

Phonological Encoding Units in Chinese Spoken Word Production: An ERP Study Using the Implicit Priming Paradigm

Authors: Wang Xuejiao, Zhang Qingfang, Zhang Qingfang

Date: 2019-12-13T00:00:00+00:00

Abstract

In this study, we selected Chinese native speakers with low English proficiency to exclude the potential influence of English phonological encoding units (phonemes) as a second language on Chinese spoken word production processes, and employed event-related potential (ERP) technology to investigate the temporal course of syllable and phoneme effects in Chinese spoken language production. The experiment utilized an implicit priming paradigm, requiring participants to produce the corresponding target word upon seeing the cue word. ERP analysis results revealed that the syllable effect emerged between 100–400 ms after cue word presentation, whereas the phoneme effect appeared between 500–600 ms, with waveforms showing the related condition as more positive than the unrelated condition. This suggests that in the phonological encoding stage following lexical selection, the unit first retrieved by speakers is the syllable, while the time window for the phoneme effect may correspond to the late stage of phonological encoding or the phonetic encoding stage, thereby validating the perspective of the appropriate encoding unit hypothesis.

Full Text

Preamble

Self-Check Report for *Acta Psychologica Sinica* Manuscripts

Please complete the following items and paste them on the first page of your manuscript.

- 1. Please list up to three innovative contributions of this study in the form of “Research Highlights,” with a total word count not exceeding 200 words.**

Acta Psychologica Sinica aims to publish cutting-edge psychological research that is “both scientifically excellent and of particularly broad interest and significance.” If your study only makes minor incremental contributions, does not attempt to open new areas of inquiry, or lacks unique and innovative perspectives—especially if it merely investigates algorithms or techniques without addressing clear psychological questions—its chance of acceptance is low. We recommend submitting such work to other journals.

Response: 1. We employed an implicit priming paradigm with Chinese native speakers of low English proficiency, thereby excluding potential influences of L2 phonological encoding units on Chinese spoken word production. 2. We simultaneously detected both syllable and phoneme effects during Chinese spoken word production, suggesting that speakers first extract syllable information after lemma selection and subsequently decompose syllables into phonemes. 3. Syllable effects occur during phonological encoding, whereas phoneme effects likely emerge during late phonological encoding or phonetic encoding stages.

2. Have you published or submitted any articles using the same dataset as this study? If yes, please attach them for review. (We discourage publishing multiple articles with identical variables from the same dataset or splitting a series of related studies into separate publications.)

Response: No.

3. Studies in management, clinical, personality, and social psychology relying solely on self-report (questionnaire) methods need to examine common method bias. What methods did you use to control for or demonstrate that such bias does not affect the validity of your conclusions? (See <http://journal.psych.ac.cn/xlkxjz/CN/abstract/abstract894.shtml> for relevant literature.) Cross-sectional studies with only self-reports and convenience samples are easy to conduct but typically lack innovative value and have low acceptance probability.

Response: This is an experimental study; this issue does not apply.

4. Clinical experiments aimed at treating diseases are recommended to pre-register before data collection. Other experimental studies are also encouraged to pre-register. Registration requires stating all research hypotheses and their rationales, plus detailed experimental/intervention procedures. Our journal’s pre-registration site is <https://osf.io/>. Pre-registration significantly increases acceptance probability. If your study was pre-registered, please provide the registration number. See <https://osf.io/5awp4/> for the importance of pre-registration.

Response: This study investigates speech production processes in healthy individuals and is not a clinical experiment for disease treatment.

5. Did you report and analyze effect sizes (e.g., Cohen’s d for t-tests, Eta-squared or partial Eta-squared for ANOVA, standardized regression coefficients)? (Many studies mechanically report effect sizes without

necessary analysis or explanation—e.g., are they small/medium/large? What theoretical or applied significance do they have?) Search “effect size calculator” on Google for convenient apps. For explanations of effect sizes, see Chinese: <http://journal.psych.ac.cn/xlkxjz/CN/abstract/abstract1150.shtml>; English: <http://www.uccs.edu/lbecker/effect-size.html>. Did you report 95% CIs for statistical analyses? (e.g., 95% CI for differences, correlation/regression coefficients). See <https://thenewstatistics.com/itns/esci/> for calculations and graphing.

Response: All ANOVA results in this study reported corresponding effect sizes (partial Eta-squared). 95% CIs were also reported; error bars in Figure 3 [Figure 3: see original paper] represent 95% CIs for syllable and phoneme effects across different time windows.

6. Please state planned and actual sample sizes. If they differ, explain why. Underpowered studies due to insufficient sample sizes are common in psychological research. We recommend explaining your sample size determination in the Methods section, based on justified effect sizes and desired power, and reporting the software/program used. See <https://osf.io/5awp4/> for sample size planning.

Response: Planned sample size: 20 participants; actual sample size: 20 participants.

7. For p-values, we require exact values (e.g., $p = 0.034$) rather than reporting ranges (e.g., $p < 0.05$), except for values below 0.001. Does your paper meet this requirement? For Bayes factors, did you report sensitivity to prior distribution assumptions?

Response: Yes, our paper meets this requirement.

8. If you excluded data in statistical analyses, did you report this in the text? What were the reasons? How would results change if included? How were missing data handled? When using scales, did you delete individual items? Why? How would results change if included? Are there unreported measures/variables? Please indicate their locations in the paper.

Response: All data exclusions in statistical analyses are detailed in paragraphs 1-3 of the Results section. No additional data were excluded beyond those reported. There were no missing data.

9. Are unpublished experimental materials, scales, or questionnaires attached for review? If not, explain why. If published, are you willing to share them with other researchers?

Response: All experimental materials used were peer-reviewed; see text for sources.

10. Can your raw data be uploaded for reviewer/editor verification? If not, explain why. If published, are you willing to share data with

other researchers?

Response: Raw data can be uploaded for verification, and we are willing to share them with other researchers.

11. For human or animal subjects, did you obtain IRB approval? If yes, please email a scanned copy to the editorial office. If no, explain why.

Response: [Not answered in original]

12. Did you write a 400-500 word English abstract following the “English Abstract Writing Guidelines” on the editorial website? Has the English title and abstract been reviewed by a native English speaker or professional SCI/SSCI editing service?

Primary Phonological Planning Units in Chinese Spoken Word Production: Evidence from an ERP Study Using the Implicit Priming Paradigm

WANG Xuejiao, ZHANG Qingfang

(Department of Psychology, Renmin University of China, Beijing 100872, China)

Abstract

This study investigated the time course of syllable and phoneme effects in Chinese spoken word production using event-related potentials (ERP), with native Chinese speakers of low English proficiency to exclude potential influences from L2 phonological encoding units. Employing an implicit priming paradigm, participants were asked to produce target words after seeing cue words. ERP analyses revealed syllable effects between 100-400 ms after cue word presentation and phoneme effects between 500-600 ms, with waveforms showing more positive amplitudes in related versus unrelated conditions. These findings indicate that after lemma selection, speakers first extract syllable information during phonological encoding, which is then decomposed into phonemes. The syllable effect occurs during phonological encoding, while the phoneme effect likely emerges during late phonological encoding or phonetic encoding, supporting the proximate unit principle.

Keywords: spoken word production; phonological encoding; proximate unit principle; syllable; phoneme

Speech production involves converting concepts or ideas into linguistic output through three stages: (1) conceptualization, where speakers identify the concept to express; (2) formulation, where appropriate lexical items, grammatical structures, and phonological forms are constructed; and (3) articulation, where selected words are expressed through motor programs (Dell, 1986; Roelofs, 1997;

Levelt, Roelofs, & Meyer, 1999). Formulation, also called lexical access, comprises two stages: lemma selection and word-form encoding. Word-form encoding further subdivides into morphophonological encoding, phonological encoding, and phonetic encoding. During morphophonological encoding, speakers select appropriate morphemes and syntactic features based on lemmas. In phonological encoding, phonological segments and metrical structures are selected and syllabified, linking segments to syllable nodes. During phonetic encoding, syllabary programs are selected to prepare for articulation (Dell, 1986; Roelofs, 1997; Levelt et al., 1999).

The processing unit of phonological encoding has been a central debate in speech production research (Indo-European languages: Dell, 1986; Levelt et al., 1999; Meyer, 1990, 1991; Forster & Davis, 1991; Kinoshita & Woollams, 2002; Malouf & Kinoshita, 2007; Schiller, 2008; Damian & Dumay, 2007, 2009; Damian & Bowers, 2003; Jacobs & Dell, 2014; Chinese: J.-Y. Chen, Lin, & Ferrand, 2003; O'Seaghdha, J.-Y. Chen, & T.-M. Chen, 2010; J.-Y. Chen, O'Seaghdha, & T.-M. Chen, 2016; Qu, Damian, & Kazanina, 2012; J. Wang, Wong, S. Wang, & H.-C. Chen, 2017; Zhang & Damian, 2019). Two major theoretical models propose different views on phonological encoding units. Based on speech error analyses, Dell's (1986) spreading activation model posits that phonological encoding units include phonemes, syllables, syllable constituents, and phonetic features. Syllable components (onset, nucleus, coda) are simultaneously activated and inserted into syllable frames. When a morpheme contains multiple syllables, syllable processing proceeds serially from left to right. Levelt et al.'s (1999) model argues that phonemes are the units, with syllable-internal phonological encoding proceeding incrementally from onset to nucleus to coda. Dell (1986) allows syllables or phonemes as units, whereas Levelt et al. (1999) restrict units to phonemes. The key difference concerns whether syllable-internal phonemes can be simultaneously activated and inserted into frames (Dell) or must be serially inserted from left to right (Levelt et al.).

Most research on Indo-European languages supports phonemes as the primary phonological encoding unit. Both speech error analyses and reaction time studies using various tasks (masked priming, implicit priming, picture-word interference) confirm this view. Speech errors primarily involve phoneme omissions, substitutions, and exchanges, with syllable errors being rare (Dell, 1986; Shattuck-Hufnagel, 1979). Using the implicit priming paradigm, researchers have found faster response times when target words share initial phonemes (Alario, Perre, Castel, & Ziegler, 2007; Damian & Bowers, 2003; Jacobs & Dell, 2014; Meyer, 1991). In adjective-noun naming tasks, naming is faster when adjectives and nouns share initial phonemes (e.g., "green goat") than unrelated conditions (e.g., "green rug") (Damian & Dumay, 2007, 2009). Similar results have been found using picture-word interference (Damian & Martin, 1999; Meyer & Schriefers, 1991) and masked priming paradigms (Forster & Davis, 1991; Kinoshita & Woollams, 2002; Malouf & Kinoshita, 2007; Schiller, 2008).

Recent research on Chinese spoken word production has yielded conflicting views

about the roles of syllables and phonemes. Chinese, a tonal language, differs substantially from alphabetic languages. First, Chinese has far fewer syllables—approximately 400 without tone and 1,200 with tone—compared to about 12,000 syllables in Dutch (Zhang & Yang, 2005; Levelt et al., 1999). Second, Chinese syllable boundaries are clear, with no resyllabification, unlike alphabetic languages where extensive cross-syllable recombination occurs. According to Levelt et al. (1999), resyllabification explains why syllables are not stored in the mental lexicon for alphabetic languages, whereas storing all syllables is more economical for Chinese. Thus, Chinese phonological characteristics suggest syllables rather than phonemes may be the primary encoding unit.

J.-Y. Chen (2000) analyzed a Mandarin speech error corpus and found significantly more syllable exchange errors than phoneme exchange errors, indicating syllables' greater importance. J.-Y. Chen et al. (2002) used the implicit priming paradigm to examine phonological encoding in Mandarin disyllabic word production, finding syllable priming effects but no phoneme priming effects, suggesting toneless syllables function as independent planning units at the phonological level. Similar results were obtained using masked priming (J.-Y. Chen et al., 2003; J.-Y. Chen et al., 2016) and picture-word interference paradigms (Zhang & Yang, 2005). You, Zhang, and Verdonschot (2012) found syllable priming effects in both picture naming and word naming tasks, providing stronger evidence for syllables as phonological encoding units. Yue and Zhang (2015) used picture-word interference with immediate naming, delayed naming, and delayed naming plus articulatory suppression to examine syllable and segment effects at different word-form encoding stages, finding syllable facilitation during phonological encoding and suggesting segments may function during subsequent phonetic encoding.

In contrast, some studies have found segment facilitation effects in Cantonese spoken word production. Wong and H.-C. Chen (2008, 2009) used picture-word interference and found faster picture naming when distractors shared rimes and tones with targets, suggesting segments might be the primary phonological unit (see also Wong, Huang, & H.-C. Chen, 2012). However, Wong, Wang, Ng, and Chen (2016) used picture-word interference with ERP and found Cantonese syllable effects between 500-650 ms. Wong et al. (2012) proposed that Cantonese speech production may differ from Mandarin due to Cantonese' s more complex phonological system: Cantonese has six codas (/p/, /t/, /k/, /m/, /n/, /ng/) compared to Mandarin' s two (/n/, /ng/), with /ng/ also occurring as an onset in Cantonese, and Cantonese has six tones versus Mandarin' s four.

Researchers have also used high-temporal-resolution ERP to examine the time course of syllable and phoneme effects. Qu et al. (2012) used an initial phoneme repetition paradigm (e.g., “yellow box” vs. “green box”) and found phoneme repetition effects between 200-300 ms after picture presentation, suggesting phoneme activation during phonological encoding (see also Yu, C. Mo, & M. Mo, 2014). O' Seaghdha et al. (2013) argued that Qu et al.' s (2012) ERP differences might reflect associative priming between color words and targets sharing ini-

tial phonemes rather than phoneme units per se. Additionally, Qu et al.'s (2012) study was conducted in the UK with highly proficient English-speaking students. Research shows L1 processing is influenced by L2 learning transfer (Jones et al., 2012; Verdonschot et al., 2013). Verdonschot et al. (2013) found initial phoneme priming in Chinese tasks with proficient English-Chinese bilinguals, indicating L2 influences L1 processing. Wang et al. (2017) used picture-picture interference with delayed picture naming and found syllable priming at 200-400 ms and 400-600 ms but no segment effects. Zhang and Damian (2019) used masked priming and, for the first time in Mandarin spoken word production, simultaneously detected syllable effects (300-400 ms) and segment effects (500-600 ms). Based on meta-analytic time courses (Indefrey, 2011), they suggested these effects occurred during phonological and phonetic encoding stages, respectively.

Roelofs' (1997) WEAVER model, built on Indo-European findings, assumes phonemes as phonological encoding units with incremental syllable-internal encoding (Levelt et al., 1999). Based on cross-linguistic differences, O' Seaghdha et al. (2010) proposed the Proximate Units Principle to explain language variation in phonological encoding units (see Figure 1 [Figure 1: see original paper]).

Proximate units are the first phonological encoding units activated after lexical morpheme selection. This principle posits that the initially selected unit varies by language: phonemes in Indo-European languages like English and Dutch, but syllables in Chinese. In Indo-European languages, speakers select phonemes, combine them with metrical information through syllabification, and retrieve syllables from a mental syllabary for articulatory programming. In Chinese, speakers select syllables first, then decompose them into phonemes or segments during phonological encoding, prepare articulatory motor programs (phonetic encoding), and finally produce speech. Roelofs (2014) validated the Proximate Units Principle through computer simulations incorporating it into WEAVER's parameters.

[Figure 1: see original paper]

Despite this, debate continues regarding syllable and phoneme roles in Chinese speech production, with inconsistencies between behavioral and ERP findings (Wong & H.-C. Chen, 2008, 2009; Wong et al., 2016) and across ERP studies (Qu et al., 2012; Wang et al., 2017; Zhang & Damian, 2019). Studies finding phoneme effects (Qu et al., 2012; Yu et al., 2014) did not compare them with syllable effects and may have been influenced by L2 transfer. Previous ERP studies used picture naming, which involves concept preparation, lemma selection, phonological encoding, phonetic encoding, and articulation—processes that interact (e.g., concept preparation affects lexical selection and phonological encoding). In ERP studies, syllable activation is typically stronger than phoneme activation, making phoneme effects difficult to detect (Wang et al., 2017), suggesting the need for more phoneme-sensitive tasks.

The implicit priming paradigm has been widely used to investigate phonological encoding. Participants learn several word pairs (e.g., “seal-sunset, memory-

daughter-in-law, emperor-happy event, shell-cell”), where the second words share the first syllable (homogeneous condition). Control sets use phonologically unrelated word pairs (heterogeneous condition). During testing, the first word (cue) is presented, and participants produce the second word (target) (Zhang, 2008). Implicit priming produces preparation effects when targets share phonological properties. In Indo-European languages, this paradigm has sensitively detected single-phoneme preparation effects (Meyer, 1990, 1991; Damian & Bowers, 2003), demonstrating its sensitivity to phonological units. In each test block, participants respond to only four cue words that have been thoroughly learned, allowing rapid target retrieval from working memory without conceptual processing (Meyer, 1990, 1991; J.-Y. Chen et al., 2002; Cholin, Schiller, & Levelt, 2004). We hypothesized that implicit priming might be more sensitive to phoneme effects than picture naming.

This study combined implicit priming with ERP to examine the time course of phoneme and syllable effects, using participants with low L2 proficiency who had no daily English exposure to eliminate L2 influence. Based on the Proximate Units Principle, we predicted earlier syllable than phoneme effects, with syllable effects during phonological encoding and phoneme effects during late phonological or phonetic encoding.

Method

Participants. Twenty adults (ages 22-46, 16 male) participated, with educational backgrounds: 2 junior high, 3 senior high, 6 college, and 9 bachelor’ s degree. All had no daily English exposure and began learning English in junior high. Before the experiment, they completed a self-rated English proficiency questionnaire (0-7 scale: 0 = none, 7 = excellent). Mean ratings were: reading 1.53 (SD = 0.94, range 0-4), writing 1.37 (SD = 1.18, range 0-3), speaking 1.42 (SD = 1.35, range 0-3), and listening 1.32 (SD = 0.98, range 0-3). All were native Mandarin speakers with standard pronunciation, normal or corrected vision, no history of brain disease, and provided informed consent before receiving compensation.

Materials. Thirty-six semantically related but orthographically and phonologically unrelated two-character word pairs were used (e.g., 神仙-凡人, “deity-mortals,” where 神仙 is the cue and 凡人 the target). Thirty-two pairs were experimental items and four were practice items. The 32 pairs were divided into four syllable-sharing sets and four initial-phoneme-sharing sets, each containing four pairs, creating four syllable-homogeneous and four phoneme-homogeneous conditions. By selecting one pair from each of the four syllable-homogeneous sets, we created four sets of four pairs with no phonological relationship (syllable-heterogeneous condition). The same method created four phoneme-heterogeneous sets (see Appendix 1). Thus, syllable-homogeneous and -heterogeneous conditions, and phoneme-homogeneous and -heterogeneous conditions, used identical word pairs, differing only in phonological relationships among targets.

To match semantic relatedness between cue-target pairs across syllable and phoneme conditions, 21 Renmin University students rated relatedness on a 5-point scale (1 = unrelated, 5 = highly related). Syllable materials ($M = 4.02$, $SD = 0.54$) and phoneme materials ($M = 4.12$, $SD = 0.52$) did not differ significantly, $F(1, 20) = 1.97$, $p = 0.176$.

To match semantic relatedness among targets within homogeneous and heterogeneous sets, 20 Renmin University students rated relatedness between target pairs (e.g., “crisis-wall,” “crisis-satellite,” “wall-tail”) on a 5-point scale. For syllable materials, both homogeneous ($M = 1.34$, $SD = 0.29$) and heterogeneous ($M = 1.32$, $SD = 0.35$) sets showed low relatedness with no significant difference, $F(1, 19) = 0.177$, $p = 0.862$. For phoneme materials, homogeneous ($M = 1.28$, $SD = 0.29$) and heterogeneous ($M = 1.27$, $SD = 0.32$) sets also showed no difference, $F(1, 19) = 0.18$, $p = 0.79$. Semantic relatedness among cue words was also low and equivalent across conditions (all p s > 0.10).

Design. The experiment used a 2 (relatedness condition: homogeneous vs. heterogeneous) \times 2 (relatedness type: syllable vs. phoneme) fully repeated-measures design, with both factors within-subjects. Each participant completed four runs of the same procedure, each containing one homogeneous and one heterogeneous block, with block order counterbalanced across participants. In each block, the four word pairs were repeated four times, yielding 16 trials per block. For each repetition and block, cue word order was randomized, with no immediate repetition of the same word. Each participant completed 256 trials (2 relatedness conditions \times 2 relatedness types \times 4 sets \times 4 pairs \times 4 repetitions).

Procedure. The experiment was programmed in E-Prime 2.0, using a PST SRBOX response box, microphone, and computer. Stimuli appeared centrally on a PIII-667 monitor. Responses were recorded via PST SRBOX-connected microphone, with presentation, timing, and response collection controlled by computer. The experimenter recorded response accuracy.

The experiment comprised three phases: learning, testing, and formal experiment. In the learning phase, participants studied 16 word pairs from syllable- or phoneme-sharing conditions plus 4 practice pairs (20 total). Cue and target words appeared onscreen, and participants controlled study duration to memorize all pairs. After learning, the testing phase presented cue words centrally in 30-point Song font, requiring participants to clearly produce the corresponding target word. Correct answers appeared after each response. Each pair was presented three times. Errors were re-learned and re-tested until all pairs were mastered.

In the formal phase, a fixation cross “+” appeared for 500 ms, followed by a 300 ms blank, then the cue word (maximum 3000 ms) in 30-point Song font. Participants were instructed to produce target words as accurately and quickly as possible. The experimenter judged response accuracy after each trial, with a 1500 ms inter-trial interval.

EEG Recording and Analysis. EEG was recorded using a Neuroscan ESI-64 system with 64 Ag/AgCl electrodes arranged in the international 10-20 configuration. The left mastoid served as reference, with ground at forehead center. Vertical EOG was recorded from electrodes above and below the left eye; horizontal EOG from electrodes 1 cm lateral to the outer canthi. Impedance was maintained below 5 k Ω . Signals were amplified with a 0.05-100 Hz bandpass filter during continuous recording at 500 Hz sampling rate.

Offline EEG processing re-referenced to bilateral mastoids (Wang, Bastiaansen, Yang, & Hagoort, 2011). Ocular artifacts were corrected using EEGLAB. Analysis applied a 0.1-30 Hz bandpass filter. Epochs were segmented from -200 to +600 ms relative to target onset, with the 200 ms pre-stimulus baseline used for correction. Trials with artifacts exceeding ± 100 V were rejected.

Six regions of interest (ROIs) were defined along sagittal and coronal axes: left anterior (F5, F7, FC5), middle anterior (Fz, FCz, Cz), right anterior (F6, F8, FC6), left posterior (P5, P7, CP5), middle posterior (CPz, Pz, POz), and right posterior (P6, P8, CP6). Mean amplitudes for each time window served as dependent variables in 2 (relatedness condition: homogeneous vs. heterogeneous) \times 2 (relatedness type: syllable vs. phoneme) \times 2 (region: anterior vs. posterior) \times 3 (hemisphere: left, middle, right) repeated-measures ANOVAs. Greenhouse-Geisser correction was applied when sphericity was violated.

Two participants were excluded due to excessive artifacts, leaving 18 participants with over 40 valid trials per condition.

Results

Behavioral Results. Data were excluded for naming errors (0.52%), non-responses or vocalizations like “um” (1.19%), and outliers beyond 2.5 SD from the mean (5.41%). ANOVAs showed no significant differences between homogeneous ($M = 844$ ms, $SD = 87$) and heterogeneous ($M = 850$ ms, $SD = 90$) conditions for syllable type, $F(1, 17) = 0.737$, $p = 0.406$, $F(1, 15) = 0.113$, $p = 0.741$. For phoneme type, homogeneous ($M = 791$ ms, $SD = 89$) and heterogeneous ($M = 788$ ms, $SD = 101$) conditions also did not differ, $F(1, 17) = 0.229$, $p = 0.638$, $F(1, 15) = 0.219$, $p = 0.647$.

ERP Results. Trials with RTs < 600 ms (5.2%) or > 1500 ms (2.7%) were excluded; only correct trials were analyzed.

Based on previous research (Qu et al., 2012; Zhang & Damian, 2019), six consecutive time windows were analyzed: 0-100 ms, 100-200 ms, 200-300 ms, 300-400 ms, 400-500 ms, and 500-600 ms. Mean amplitudes in each window were analyzed with 2 (relatedness type: syllable vs. phoneme) \times 2 (relatedness condition: homogeneous vs. heterogeneous) \times 2 (region: anterior vs. posterior) \times 3 (hemisphere: left, middle, right) repeated-measures ANOVAs, with Greenhouse-Geisser correction and FDR correction using R's `fdrtool` (see Table 1).

Table 1 ANOVA results for relatedness type, condition, region, and hemisphere across 0-600 ms

Time window (ms)	Condition (1,17)	Type (1,17)	Condition × Type (1,17)	Region (2,34)	Hemisphere (1,17)	Region × Hemisphere (2,34)	Type × Hemisphere (1,17)	Region × Hemisphere (2,34)	Type × Region (1,17)	Type × Hemisphere × Region (1,17)
0-100	7.00*									
100-200	5.34*		4.01↑							
200-300	4.01↑									
300-400	4.21↑		6.31*							
400-500	5.09*									
500-600	6.31*	4.30*	4.49↑							

Note: ↑ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, p-values FDR-corrected.

100-200 ms window. The Relatedness Type × Condition × Region interaction was marginally significant, $F(1, 17) = 4.01$, $p = .061$, $p^2 = 0.191$. Follow-up tests (Table 2) showed significant differences between syllable homogeneous and heterogeneous conditions, with more positive amplitudes for homogeneous conditions (see Figure 2 [Figure 2: see original paper]). Differences reached significance (FDR-corrected) in left anterior, $t(17) = 3.02$, $p = .024$, and right anterior regions, $t(17) = 3.35$, $p = .020$. No phoneme effects were found (all $ps > .462$).

200-300 ms window. No significant Relatedness Type × Condition interactions emerged ($ps > .105$). However, follow-up ROI analyses (Table 2) revealed significant syllable homogeneous vs. heterogeneous differences, with more positive amplitudes for homogeneous conditions (Figure 2). Significant differences (FDR-corrected) appeared in left anterior, $t(17) = 2.55$, $p = .032$, middle anterior, $t(17) = 2.19$, $p = .043$, and right anterior regions, $t(17) = 3.82$, $p = .012$. No phoneme effects were observed ($ps > .230$).

300-400 ms window. The Relatedness Type × Condition interaction was significant, $F(1, 17) = 6.31$, $p = .022$, $p^2 = 0.271$. Follow-up tests (Table 2) showed syllable homogeneous vs. heterogeneous differences with more positive amplitudes for homogeneous conditions (Figure 2). Significant differences (FDR-corrected) appeared in five ROIs: left anterior, $t(17) = 2.38$, $p = .035$; middle anterior, $t(17) = 2.63$, $p = .031$; right anterior, $t(17) = 3.26$, $p = .020$; middle

posterior, $t(17) = 2.28$, $p = .039$; and right posterior, $t(17) = 2.67$, $p = .031$. No phoneme effects were found ($ps > .315$).

400-500 ms window. No significant Relatedness Type \times Condition interactions were found ($ps > .271$).

500-600 ms window. No significant Relatedness Type \times Condition interactions emerged ($ps > .233$). However, follow-up ROI analyses revealed phoneme homogeneous vs. heterogeneous differences with more positive amplitudes for homogeneous conditions (Figure 2). Significant differences (FDR-corrected) appeared in right anterior, $t(17) = 2.71$, $p = .031$, and right posterior regions, $t(17) = 2.43$, $p = .035$. No syllable effects were observed ($ps > .227$).

Table 2 Syllable and phoneme effects across time windows

Time window (ms)	Syllable effect $t(17)$	Phoneme effect $t(17)$
0-100	2.55*	
100-200	2.19*	
200-300	3.82*	
300-400	2.38*	
400-500	2.63*	
500-600	3.26*	2.71*

Note: $\uparrow p < 0.1$, $*p < 0.05$, $**p < 0.01$, p-values FDR-corrected.

Figure 2 Average waveforms for syllable and phoneme effects across six ROIs (boxes indicate significant condition differences)

Figure 3 Mean amplitudes and topographic maps for syllable and phoneme effects across time windows (error bars represent 95% CI)

Discussion

Using the implicit priming paradigm, we examined phonological encoding units and compared the time course of syllable and phoneme retrieval in Chinese speech production. Behavioral results showed no syllable or phoneme effects. ERP analyses revealed syllable effects between 100-400 ms after cue presentation, initially in left and right anterior regions, then middle anterior, and finally spreading to anterior and right middle-posterior regions. Phoneme effects appeared later, between 500-600 ms, localized to right anterior and posterior regions. Both effects showed more positive waveforms for related versus unrelated conditions. These findings indicate that in implicit priming, syllable information is retrieved before phoneme information, consistent with the Proximate Units Principle (O' Seaghdha et al., 2010): after lexical retrieval, speakers first extract syllables during phonological encoding rather than phonemes.

The absence of behavioral syllable effects contrasts with previous studies using the same task (J.-Y. Chen et al., 2002; Zhang, 2008). To verify our materi-

als, six Renmin University graduate students who did not participate in the formal experiment completed the task. All showed syllable priming: syllable-homogeneous ($M = 745$ ms, $SD = 153$) was faster than heterogeneous ($M = 765$ ms, $SD = 157$), $t(5) = 3.10$, $p = .022$, with no phoneme priming ($p > 0.05$), consistent with prior research and confirming material validity. Notably, previous studies used college students with substantial daily English exposure, suggesting that education level and L2 experience may influence L1 speech production. Our participants showed a weak syllable priming trend (homogeneous faster than heterogeneous), whereas graduate students showed ~ 20 ms facilitation, indicating greater sensitivity to syllable information. Additionally, our participants' mean naming latency was ~ 840 ms versus ~ 750 ms for the material validation group. In implicit priming, participants must memorize cue-target associations; weaker memory may have reduced the homogeneous-heterogeneous difference when responses were slower. This suggests that education level and literacy experience may affect speech production mechanisms, warranting further investigation.

Implicit priming studies typically do not match cue-target word frequency or stroke count, focusing instead on whether homogeneous and heterogeneous conditions differ within each phonological condition. This logic has been validated across numerous studies (J.-Y. Chen et al., 2002; Cholin et al., 2004; Damian & Bowers, 2003; Meyer, 1990, 1991; Zhang, 2008). For example, J.-Y. Chen et al. (2002) reported syllable-homogeneous RTs of 658 ms (Experiment 2), 666 ms (Experiment 3), and 662 ms (Experiment 4), versus phoneme-homogeneous RTs of 622 ms (Experiment 5). Zhang (2008) found phoneme-homogeneous RTs of 620 ms versus syllable-homogeneous RTs of 633 ms. These material-dependent RT differences do not affect the detection of syllable or phoneme effects, as comparisons are made within, not across, conditions.

Our ERP results show earlier syllable than phoneme retrieval in implicit priming, consistent with masked priming ERP findings (Zhang & Damian, 2019) where syllable effects occurred at 300-400 ms. Based on meta-analytic time courses of picture naming (Indefrey, 2011), syllable effects were interpreted as occurring during phonological encoding. Wang et al. (2017) found syllable priming at 200-400 ms and 400-600 ms in Cantonese picture-picture interference with delayed naming, interpreting these as phonological and phonetic encoding stages, respectively, but found no phoneme effects. Previous time courses were based on picture naming, which includes concept preparation, lemma selection, phonological encoding, phonetic encoding, and articulation. Implicit priming excludes concept preparation, proceeding from lemma selection to articulation. Thus, phonological encoding occurs relatively earlier in implicit priming than picture naming, consistent with our syllable effects beginning around 100 ms. Syllable effects persisted until 400 ms, spreading from anterior to right middle-posterior regions, indicating that syllable retrieval occurs during phonological encoding.

Phoneme effects appeared later, at 500-600 ms, after syllable effects. According to O' Seaghdha et al.' s (2010) model (Figure 1), after syllable retrieval during phonological encoding, syllables are decomposed into phonemes, which are

then retrieved in parallel (Roelofs, 2015) or serially left-to-right (J.-Y. Chen et al., 2016). Metrical information is also retrieved to create word-form frames, phonemes are inserted, and phonetic encoding prepares articulation. Two interpretations are possible: (1) phoneme effects occur during late phonological encoding, after syllable retrieval; or (2) they occur during phonetic encoding. Zhang and Damian (2019) found segment effects at 500-600 ms in picture naming and, comparing with meta-analytic time courses, suggested they reflected phonetic encoding. Wang et al. (2017) simultaneously examined syllable and phoneme conditions but found only syllable effects. Qu et al. (2012) and Yu et al. (2014) examined only phoneme conditions, preventing comparison of relative time courses. Our study cannot distinguish these interpretations; tasks like immediate naming, delayed naming, and delayed naming with articulatory suppression are needed.

Our phoneme effect time window is later than Qu et al.'s (2012) 200-300 ms and Yu et al.'s (2014) 180-300 ms, which they interpreted as phoneme activation during phonological encoding. In picture naming with ~600 ms latency, 180-300 ms corresponds to lemma selection (Indefrey, 2011), with phonological encoding at 275-400 ms. Qu et al.'s (2012) mean naming RT was 976 ms, yet phoneme effects appeared early. O' Seaghdha et al. (2013) suggested their findings might reflect phonological association rather than phoneme unit activation during encoding—a claim requiring further verification. Task differences may also explain divergent time windows. Additionally, previous ERP studies did not correct for multiple comparisons across ROIs (Qu et al., 2012; Yu et al., 2014; Wong et al., 2016; Wang et al., 2017). Applying FDR correction to their reported p-values renders most non-significant ($p > 0.05$). Our effects remained significant after correction, demonstrating robust phoneme effects in implicit priming.

Both syllable and phoneme effects showed more positive waveforms for homogeneous than heterogeneous conditions, consistent with other ERP studies (Qu et al., 2012; Yu et al., 2014; Wong et al., 2016; Wang et al., 2017; Verdon-schot, Tokimoto, & Miyaoka, 2019), reflecting facilitation when syllables or phonemes overlap. Syllable facilitation at 100-400 ms aligns with classic P2 and P3 time windows (Hackley, Woldorff, & Hillyard, 1990; Donchin & Coles, 1988). P2 is associated with visual feature detection and is sensitive to orthographic, phonological, semantic, and lexical classification tasks (Luck & Hillyard, 1994). P3 amplitude relates to attention (Gray & Burgess, 2004), cognitive closure after decision-making (Desmedt, 1980; Verleger, 1998), or conscious information access (Picton, 1992). P3 comprises P3a (frontal, elicited by unexpected events, reflecting automatic attention) and P3b (parieto-occipital, related to attention, working memory, and complex cognition including language). This suggests syllable homogeneous conditions may elicit greater attention, facilitating speech production. How attention modulates phonological encoding and waveform changes requires further study.

ERP studies in spoken word production typically compare experimental and baseline conditions across time windows, with considerable variation in window

selection. Caution is needed when comparing across studies with different tasks and participants. However, the temporal pattern of effects is comparable across studies. Our finding of earlier syllable than phoneme effects aligns with masked priming results (Zhang & Damian, 2019).

In summary, we are the first to combine implicit priming with ERP, using verbal materials and low-English-proficiency Chinese speakers to exclude L2 influence, while still finding earlier syllable (100-400 ms) than phoneme (500-600 ms) effects. This indicates that after lemma selection in Chinese spoken word production, speakers first extract syllable information before decomposing it into phonemes, supporting the Proximate Units Principle.

References

- Alario, F. X., Perre, L., Castel, C., & Ziegler, J. C. (2007). The role of orthography in speech production revisited. *Cognition*, *102*, 464-482.
- Chen, J.-Y. (2000). Syllable errors from naturalistic slips of the tongue in Mandarin Chinese. *Psychologia: An International Journal of Psychology in the Orient*, *43*, 15-26.
- 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*, 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*, 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*, 825-836.
- Cholin, J., Schiller, N. O., & Levelt, W. J. M. (2004). The preparation of syllables in speech production. *Journal of Memory and Language*, *50*(1), 47-61.
- Damian, M. F., & Bowers, J. S. (2003). Effects of orthography on speech production in a form-preparation paradigm. *Journal of Memory and Language*, *49*, 119-132.
- Damian, M. F., & Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language*, *57*, 195-209.
- Damian, M. F., & Dumay, N. (2009). Exploring phonological encoding through repeated segments. *Language and Cognitive Processes*, *24*, 685-712.
- Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 345-361.

- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283-321.
- Desmedt, J. E. (1980). P300 in serial tasks: An essential post-decision closure mechanism. *Progress in Brain Research*, *54*, 682-686.
- Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of cognitive updating? *The Behavioral and Brain Sciences*, *11*, 357-427.
- 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*, 1-25.
- Gray, J. & Burgess, G. C. (2004). Personality differences in cognitive control? BAS, processing efficiency, and the prefrontal cortex. *Journal of Research in Personality*, *38*, 35-36.
- Hackley, S. A., Woldorff, M., & Hillyard, S. A. (1990). Cross-modal selective attention effects on retinal, myogenic, brainstem and cerebral evoked potentials. *Psychophysiology*, *27*, 195-208.
- Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, *2*, 1-16.
- Jacobs, C. L., & Dell, G. S. (2014). “hotdog”, not “hot”“dog”: The phonological planning of compound words. *Language, Cognition and Neuroscience*, *29*, 512-523.
- Kinoshita, S., & Woollams, A. (2002). The masked onset priming effect in naming: Computation of phonology or speech planning. *Memory and Cognition*, *30*, 237-245.
- Levelt, W. J. M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1-75.
- Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, *31*, 291-308.
- Malouf, T., & Kinoshita, S. (2007). Masked onset priming effect for high-frequency words: Further support for the speech-planning account. *Quarterly Journal of Experimental Psychology*, *60*, 1155-1167.
- Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory & 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*, 69-89.
- Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of

interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 1146-1160.

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, 282-302.

O' Seaghdha, P. G., Chen, J.-Y., & Chen, T.-M. (2013). Close but not proximate: The significance of phonological segments in speaking depends on their functional engagement. *Proceedings of the National Academy of Sciences of the United States of America*, 110, E3.

Parker Jones, O., Green, D. W., Grogan, A., Pliatsikas, C., Filippopolitis, K., Ali, N., . . . Price, C.J. (2012). Where, when and why brain activation differs for bilinguals and monolinguals during picture naming and reading aloud. *Cerebral Cortex*, 22(4), 892-902.

Picton, T. W. (1992). The P300 wave of the human event related potential. *Clinical Neurophysiology*, 9(1), 456-479.

Qu, Q., Damian, M. F. & Kazanina, N. (2012). Sound-sized segments are significant for Mandarin speakers. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 14265-14270.

Qu, Q., Zhang, Q., & Damian, M. F. (2016). Tracking the time course of lexical access in orthographic production: An event-related potential study of word frequency effects in written picture naming. *Brain and Language*, 159, 118-126.

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

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

Schiller, N. O. (2008). The masked onset priming effect in picture naming. *Cognition*, 106, 952-962.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Erlbaum.

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

Verleger, R. (1988). Towards an integration of P3 research with cognitive neuroscience. *Behavioral and Brain Sciences*, 21, 150-154.

Verdonschot, R. G., Tokimoto, S., & Miyaoka, Y. (2019). The fundamental

phonological unit of Japanese word production: An EEG study using the picture-word interference paradigm. *Journal of Neurolinguistics*, 184-193.

Wang, L., Bastiaansen, M., Yang, Y. F., & Hagoort, P. (2011). The influence of information structure on the depth of semantic processing: How focus and pitch accent determine the size of the N400 effect. *Neuropsychologia*, 49(5), 813-820.

Wang, J., Wong, W. K., Wang, S., & Chen, H. C. (2017). Primary phonological planning units in spoken word production are language-specific: Evidence from an ERP study. *Scientific Reports*, 7, 5815-5822.

Wong, A. W. K., & Chen, H.-C. (2008). Processing segmental and prosodic information in Cantonese word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1172-1190.

Wong, A. W. K., & Chen, H.-C. (2009). What are effective phonological units in Cantonese spoken word planning? *Psychonomic Bulletin & Review*, 16, 888-892.

Wong, A. W. K., Huang, J., & Chen, H.-C. (2012). Phonological units in spoken word production: Insights from Cantonese. *PLoS One*, 7(11):e48776.

Wong, A. W. K., Wang, J., Ng, T. Y., & Chen, H.-C. (2016). Syllabic encoding during overt speech production in Cantonese: Evidence from temporal brain responses. *Brain Research*, 1648, 101-109.

You, W., Zhang Q., & Verdonschot, R. G. (2012). Masked syllable priming effects in word and picture naming in Chinese. *PLoS ONE*, 7(10):e46595.

Yu, M., Mo, C. & Mo, L. (2014). The role of phoneme in Mandarin Chinese production: Evidence from ERPs. *PLoS One*, 9(9):e106486.

Yue, Y., & Zhang, Q.F. (2015). Syllable and segments effects in Mandarin Chinese spoken word production. *Acta Psychologica Sinica*, 47(3), 319-328.

Zhang, Q.F., & Yang, Y. (2005). The Phonological Planning Unit in Chinese Monosyllabic Word Production. *Psychological Science (China)*, 28, 374-378.

Zhang, Q. F. (2008). Phonological encoding in monosyllabic and bisyllabic Mandarin word production: Implicit priming paradigm study. *Acta Psychologica Sinica*, 40(3), 253-262.

Zhang, Q., & Damian, M. F. (2019). Syllables constitute proximate units for Mandarin speakers: Electrophysiological evidence from a masked priming task. *Psychophysiology*, 56(6), e13317.

Appendix 1: Experimental Materials

Group	Syllable Homogeneous	Cue-Target Pairs
/wei/	灾祸-危机 (zai1huo4-wei1ji1, disaster-crisis)	院子-围墙 (yuan4zi-wei2qiang2, yard-wall)
	屁股-尾巴 (pi4gu-wei3ba, buttocks-tail)	人造-卫星 (ren2zao4-wei4xing1, artificial-satellite)
/guo/	收获-果实 (shou1huo4-guo3shi2, harvest-fruit)	惩罚-过错 (cheng2fa2-guo4cuo4, punishment-fault)
	烧水-锅炉 (shao1shui3-guo1lu2, boil water-boiler)	熊猫-国宝 (xiong2mao1-guo2bao3, panda-national treasure)
/fan/	神仙-凡人 (shen2xian1-fan2ren2, deity-mortals)	化学-反应 (hua4xue2-fan3ying4, chemistry-reaction)
	食物-饭馆 (shi2wu4-fan4guan3, food-restaurant)	航海-帆船 (hang2hai3-fan1chuan2, sailing-sailboat)
/tong/	地下-通道 (di4xia4-tong1dao4, underground-passage)	师傅-徒弟 (shi1fu-tu2di4, master-apprentice)
	运动-体育 (yun4dong4-ti3yu4, exercise-physical education)	宇宙-太空 (yu3zhou4-tai4kong1, universe-space)
/lv/	辩护-律师 (bian4hu4-lv4shi1, defense-lawyer)	废品-垃圾 (fei4pin3-la1ji1, waste-garbage)
	闪电-雷声 (shan3dian4-lei2sheng1, lightning-thunder)	生日-礼物 (sheng1ri4-li3wu4, birthday-gift)
/dou/	披风-斗篷 (pi1feng1-dou3peng2, cloak-cloak)	和尚-道士 (he2shang4-dao4shi4, monk-Taoist priest)
	榔头-钉子 (lang2tou-ding1zi, hammer-nail)	鸦片-毒品 (ya1pian4-du2pin3, opium-drug)
/ming/	排序-名次 (pai2xu4-ming2ci4, ranking-placement)	帅哥-美女 (shuai4ge1-mei3nv3, handsome guy-beautiful girl)
	视线-目光 (shi4xian4-mu4guang1, line of sight-gaze)	擦拭-抹布 (ca1shi4-ma1bu4, wipe-rag)

Primary phonological planning units in Chinese spoken word production: Evidence from an ERP study with implicit priming paradigm

WANG Xuejiao, ZHANG Qingfang

(Department of Psychology, Renmin University of China, Beijing 100872, China)

Abstract Languages may differ in terms of the proximate units of phonological encoding in spoken word production. While phonemes are widely accepted as primary units in Indo-European languages, the functional units (syllables or phonemes) in Chinese spoken word production remain controversial. Within the framework of the proximate unit principle for Mandarin Chinese, phonological content is activated when the syllabic frame is retrieved. Both syllabic content and frame are linked sequentially, and tone is specified at this point. Crucially, syllables are retrieved early in word-form encoding, after which phonemes are retrieved in parallel and linked to their positions in the syllable frame. This study investigated the temporal courses of syllables and phonemes using an implicit priming paradigm with concurrent electrophysiological recording. Participants were young native Chinese speakers with low English ability to exclude L2 influence.

In the implicit priming paradigm, four sets of word pairs served as stimuli, each containing four pairs. The first word was the cue, the second the response target. In homogeneous conditions, four response words shared phonological properties (e.g., 危机/wei1ji1/, 围墙/wei2qiang2/, 尾巴/wei3ba/, 卫星/wei4xing1/ share the first syllable but not tone). Heterogeneous conditions used the same pairs but rearranged into sets without shared properties. Two homogeneous conditions were used: syllable-shared and initial-phoneme-shared. Participants memorized the four word pairs per set, then produced targets when presented with cues.

Behavioral results showed no syllable or phoneme effects. EEG results revealed significant syllable overlap effects at 100-400 ms, first in left and right anterior regions, then middle anterior, and finally in anterior and right middle-posterior regions. Phoneme effects appeared only at 500-600 ms after cue onset, localized to right anterior and posterior regions. Both syllable and phoneme homogeneous conditions elicited more positive ERPs than heterogeneous conditions. We propose that syllable overlap effects arise during early phonological encoding, whereas phoneme overlap effects likely emerge during late phonological encoding or phonetic encoding in Chinese spoken word production. These findings support the proximate unit principle.

Keywords: spoken word production; phonological encoding; proximate unit principle; syllable; phoneme

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv – Machine translation. Verify with original.