical analyses were conducted using R (R Core Team, 2014) and the lme4 package (Bates et al., 2014). In our main analysis, following Sevald and Dell (1994), we focused on average production times and used a linear mixed-effects model to analyze the effects of C pattern, V pattern, tone pattern, and their interactions. We attempted to use the maximal random effects structure for the models, and the models that converged included random intercepts for participants and items (i.e., lmer (AvgTime ~ O * V * Tone +(1|participant)+(1|item)). Each item refers to each four-syllable sequence.

Data and analysis scripts are publicly available at https://osf.io/hqde6/.

Results

Our planned analysis revealed significant main effects of repetition pattern for all three positions: C, V, and tone. However, unlike English, more frequent repetition did not necessarily lead to faster speech rate. Table 3 shows the by-subject mean production times for all 27 conditions, and Table 4 shows the marginal means for C,V, and tone repetition patterns. Most strikingly, when C, V, and tone patterns were all AAAA, the speech rate was *not* the fastest ($M = 278$ ms; e.g., when both V and tone patterns were ABBA and the C pattern was AAAA, $M = 269$ ms). As can be seen, speech rate was the fastest when the C pattern was ABBA (280 ms), followed by ABAB (285 ms). Surprisingly, AAAA was the slowest (290 ms). The main effect of the C repetition pattern was statistically significant, $\chi^2(2) = 12.08$, $p = .002$. Similarly for the V pattern, ABBA was the fastest (279 ms), followed by ABAB (283 ms), and finally AAAA (292 ms); the main effect of V repetition pattern was statistically significant, $\chi^2(2) = 21.07$, $p < .001$. Meanwhile, the tone repetition pattern followed a different trend, with AAAA being the fastest (280 ms), then ABAB (286 ms), and ABBA (290 ms); the main effect of tone repetition pattern was statistically significant, $\chi^2(2) = 12.57$, $p = .002$.

The proportion of invalid trials of each condition are shown in Tables 3 and 4. Although there is some trade-off between speed and accuracy/disfluency (e.g., the C pattern), the overall pattern of results cannot be explained as a speed-accuracy trade-off (e.g., when both C and tone patterns were AAAA and the V pattern was ABAB, speed and accuracy were similar to AAAA). Given that our dataset contained far fewer errors in each condition than previous work using the same paradigm that focuses on error type analysis, we did not have enough power to conduct systematic comparisons between error types that may reveal time course of tone vs. segment encoding.

Table 3
Average production times and proportion of excluded trials of the 27 conditions.

C pattern	V pattern	Tone pattern	CV pattern	Average production time (ms)	Proportion of excluded trials
AAAA	AAAA	AAAA	AAAA	278	4 %
AAAA	AAAA	ABBA	AAAA	315	10 %
AAAA	AAAA	ABAB	AAAA	304	7 %
AAAA	ABBA	AAAA	ABBA	283	11 %
AAAA	ABBA	ABBA	ABBA	269	11 %
AAAA	ABBA	ABAB	ABBA	299	15 %
AAAA	ABAB	AAAA	ABAB	281	5 %
AAAA	ABAB	ABBA	ABAB	297	11 %
AAAA	ABAB	ABAB	ABAB	287	11 %
ABBA	AAAA	AAAA	ABBA	282	34 %
ABBA	AAAA	ABBA	ABBA	283	27 %
ABBA	AAAA	ABAB	ABBA	311	23 %
ABBA	ABBA	AAAA	ABBA	270	18 %
ABBA	ABBA	ABBA	ABBA	275	14 %
ABBA	ABBA	ABAB	ABBA	276	13 %
ABBA	ABAB	AAAA	ABCD	284	19 %
ABBA	ABAB	ABBA	ABCD	281	17 %
ABBA	ABAB	ABAB	ABCD	273	17 %
ABAB	AAAA	AAAA	ABAB	279	11 %
ABAB	AAAA	ABBA	ABAB	315	24 %
ABAB	AAAA	ABAB	ABAB	277	8 %
ABAB	ABBA	AAAA	ABCD	285	15 %
ABAB	ABBA	ABBA	ABCD	286	22 %
ABAB	ABBA	ABAB	ABCD	279	15 %
ABAB	ABAB	AAAA	ABAB	281	9 %
ABAB	ABAB	ABBA	ABAB	297	11 %
ABAB	ABAB	ABAB	ABAB	273	14 %

Table 4
Mean production time (ms) and proportion of excluded trials (in parentheses) as a function of C, V, and tone repetition pattern.

Pattern	C	V	Tone
AAAA (identical)	290 (9 %)	292 (16 %)	280 (14 %)
ABBA (near repetition)	280 (20 %)	270 (15 %)	290 (16 %)
ABAB (far repetition)	285 (14 %)	283 (12 %)	286 (13 %)

Our results thus far have shown that speech rates in Mandarin production cannot be reliably predicted by repetition of each of the phonological elements alone. In addition, we obtained significant interactions between C and tone, $\chi^2(4) = 23.64$, $p < .001$; V and tone, $\chi^2(4) = 27.12$, $p < .001$; as well as a three-way interaction between C, V, and tone, $\chi^2(8) = 23.04$, $p = .003$. As described above, the repetition patterns in these three positions trended in different directions, rendering these interactions difficult to interpret in terms of repetition frequency alone. Moreover, these interactions suggest that perhaps the three phonological elements should not be treated as completely independent of each other. Thus, we conducted additional post-hoc analyses to explore whether the current data set could help to inform how phonological units are encoded in Mandarin production, by analyzing what groupings of C, V, and tone could better predict speech rate.[1]

Post-hoc analyses

In light of the surprising findings from our planned analysis, particularly that the repetition frequency of the phonological elements (when they were treated as independent units) did not reliably predict speech rate, we explored the potential dependence between C, V, and

[1] We also analyzed CV repetition pattern, which was not manipulated when we design Experiment 1. See the third post-hoc analysis below.

tone in three post-hoc analyses.

First, we explored the possibility of a speech rate advantage when there are covarying repetition patterns. That is, perhaps Mandarin speech production is more affected by whether different phonological units share the same repetition pattern or not, more so than what the exact pattern is. At this point, we have not yet established how exactly phonological units are encoded in Mandarin speech production, other than suggesting that (given the results of the planned analyses) it is highly unlikely that C, V, and tone function as independent units. Thus, in order to explore if and how Mandarin phonological elements may be dependent on each other, we recoded our data by analyzing two of the elements at a time (i.e., C and V, V and tone, and C and tone) and comparing production times in trials where those two elements either shared the same repetition pattern or not. The results showed that whether C and V shared the same pattern made no difference to speech rate (285.51 ms when C and V shared the same pattern and 285.17 ms they did not; 0.35 ms difference, z = -0.09, p = .93), whereas significantly faster speech rates were observed when C and tone shared the same pattern (277.73 ms for same and 289.06 for different; 11.33 ms difference, z = 3.05, p = .002), and when V and tone shared the same pattern (276.93 ms for same and 289.46 ms for different; 12.54 ms difference, z = 3.25, p = .001). Because C and V seemed to behave similarly in relation to tone, we additionally explored the notion that CV may together function as a syllabic unit. We once again recoded the data and compared trials in which tone shared the same repetition pattern with the syllables (i.e., CV treated as one combined unit) with other trials. We found that speech rates were also faster when CV shared the same pattern with tone (274.30 ms for same and 286.66 ms for different; 12.36 ms difference, z = 2.15, p = .03).

The set of comparisons above points towards two points: One, tone seemed to have a privileged status in speech production, which was easier when tone was "attached" to another element (i.e., shared the same repetition pattern with another element). Two, given that the speech advantages we observed above were around the same magnitude regardless of whether tone was attached to C or V alone or CV as a unit, we suspected that C and V may not be represented independently in Mandarin, unlike in English. However, these results remained preliminary, as they were inferences that were not obtained from direct comparisons. Our second post-hoc analysis attempted to address this limitation by using a single model. Specifically, we re-classified the trials into five categories: trials in which only one tone was involved (tone pattern was AAAA), trials in which there were two tones (i.e., tone pattern was either ABBA or ABAB) that covaried with (i.e., shared the same repetition pattern with) C alone, V alone, both C and V, and neither C nor V.

We compared the production times across the five covariation categories, in order to assess whether trials involving two tones were slower than those involving only one, and whether covariation between tone and other elements produced any speech rate advantage. The results are shown in Fig. 1. Our model revealed that there was a statistically significant difference in speech rate among these five categories ($\chi^2(4)$ = 41.80, p < .001). Specifically, the effect was solely driven by the slow production times in the "neither" condition. Pairwise comparisons showed that there was no difference between trials that only contained one tone (279 ms; 1023 trials) and trials that contained two tones sharing the same repetition pattern as C alone (279 ms; 439 trials), V alone (279 ms; 448 trials), or both C and V (272 ms; 226 trials); pairwise comparisons ps > .05. Importantly, the trials involving two tones that did not share repetition pattern with any other element (302 ms; 905 trials) were significantly slower than all other conditions: 22.39 ms slower than one tone (z = -5.22, p < .001), 22.24 ms slower than sharing with C alone (z = -4.98, p < .001), 23.06 ms slower than sharing with V alone (z = -4.95, p < .001), and 29.7 ms slower than sharing with both C and V (z = -4.89, p < .001).

These comparisons confirmed that speech production was significantly more difficult when tone did not covary with other elements.

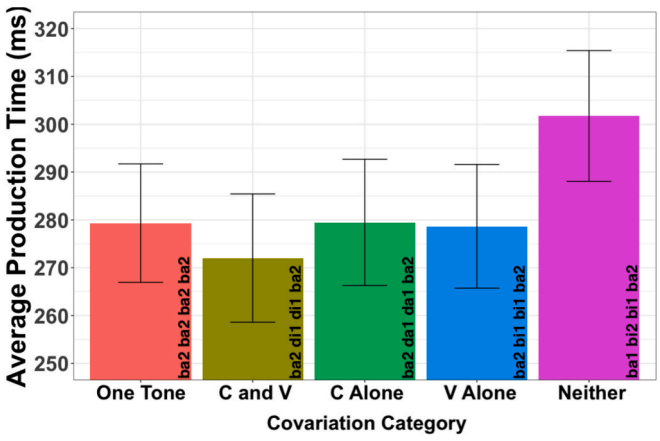

Fig. 1. Average production times for trials with one tone only (e.g., ba2 ba2 ba2 ba2), two tones that covaried with both C and V (e.g., ba2 di1 di1 ba2), C Alone (e.g., ba2 da1 da1 ba2), V Alone (e.g., ba2 bi1 bi1 ba2), and Neither (e.g., ba1 bi2 bi1 ba2)Note. Error bars represent standard errors.

Additionally, based on the lack of additive speech rate advantage when tone shared the repetition with both C and V, compared to with C or V alone, C and V may form a single proximate unit together in Mandarin. Notably, trials involving one tone were not significantly different from trials involving two tones, with the exception of the "neither" condition, suggesting that it may not be the number of unique phonological units but rather the CV-tone pairing that affects Mandarin production. In particular, we speculated that perhaps trials the "neither" condition (e.g., *ba2 bi1 ba1 bi2*: AAAA for C, ABAB for V, and ABBA for tone) was particularly difficult because it is only in this condition that each CV was paired with two different tones in the sequence (e.g., ABAB for CV and ABBA for tone), such that speakers had to repeatedly "detach" a tone from a CV syllable reattach another tone in preparation for the next time it appears in the sequence, slowing production (ba was paired with tone 2 and then tone 1 within the same production sequence). In all other conditions, each CV syllable is only paired with one tone (see Fig. 1). We refer to this speculation as the *reattachment hypothesis* and explored the validity of it in the third post-hoc analysis.

In the third and final post-hoc analysis, we redefined repetition using CV syllable as a unit. Notably, in addition to the previous patterns (AAAA, ABBA, and ABAB), this way of coding now presents a new possibility of the ABCD pattern (e.g., *ba_di_da_bi_*; previously coded as ABBA for C and ABAB for V; see Table 2). More importantly, we also classified trials into conditions where "reattaching" is required versus not. Reattaching is defined by whether each unique CV syllable is only paired with one (i.e., no reattaching) versus two tones in the sequence (and thus requiring detaching from one tone and reattaching to another repeatedly).

We assessed the effects of CV repetition pattern, reattaching, and their interaction on production times. Fig. 2 shows the results. Consistent with what we have reported thus far, CV repetition pattern did not affect speech rate ($\chi^2(3)$ = 6.15, p = .10) and there was thus no interaction with reattaching ($\chi^2(2)$ = 1.83, p = .40). Critically, we observed a main effect of reattaching ($\chi^2(1)$ = 16.37, p < .001), such that speech rate was on average 23 ms slower when each CV syllable was required to reattach to a different tone within the sequence (301 ms), compared to when no reattaching was required (278 ms).

Discussion

The planned analyses in Experiment 1 showed that the repetition patterns in C, V, and tone trended in different directions, rendering interactions between these phonological elements. Most strikingly, production was not the fastest when all elements repeat in the AAAA

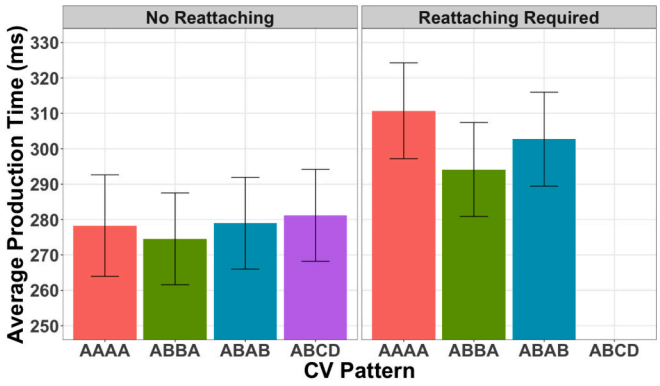

Fig. 2. Average Production Times for Trials of Various CV Repetition Patterns that Required Reattaching Versus Not Note. Error bars represent standard errors.

pattern. These results suggest that the three phonological elements should not be treated as completely independent of each other. Supportively, the post-hoc analyses suggested that participants seemed to benefit from the conditions in which tone covaried with other phonological elements. C and V may function together as a single phonological unit in production, whose production difficulty relies on the pairing with tone. This hypothesis was supported by two key findings from post-hoc analyses: First, the covarying advantage was not additive. Tone covarying with one segment or with the whole atonal syllable yielded the same benefits. Because changing one segment also changes the whole atonal syllable (e.g., ba vs. da are two different syllables, just as ba vs. di), this lack of additivity suggests that the critical point is whether the whole CV syllable changes. Second, regardless of the CV repetition pattern, production was slowed when a particular CV was paired with more than one tone in a sequence. In addition, the CV repetition pattern itself did not predict production speed. Therefore, we developed the *Reattachment Hypothesis*, which speculates that Mandarin speakers represent CVs (or atonal syllables) as "chunks," programming each with tone upon phonetic encoding, so that producing the same chunk with more than one tone in a sequence is difficult. We simplified the design of Experiment 1 to directly test the reattachment hypothesis in Experiment 2.

Experiment 2: Investigation of the reattachment hypothesis

In Experiment 2, we simplified the design by treating C and V as a single phonological unit, so that C and V always repeated in the same pattern. Adopting a full factorial design with this change (thus fewer combination possibilities or conditions), we first repeated the planned analyses in Experiment 1 to examine whether the repetition pattern of CV and/or tone might predict production speed, and tested the reattachment hypothesis to see whether participants would benefit from tone covarying with CV.

Method

Participants

Our pre-registered target sample size (based on a power analysis[2] conducted with data from Experiment 1) was 26 participants. Forty-four undergraduates from the University of California San Diego participated in the experiment in exchange for course credit. Eighteen were excluded from our analyses because they either did not produce more than 50 % usable data (based on the same exclusion criteria in Experiment 1, noted above in Coding and Data Analysis), had extensive prior knowledge of tonal languages other than Mandarin, including Cantonese, or had unstable internet connections. All 26 remaining participants indicated that they were native Mandarin Chinese speakers with little to no exposure to other tonal languages, and that they moved to the US after the age of 15.

Materials and design

The primary goal of Experiment 2 was to test the reattachment hypothesis with a simpler design. That is, we investigated whether speech rate was indeed slower when each CV was paired with more than one tone in a sequence (i.e., requiring reattaching). We used the same word sequence repetition task as Experiment 1, with the logic that slower speech rate indicates greater encoding difficulty. We adopted a within-subjects, 2 (CV Number) x 2 (Tone Number) x 2 (Repetition Pattern) factorial design, such that the four-syllable sequences comprised either one or two CV units; either one or two tones; and when two CV units or two tones were involved, they either followed an ABBA or ABAB pattern. Out of the eight conditions, the two conditions that paired one CV with two tones were further coded as "requiring reattaching", while the other six were recoded as "no reattaching required". The materials were constructed using the nine phoneme sets in Table 5, counterbalancing across participants which sound was assigned as Sound A or Sound B in each repetition pattern. In sum, each participant contributed to 72 trials (i.e., 8 conditions x 9 Phoneme Sets). Table 6 shows the eight experimental conditions from Phoneme Set 1. Taking this set as an example, "拔巴巴拔" (*ba2 ba1 ba1 ba2*) and "笛低笛低" (*di2 di1 di2 di1*) were the two conditions that required each CV to detach from a tone and reattach to another within the sequence, with the former reattaching less frequently than the latter (i.e., reattaching every other word versus every word).

Procedure

The procedure was exactly the same as Experiment 1, except that the experiment was administered remotely over the internet, via Zoom. Instead of having the participants control the pace of the experiment in between trials, we constructed a video such that each trial was presented one after another without breaks. An experimenter showed the video to each participant using the screen share function, and the participant's spoken responses were recorded. The preparation and the production durations remained unchanged from Experiment 1 (i.e., eight seconds each).

Coding and data analysis

Similar to Experiment 1, we manually counted the number of syllables produced within the eight second production period, excluding any production that overlapped with the beginning or terminal beep signal of each trial. Based on our goal to indirectly infer the proximate units of speech using speech rate when a sequence was correctly produced, we excluded 331 (17.68 %) trials using the same exclusion criteria as

Table 5
Sounds in each phoneme set.

Phoneme Set	CV	Tone	Phoneme Set	CV	Tone
1	ba	2	6	du	4
	di	1		mi	2
2	ni	2	7	ke	2
	tu	4		ju	4
3	la	1	8	pa	2
	pu	4		zu	1
4	ti	2	9	ji	2
	qu	4		hu	4
5	ma	2			
	fu	4			

[2] The power analysis was conducted through simulation based on the first experiment's data, with the SimR package in Rstudio and using Monte Carlo simulations on a range of sample sizes (Green & MacLeod, 2016). Each simulation was run with a linear mixed effect model structure on the reattachment main effect, and results of all analyses showed that 26 participants would be enough to reach the 80% power threshold.

Table 6
Example trials of the 8 conditions from phoneme set 1.

Pinyin	Characters	No. of CV	No. of Tone	CV or tone repetition pattern if not AAAA	Reattaching
ba2 ba2 ba2 ba2	拔拔拔拔	One	One	N/A	No
ba2 di2 di2 ba2	拔笛笛拔	Two	One	ABBA	No
ba2 ba1 ba1 ba2	拔巴巴拔	One	Two	ABBA	Yes
ba2 di1 di1 ba2	拔低低拔	Two	Two	ABBA	No
di2 di2 di2 di2	笛笛笛笛	One	One	N/A	No
di2 ba2 di2 ba2	笛拔笛拔	Two	One	ABAB	No
di2 di1 di2 di1	笛低笛低	One	Two	ABAB	Yes
di2 ba1 di2 ba1	笛巴笛巴	Two	Two	ABAB	No

Experiment 1, resulting in 1541 (82.32 %) analyzable trials. Two counters (native Mandarin speakers) were each responsible for counting half of the trials and cross-checking the other half that the other counter initially counted. Of the analyzable trials, only 25 trials led to inter-counter disagreements, which were all resolved with a third round of counting. The syllable counts were then transformed into average production times per syllable by dividing 8,000 ms by the syllable count.

There were two major analyses in this experiment. First, we used a linear mixed-effects model to analyze the effects of the number of CV, Tone, Repetition Pattern, and their interactions on average production times. We used the maximal random effects structure, such that the resulting model included all the fixed effects as by-subject random slopes and the counterbalancing lists as the by-item random slope. This analysis allowed us to (as in Experiment 1) investigate whether the number of different phonological units involved and their repetition frequency affected speech rate. Second, we used another linear mixed-effects model to analyze the effects of reattaching and its interaction with the repetition pattern on speech rate. Again, we used the maximal random effects structure, which included reattaching, repetition pattern, and their interaction as by-subjects random slopes and the counterbalancing lists as the by-item random slope.

Results

Table 7 and Fig. 3 summarize the average production times for each of the eight conditions of the number of CVs, tones, and repetition pattern. The first analysis examined whether the number of different CVs and tones in a sequence, as well as their repetition pattern, affected speech rate. Therefore, CV number (1 vs. 2), tone number (1 vs. 2), repetition pattern (ABBA vs. ABAB), and all the two-way and three-way interactions were entered as fixed effects. Subjects and items were entered as random intercepts with related random slopes. We found a significant main effect of the number of distinct CVs on speech rate ($\chi^2(1) = 28.57$, $p < .001$), but not of the number of tones, and neither interacted with repetition pattern (*ps* > .05). Participants were about 20 ms faster at producing sequences with two CVs versus one, regardless of the repetition pattern. Specifically, the average production times were 322 ms for one CV versus 300 ms for two when the repetition pattern was ABAB (i.e., 22 ms difference; $t = 3.94$, $p < .001$), and 318 ms for one CV versus 293 ms for two when the repetition pattern was ABBA (i.e., 25 ms difference; $t = 4.27$, $p < .001$). In contrast, the number of tones did not seem to significantly affect speech rate. The average production times were 308 ms for one tone versus 314 ms for two tones when the

Table 8
Example trials of the 8 Examples (4 conditions) from Phoneme Set 1.

Pinyin	Characters	CV repetition	Tone repetition	Reattaching
ba1 di2 ba1 di2 ba2 di1 ba2 di1	巴笛巴笛 拔低拔低	ABAB	ABAB	No
ba1 di2 ba2 di1 ba2 di1 ba1 di2	巴笛拔低 拔低巴笛	ABAB	ABBA	Yes
ba1 di2 di2 ba1 ba2 di1 di1 ba2	巴笛笛巴 拔低低拔	ABBA	ABBA	No
ba1 di2 di1 ba2 ba2 di1 di2 ba1	巴笛低拔 拔低笛巴	ABBA	ABAB	Yes

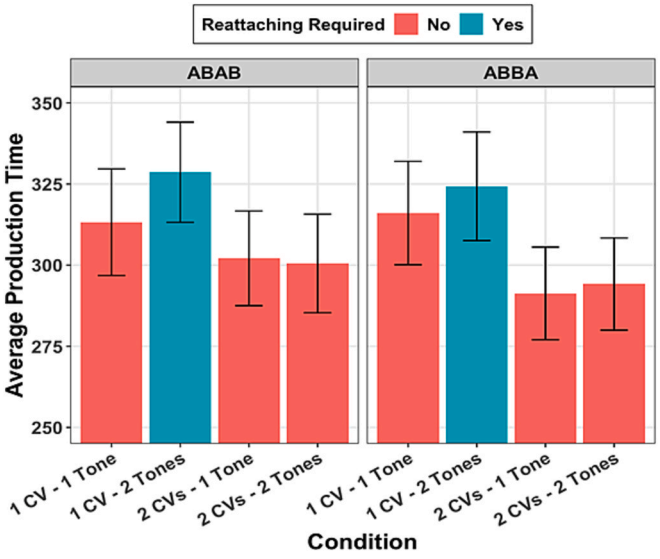

Fig. 3. Average production times for the eight conditions of CVs, tones, and repetition patterns in experiment 2 note. Error bars represent standard errors. The blue bars indicate conditions that required reattaching, whereas red bars indicate conditions that did not require reattaching. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 7
Example trials from phoneme set 1, average production times, and standard errors of the 8 conditions in experiment 2.

Pinyin	Characters	No. of CV	No. of tone	Repetition pattern	Reattaching	Average production time
ba2 ba2 ba2 ba2	拔拔拔拔	One	One	N/A	No	316 (16)
ba2 di2 di2 ba2	拔笛笛拔	Two	One	ABBA	No	291 (14)
ba2 ba1 ba1 ba2	拔巴巴拔	One	Two	ABBA	Yes	324 (17)
ba2 di1 di1 ba2	拔低低拔	Two	Two	ABBA	No	294 (14)
di2 di2 di2 di2	笛笛笛笛	One	One	N/A	No	313 (16)
di2 ba2 di2 ba2	笛拔笛拔	Two	One	ABAB	No	302 (15)
di2 di1 di2 di1	笛低笛低	One	Two	ABAB	Yes	329 (14)
di2 ba1 di2 ba1	笛巴笛巴	Two	Two	ABAB	No	301 (15)

repetition pattern was ABAB (i.e., −5.75 ms difference, t = -1.12, p = .27), and 303 ms for one tone versus 308 ms for two when the repetition pattern was ABBA (i.e., −4.49 ms difference, t = -0.83, p = .40).

Consistent with the results of Experiment 1, repetition pattern did not reliably predict speech rate ($\chi^2(1) = 2.31, p = .13$). Additionally, we found that participants in Experiment 2 were slower at producing sequences with one CV rather than two. At first glance, this finding seems at odds with Experiment 1′s post-hoc results showing similar average production times for sequences with one, two, and four different CVs. It is important to note that all of the trials that required reattaching in Experiment 2 were sequences with one CV and two tones. If reattaching indeed reliably slowed production times, this feature of the design could have explained the slower average production times for sequences with one CV.

The second analysis sought to confirm that reattaching reliably slowed speech rate, and whether the effect interacted with reattaching frequency. In this analysis, reattaching (yes vs. no), repetition pattern (ABBA vs. ABAB), and their interaction were entered as fixed effects. Subjects and items were entered as random slopes with related random slopes. Recall that sequences in conditions with one CV and two tones require reattaching, whereas the rest do not. We found a main effect of reattaching ($\chi^2(1) = 19.77$, $p < .001$), but it did not interact with repetition pattern ($\chi^2(1) = 0.01$, $p = .93$). Specifically, the average production time was 328 ms when participants had to reattach tone to CV for every word, versus 306 ms when reattaching was not required (i.e., 22.5 ms difference; $t = 3.38$, $p = .001$). Similarly, the average production time was 322 ms when participants had to reattach for every other word, versus 300 ms when reattaching was not required (i.e., 21.8 ms difference; t = -3.00, $p = .004$). The main effect of repetition pattern was still not significant ($\chi^2(1) = 1.38$, $p = .24$). Note that most excluded trials were from the reattaching condition (30.8 % out of all excluded trials), which further suggested that reattachment was difficult for participants so that they produced more errors or disfluencies. Within the conditions that did not require reattachment, the one CV-one tone condition had the smallest proportion of excluded trials (19.3 %), while the two CVs-one tone and two CVs-two tones conditions had similar proportions of excluded trials (25.4 % and 24.5 %, respectively). This difference might explain why the one CV-one tone trials were produced more slowly than trials in the other two conditions (see Fig. 3), that there seemed to be a speed-accuracy trade off. But most critically, reattaching trials elicited more errors/disfluencies and slower production speed.

Discussion

In Experiment 2, we found evidence supporting the reattachment hypothesis. First, the CV repetition pattern did not reliably predict speech rate. Second, Mandarin speakers are reliably slower at producing sequences in which a CV is paired with more than one tone, potentially due to having to detach a tone and reattach another one to the CV. However, we found that engaging in reattaching for every word versus every other word made minimal difference to production times. Before discussing these results in detail, we conducted Experiment 3 to address several limitations in Experiment 2.

Experiment 3: Validation of the reattachment hypothesis

One limitation of Experiment 2 was that we did not have an equal number of reattached versus not-reattached trials. In addition, trials requiring reattachment always include one CV and two tones, whereas those not requiring reattachment may include one or two CVs, and one or two tones (so that we could examine again the effects of repetition pattern). In Experiment 3, we addressed the above confounds by only including two-CV-two-tone sequences with equal number of reattaching and not-reattaching trials.

Method

Participants

Forty undergraduates from the University of California San Diego participated in the experiment in exchange for course credit. Ten were excluded from our analyses because they either did not produce more than 50 % usable data (based on the same exclusion criteria in Experiments 1 and 2), had extensive prior knowledge of tonal languages other than Mandarin (e.g., Cantonese), or could not read simplified Chinese characters fluently. All 30 remaining participants indicated that they were native Mandarin Chinese speakers with little to no exposure to other tonal languages, and that they moved to the US after the age of 15.

Materials, design, and procedure

The goal of Experiment 3 was to test the reattachment hypothesis with a more balanced design that better targeted the reattachment hypothesis. We used the same word sequence repetition task as Experiments 1 and 2, with the logic that slower speech rate indicates greater encoding difficulty. In this experiment, all sequences included two CVs and two tones. When CV and tone repeated in the same pattern (ABAB or ABBA), reattachment was not needed; when CV and tone repeat in different patterns (one was ABAB and the other was ABBA), reattachment was needed. As in Experiment 2, the materials were constructed using the nine phoneme sets in Table 4, and each set included all possible eight CV-tone-repetition pattern combinations, so that each participant completed 72 trials. Table 7 shows the eight combinations from Phoneme Set 1. Taking this set as an example, “巴笛拔低” (*ba1 di2 ba2 di1*) and “巴笛低拔” (*ba1 di2 di1 ba2*) were the two conditions that required each CV to detach from a tone and reattach to another within the sequence, with the former reattaching less frequently than the latter condition (i.e., reattaching every other word versus every word). “巴笛巴笛” (*ba1 di2 ba1 di2*) and “巴笛笛巴” (*ba1 di2 di2 ba1*) were the two conditions not requiring CV-tone reattachment. Participants repeated the same tonal syllable every other word in the former but repeated the same tonal syllable twice consecutively in the latter condition. The procedure was exactly the same as Experiment 1, and so was administered in person.

Coding and data analysis

Similar to Experiments 1 and 2, we manually counted the number of syllables produced within the eight second production period, excluding any production that overlapped with the beginning or terminal beep signal of each trial. Based on our goal to indirectly infer the proximate units of speech using speech rate when a sequence was correctly produced, we excluded 296 (13.70 %) trials using the same exclusion criteria as Experiments 1 and 2, resulting in 1,864 (86.30 %) analyzable trials. Again, two counters (native Mandarin speakers) were each responsible for counting half of the trials and cross-checking the other half that the other counter initially counted. Of the analyzable trials, only 18 trials led to inter-counter disagreements, which were all resolved with a third round of counting. The syllable counts were then transformed into average production times per syllable by dividing 8,000 ms by the syllable count.

We used a linear mixed-effects model to analyze the effects of the CV repetition pattern, reattachment, and their interactions on average production times (tone repetition could be determined based the levels of these two factors). We used the maximal random effects structure, such that the resulting model included all the fixed effects as by-subjects random slopes and no by-item random slopes.

Results

Table 9 and Fig. 4 summarize the average production times for each of the four conditions. We found main effects for both CV repetition pattern and reattachment. Sequences requiring reattachment were produced 14 ms more slowly than those not requiring reattachment (*M*

Table 9
Example trials from Phoneme set 1, average production times by subject (standard errors in parentheses), and error rates by subject (standard errors in parentheses) of the 4 conditions in experiment 3.

Pinyin	Characters	CV repetition	Tone repetition	Reattaching	Average production time	Error rate (%)
ba1 di2 ba1 di2	巴笛巴笛	ABAB	ABAB	No	285 (11)	7.6 (1.6)
ba1 di2 ba2 di1	巴笛拔低	ABAB	ABBA	Yes	295 (12)	14.8 (2.3)
ba1 di2 di2 ba1	巴笛笛巴	ABBA	ABBA	No	273 (11)	9.8 (2.0)
ba1 di2 di1 ba2	巴笛低拔	ABBA	ABAB	Yes	291 (12)	22.6 (2.5)

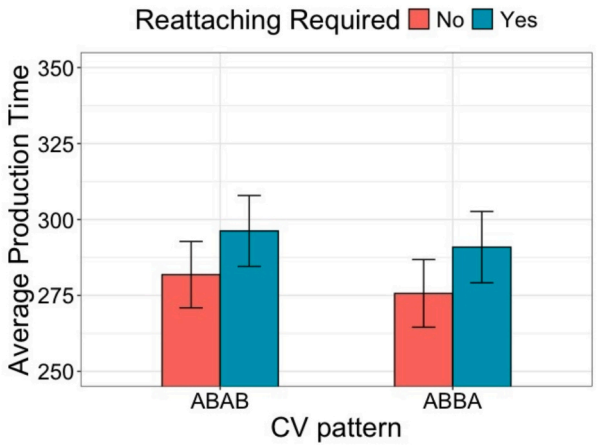

Fig. 4. Average production times for the two CV repetition patterns with vs. without reattachment in experiment 3 note. Error bars represent standard errors.

= 293 ms vs. 279 ms; $\chi^2(1) = 27.82$, $p < .001$); CVs repeated in an ABAB pattern were produced 8 ms more slowly than those in an ABBA pattern (M = 290 ms vs. 282 ms; $\chi^2(1) = 12.15$, $p < .001$). The interaction between these two factors was marginally significant ($\chi^2(1) = 3.82$, $p = .051$). Further analyses suggested that the reattachment effects were significant in both CV repetition patterns ($ps < .01$), but was larger when the CV repeated in an ABBA pattern (*M difference* = 18 ms vs. 10 ms). This suggests that different from in Experiment 2, the need to immediately repeat the same CV with two different tones perhaps elicited greater difficulty. Lastly, more reattachment trials were excluded than those not requiring reattachment (M = 19.7 % vs. 9.7 %; $\chi^2(1) = 26.34$, $p < .001$), suggesting the reattachment costs in production duration were not a result of speed-accuracy trade-off. CVs repeated in an ABBA pattern were excluded more than those repeated in an ABAB pattern (M = 16.2 % vs. 11.2 %; $\chi^2(1) = 4.61$, $p = .032$), a speed-accuracy trade-off in CV repetition pattern (main effect). There was no significant interaction between CV pattern and reattachment in the analysis of trial exclusion rates ($\chi^2(1) < 1$).

Discussion

The results of Experiment 3 further supported our reattachment hypothesis. With a better controlled design (all trials have two CVs and two tones, equal number of reattaching and non-reattaching trials), reattachment still led to significant costs, reliably slowing production and eliciting more invalid trials. In addition, reattachment costs in this experiment were larger when the CV repeated in ABBA (reattaching required every syllable) than in ABAB (reattaching occurs every other CV), though the relevant interaction term was only marginally significant ($p = .051$). While speech rate showed a main effect of CV repetition pattern, it might be a speed-accuracy trade-off. Of course, after controlling the number of unique CVs and tones and the number of trials in each condition, a confound remains: The reattaching trials included four unique tonal syllables, and the non-reattaching trials only included two unique tonal syllables (see Tables 7 and 8). However, Experiments 1 and 2 showed that the number of unique syllables did not predict production difficulty. For example, in Experiment 2, *ba2 ba1 ba1 ba2* (two unique tonal syllables with reattachment) elicited significant slower speech rate than *ba2 di1 di1 ba2* (two unique tonal syllables no reattachment); in Experiment 1, *ba2 bi2 ba2 bi2* (two unique tonal syllables) elicited similar speech rate compared to *ba2 ba2 ba2 ba2* (only one unique tonal syllable). Therefore, although it is impossible to control everything in a single experiment, the three experiments jointly suggest that processing costs in the present study are driven by tone-syllable reattaching rather than other factors.

General discussion

In this study, we report three experiments that investigated Mandarin phonological production using a word sequence speeded repeated production task. We measured the average production times per syllable in four-word sequences containing phonological units of various repetition patterns and examined how phonological units are encoded in Mandarin, as well as how they interacted with each other in production.

Experiment 1 explored whether consonants, vowels, and tones function as independent speech units in phonological encoding in Mandarin, or if some combinations of them are selected and processed together as a unit. The planned analysis revealed a major cross-linguistic difference between Mandarin (as investigated here) and English (as reported in the literature): In Mandarin, neither the number of distinct units in a sequence nor their repetition frequency reliably predicted speech rate. In contrast, previous English experiments reported that speakers were generally faster at producing sequences with fewer distinct units and a higher repetition frequency, compared to sequences with more distinct units and a lower repetition frequency (Sevald & Dell, 1994 and Sevald et al., 1995). This suggests that Mandarin consonants, vowels, and tones may not be completely independent units, or they may not show qualitatively similar patterns of repetition benefits as in previous English data.

Given that repetition frequency did not reliably predict speech rate, we conducted two sets of post-hoc analyses to explore the idea that it is perhaps the quality (e.g., tone-syllable co-varying) rather than the quantity of repetition (e.g., the number of unique tones) across units that affects speech rate. That is, even though the frequency with which each phonological unit repeats does not necessarily affect production difficulty, whether different units share the same or different repetition patterns might. Indeed, we found evidence based on exploratory analyses that production times were faster when tone shared the same pattern with another phonological unit (C or V alone, or both of them), compared to when it did not, suggesting that tone may be attached to another phonological unit. Additionally, when tone shared the same pattern as both C and V, it did not lead to additive benefit compared to sharing with C or V alone, suggesting that C and V are unlikely to be represented independent of each other either.

Experiment 2 sought to confirm two key insights from Experiment 1. The first was whether CVs are represented together as single units in Mandarin, separate from tone. The second was whether there is a process in phonological encoding that involves the pairing of selected CVs and tones, such that changes in the set of CV-tone pairings in a sequence leads to production difficulty. Indeed, the results once again showed that the number of unique CVs or tones in a sequence did not affect speech

rate; but when a given CV was associated with more than one tone (i.e., requiring the production system to detach the tone from the CV and attach to a new tone), speech rate was slow. We coined this *the reattachment hypothesis*, which was further supported by results in Experiment 3: After controlling the number of CVs, the number of tones (all trials were two CVs-two tones), and the proportion of trials requiring reattachment, we still showed robust reattachment costs.

The atonal syllable as the proximate unit in phonological encoding

Our results add to the evidence of cross-linguistic differences in proximate units suggested by form preparation (see Meyer 1990, 1991; Roelofs & Meyer, 1998 for atonal Indo-European languages, Chen et al., 2002 for Mandarin, and O'Seaghdha et al., 2010 for direct cross-linguistic comparisons) and masked priming paradigms (Chen, 2003; Chen et al., 2016; Forster & Davis, 1991; Verdonschot et al., 2013). Specifically, our findings suggest that C and V together form phonological units that are separate from tones and selected at an early stage before tones are integrated, as the speech rate advantage was not additive when tone shared the repetition with both C and V compared to with C or V alone. This supported the secondary role of individual segments in Mandarin production, and is compatible with O'Seaghdha et al.'s (2010) notion that the proximate unit (i.e., the first selectable phonological unit below the word level) in Mandarin is the atonal syllable. We speculate that Mandarin speakers represent all possible CV combinations in the language in a syllabary, selecting the CV(s) corresponding to each word before passing the syllable(s) on for further phonological encoding (including integrating with tone(s)) and determining phonetic details).

Note that the results rule out the possibility that Mandarin speakers plan speech through retrieval of fully specified phonetic forms, which would predict faster production for repeated syllable-tone combinations (e.g., ba2 ba2 ba2 ba2) than for sequences with different syllables (e.g., ba2 di2 di2 ba2 or ba2 di1 di1 ba2). However, this was not the case (see results in Experiment 2). Instead, across all three experiments, speech rate slowed specifically when the same CV appeared with different tones, suggesting that tone is integrated during phonological encoding in a way that disrupts the reuse of syllable frames. This supports a model of dynamic phonological planning rather than one based on retrieving fixed phonetic forms.

Note, however, that we did not reliably observe benefits similar to those reported in previous form preparation paradigms. Recall that it is commonly found that speakers benefit from knowing in advance the onset segment of the target response in English and Dutch (Meyer, 1990, 1991; Roelofs & Meyer, 1998), or the initial syllable (atonal CV) in Mandarin (Chen et al., 2002; O'Seaghdha et al., 2010), and thus produce the target quicker in those conditions, compared to conditions where little to no information is available for advance planning or activation. If those form preparation benefits were directly transferable to our paradigm, we might have expected to see faster speech rates for sequences that contained fewer unique phonological units (e.g., fewer unique CVs), as they supposedly required less advance planning. On the contrary, we found that repeatedly producing the same CV and tone was not necessarily faster than producing two or more distinct CVs in a sequence. Also, producing one CV that was paired with different tones in a sequence was much slower than producing two CVs that were paired with the same tones.

However, there are many factors that could potentially affect whether form preparation benefits are observed or not. In particular, a long preparation period may obscure any subtle preparation benefits, which may have occurred in our experiments. With the eight-second silent preparation period, speakers may have had more than enough time to complete any advance planning possible to produce the target responses in our experiments, which is in stark contrast to having no preparation period between the presentation of the prompt and the production of the target in the form preparation paradigm. If the preparation period indeed absorbed any form preparation benefits, then we may only be able to observe significant production difficulties that advance planning may not effectively overcome, such as reattaching CVs and tones on-the-fly. Moreover, the form preparation paradigm measures onset latency in a one-off manner per trial, which makes it sensitive to speech preparation. On the other hand, our paradigm measures average production time, which includes both preparation and execution time, potentially obscuring some preparation benefits.

While the difference between our results and those in form preparation studies in Mandarin may be due to task differences, we suggest that the difference between our findings and those in Sevald and Dell (1994) are due to language differences. Sevald and Dell (1994) showed clear repetition benefits in the same speeded repeated production paradigm, but they also showed some inhibitory effects and suggested that the inhibition might be due to competition between phonologically similar but not identical candidates. In Mandarin production, however, segment differences seem not to be the most critical issue, probably because CVs serve as whole units; instead, CV-tone pairing seems critical, probably because tone is differently represented compared to segments (i.e., the reattachment issue, see our discussion below). This may obscure some preparation benefits in speeded repeated production. Future research that directly compares the two paradigms may better reveal why CV preparation benefits were not observed in speeded repeated production.

Tone-to-syllable integration through detaching and reattaching

Despite the limitations brought about by a long preparation period, the speeded repeated production paradigm shed light on the interaction of syllables and tones in online processing through measuring a relatively long stretch of execution time (i.e., eight seconds). Specifically, it provided novel insights about the mechanism of how tones are integrated with syllables in continuous speech, namely the reattachment hypothesis. This hypothesis is somewhat similar to the phonological competition account of inhibitory effects in producing beginning-related words in Sevald and Dell (1994), but these two accounts are not identical. The former is about assignment of a component of the metrical frame after the encoding of proximate units, being specific to the atonal syllable and the tone, whereas the latter is about sequentially-cued activation of any competing segments. If the tone-reattachment costs were a result of sequential cued competition, other beginning-related trials should elicit interference. For example, when English speakers repeatedly produce *pick pun pick pun* or *pick pin pick pin* (initial C or C + V stay the same), competition occurs between ending segments like *−ck* vs. *−n* or *−ick* vs. *−un,* leading to processing costs in both cases. In contrast, when Mandarin speakers repeatedly produce *ba1 bi1 ba1 bi1 or ba1 bi2 ba1 bi2* (C + tone stay the same and only the V changes or only C stays the same), competition does not occur between ending segments *a* vs. *i or a1* vs. *i2,* and no reattachment is required between CV and tone, so that no processing difficulties are elicited. Therefore, the tone-reattachment costs could not be explained by a general sequential cued competition account. In addition, tone in Mandarin and segments in English are encoded at different stages in production (see below in the discussion about the modified Levelt-type production model). Therefore, reattachment and sequentially cued competition are different processes. While reattachment costs were due to an additional process (where an earlier tone-to-syllable pairing is dissolved), sequentially cued competition in general is due to activation of previously produced components. The production consequences of reattaching would have been difficult, if not impossible, to observe in paradigms such as form preparation and masked priming, which do not offer the opportunities for tones to dynamically attach, detach, and reattach to syllables in close succession. In sum, the comparison of paradigms highlights the value of using a diverse set of tasks to investigate an issue.

If Mandarin CVs and tones are represented separately but still

depend on each other during encoding, is tone processed relatively early (at the same time as segments) or late (after segmental content has been selected and serially organized)? The best evidence for early encoding accounts comes from the comparison of error types and frequencies between segments and tones (e.g., Wan & Jaeger, 1998; Alderete et al., 2019). Some evidence suggests that tones are retrieved earlier than syllables – in primed picture naming, effects of tonal relatedness emerged in an earlier ERP time window than syllabic relatedness, though tone-to-syllable integration (i.e., tone*syllabic relatedness interaction) occurred in a later ERP time window than tone retrieval (i.e., two-stage tonal encoding, Chen & Zhang, 2025). However, due to the nature of the paradigm, the majority of our data came from error-free production without time-sensitive ERP data, and the small number of errors we obtained did not offer enough power to conduct systematic comparisons between error types that may reveal time course of tone vs. segment encoding. Hence, we did not find strong and direct evidence against tones being selected at the same time as segments. Therefore, future research is needed to further investigate this issue.

However, the process of integrating CVs and tones highlighted by the reattachment hypothesis points towards tones being involved late during encoding. The crucial piece of evidence is that the production difficulty associated with reattaching is anchored on CVs and not on tones. That is, what determines whether a sequence is difficult is whether each CV is associated with multiple tones, but not whether each tone is associated with multiple CVs. In fact, a single tone being associated with multiple distinct CVs did not lead to production difficulties. These observations illustrate that CVs may have a privileged status in phonological encoding, such that they are selected first and later integrated with tones. If tones had been selected before CVs or simultaneously involved in early encoding, we might expect active competition in selection between tones such that producing the same tone with multiple CVs in a sequence would have led to similar production difficulties.

Our results are most compatible with modified Levelt-type production model that assumes tone to be part of a metrical frame (e.g., Chen et al., 2002; O'Seaghdha et al., 2010; Chen & Chen; 2013). As introduced earlier, the model proposes that while atonal CV representations are selected first in word form encoding, their tonal counterparts are penultimate representations that precede actual articulation. In between those two steps lie the processes of linearization and tone value assignment, during which selected CVs are slotted into structural frames with diacritics indicating the tone values inherent to the frames (i.e., unit-to-frame association). When syllable-tone reattachment is needed, speakers have to repeatedly (in this task) detach a selected CV from a frame and reattach it to another, leading to processing costs.

An intriguing question concerns the nature of the process of attaching and, more distinctively, detaching tones. Based on these experiments, we can only speculate regarding specifics of such hypothetical processes. The experiments here provide quite strong evidence (converging with evidence in the literature) that in Mandarin production, something like syllable chunks operate, given that we found no evidence of any benefit or penalty to production for repetition of C or V alone. Attaching a tone may happen in terms of insertion into a structural slot that itself carries the instructions for tone production (with different slots for each tone that can be recursively deployed, just like other structural representations). If so, and if that same syllable chunk must subsequently be produced with a different tone, then that same chunk must be removed from the slot that supported the articulation of the first tone so that it can be inserted into a slot that supports articulation of the second tone. In contrast, if the same syllable chunk is produced with the same tone, even non-contiguously (i.e., even when another syllable chunk with its own tone intervenes), then such detachment of the chunk from the tone slot will not be needed within the 8-second production sequence, permitting production to be more efficient.

While the speeded repeated production task in our study revealed a robust reattachment penalty, such a penalty may not be as apparent in everyday speech. The task strongly encourages the re-use of planning units, allowing us to observe how intensive short-term and repeated detachment of one tone from a particular syllable chunk and subsequent attachment of another tone to that same syllable chunk affects speech. In contrast, real-life speech production involves much fewer repetitions compared to those in designed tongue twister experiments, reducing the need to detach and re-attach different tones to the same syllable chunk in a short sequence. As a result, the opportunity to observe a reattachment penalty is limited. However, an intriguing exception highlights this effect—a well-known Chinese story consisting of 91 syllables, all pronounced as *shi* with different tones (see Appendix). This story is notoriously difficult to read aloud, even without considering comprehension, as readers constantly detach and reattach different tones to the *shi* syllable, illustrating the challenges imposed by reattachment.

For nontonal languages, the tone detaching and reattaching processes may be analogous to using different pitch for the same word. For example, while the rising pitch may indicate criticism or doubt (e.g., beer?), the falling pitch may indicate naming (e.g., beer!), so that the same word with different pitch can reliably express different intentions (Hellbernd & Sammler, 2016). This raises the interesting possibility that reattachment effects, as observed in this study, may also manifest in nontonal languages. Specifically, if participants are asked to repeatedly produce sequences of words, sequences requiring pitch-word reattachment (e.g., beer? beer! wine? wine!) may be produced more slowly than those that do not (e.g., beer? wine! beer? wine!), offering a potential avenue for future investigation.

Limitations and future directions

A question still remains and warrants further investigation in future research: If reattaching is costly, why is there no difference between reattaching more often versus less in Experiment 2? Specifically, the average production time for reattaching every word (ABAB pattern, e.g., "笛低笛低" *di2 di1 di2 di1*) was comparable to that for reattaching every other word (ABBA pattern, e.g., "拔巴巴拔" *ba2 ba1 ba1 ba2*). One possibility is that Mandarin speakers use the "disyllabic chunking strategy". Most of words are disyllabic in Chinese (about 72 %; Lexicon of Common Words in Contemporary Chinese Research Team, 2008), so that participants treated ABAB as repeating the same "disyllabic word" twice, whereas ABBA was treated as producing two different "disyllabic words". Producing more distinct "words" can be more costly than producing a single "word", and this cost may offset the benefits from reattaching less frequently in certain contexts. Accordingly, there was not a clear difference between ABAB and ABBA CV patterns (main effect) in any experiment. This chunking strategy may also explain why producing the same syllable with only one tone did not elicit the fastest production speed (both CV and tone repeated in AAAA) in Experiment 2. Participants may treat one CV-one tone sequences as repeating one monosyllabic word four times, leading to more repetition than producing two disyllabic words and some processing costs, or treating it as two disyllabic words (AA|AA), which is similar to AB|AB.

Conclusion

The current study reported corroborating evidence for the notion that atonal syllables function as proximate units in Mandarin Chinese production. More importantly, based on data showing that speakers were slower to produce sequences that required reattaching between CVs and tones, we suggest that tone is represented differently from phonological content, specifically, atonal syllables. Tone appears to be represented as part of a metrical frame, and its integration with the atonal syllable occurs relatively late in phonological processing, supporting late encoding accounts.

CRediT authorship contribution statement

Chuchu Li: Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Sin Hang Lau:** Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Victor S. Ferreira:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Already shared on OSF

References

Alderete, J., & Davies, M. (2019). Investigating perceptual biases, data reliability, and data discovery in a methodology for collecting speech errors from audio recordings. *Language and Speech, 62*(2), 281–317.

Alderete, J., Chan, Q., & Yeung, H. H. (2019). Tone slips in Cantonese: Evidence for early phonological encoding. *Cognition, 191*, Article 103952.

Alderete, J., & O'Séaghdha, P. G. (2022). Language generality in phonological encoding: Moving beyond Indo-European languages. *Language and Linguistics Compass, 16*(7), Article e12469.

Bates, D., Mäechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version, 1*, 7.

Chen, J. Y. (1999). The representation and processing of tone in Mandarin Chinese: Evidence from slips of the tongue. *Applied psycholinguistics, 20*(2), 289–301.

Chen, J. Y., Chen, T. M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language, 46*(4), 751–781.

Chen, J. Y., Lin, W. C., & Ferrand, L. (2003). Masked priming of the syllable in Mandarin Chinese speech production. *Chinese Journal of Psychology, 45*(1), 107–120.

Chen, J. Y., O'Séaghdha, P. G., & Chen, T. M. (2016). The primacy of abstract syllables in Chinese word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(5), 825.

Chen, X., & Zhang, C. (2025). Setting the "tone" first and then integrating it into the syllable: An EEG investigation of the time course of lexical tone and syllable encoding in Mandarin word production. *Journal of Memory and Language, 140*, Article 104575.

Forster, K. I., & Davis, C. (1991). The density constraint on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language, 30*(1), 1–25.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language*, 27–52.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498.

Halle, M., & Stevens, K. N. (1971). A note on laryngeal features. *MIT Quarterly Progress Report, 101*, 198–212.

Kember, H., Croot, K., & Patrick, E. (2015). Phonological encoding in Mandarin Chinese: Evidence from tongue twisters. *Language and Speech, 58*(4), 417–440.

Leben, W. (1978). The Representation of Tone. In V. Fromkin (Ed.), *Tone: A Linguistic Survey* (pp. 177–220). New York: Academic Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1–38.

Lexicon of Common Words in Contemporary Chinese Research Team. (2008). Lexicon of common words in contemporary Chinese. *Beijing: Commercial Press*.

Lin, C. Y., Wang, M., & Shu, H. (2013). The processing of lexical tones by young Chinese children. *Journal of Child Language, 40*(4), 885–899.

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language, 29*(5), 524–545.

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language, 30*(1), 69–89.

Moser, D. (1991). *Slips of the tongue and pen in Chinese*. (Sino-Platonic Papers, No. 22). Department of Oriental Studies, University of Pennsylvania, Philadelphia, PA.

Odden, D. (1995). Tone: African languages. In J. A. Goldsmith (Ed.), *Handbook of phonological theory* (pp. 444–475). Oxford: Blackwell.

O'Seaghdha, P. G. (2015). Across the great divide: Proximate units at the lexical-phonological interface. *Japanese Psychological Research, 57*(1), 4–21.

O'Seaghdha, P. G., Chen, J. Y., & Chen, T. M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition, 115*(2), 282–302.

O'Seaghdha, P. G., Dell, G. S., Peterson, R. R., & Juliano, C. (1992). Models of form-related priming in comprehension and production. *Connectionist approaches to natural language processing, 1*, 373–408.

Peterson, R. (1991). A phonological competition model of form-related priming effects. *Unpublished doctoral dissertation, University of Rochester, Rochester, New York*.

Peterson, R. R., Dell, G. S., & O'Seaghdha, P. G. (1989). A Conncetionist Model of Form-related Priming Effects. In: *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 11).

R Core Team. (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.

Roelofs, A. (2015). Modeling of phonological encoding in spoken word production: From Germanic languages to Mandarin Chinese and Japanese. *Japanese Psychological Research, 57*(1), 22–37.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition, 64*(3), 249–284.

Roelofs, A., & Meyer, A. S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(4), 922–939.

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition, 53*(2), 91–127.

Shen, J. (1993). Slips of the tongue and the syllable structure of Mandarin Chinese. In S.-.-C. Yau (Ed.), *Essays on the Chinese language by contemporary Chinese scholars* (pp. 139–161). Paris: Centre de Recherche Linguistiques sur L'Asie Orientale. École des Hautes Études en Sciences Sociales.

Verdonschot, R. G., Nakayama, M., Zhang, Q., Tamaoka, K., & Schiller, N. O. (2013). The proximate phonological unit of Chinese-English bilinguals: Proficiency matters. *PLoS One1, 8*(4), Article e61454.

Wan, I. P., & Jaeger, J. (1998). Speech errors and the representation of tone in Mandarin Chinese. *Phonology, 15*(3), 417–461.

Wong, A. W. K., Chiu, H. C., Tsang, Y. K., & Chen, H. C. (2023). Tonal and syllabic encoding in overt Cantonese Chinese speech production: An ERP study. *PLoS One1, 18*(12), Article e0295240.

Yip, M. (1995). Tone in East Asian languages. In J. A. Goldsmith (Ed.), *Handbook of phonological theory* (pp. 476–494). Oxford: Blackwell.