
AI translation · View original & related papers at
chinaxiv.org/items/chinaxiv-202006.00119

The theta frequency band reflects the syllable encoding process

Authors: Jiang Yuchen, Cai Xiao, Zhang Qingfang, Zhang Qingfang

Date: 2020-06-16T00:00:00+00:00

Abstract

Neural oscillations in the brain often reflect various human cognitive activities. Research on language comprehension has found that activity in the theta band is closely related to syllable processing, yet no studies have investigated the relationship between brain activity in specific frequency bands and syllable processing during speech production. Using EEG time-frequency analysis and a masked priming paradigm, we examined the relationship between theta band activity and syllable processing during picture naming in 23 healthy participants. Behavioral results revealed that when a syllable relationship existed between the prime and target picture name, individuals' naming latencies were faster than in the syllable-unrelated condition, whereas naming latencies in the phoneme-related condition were slower than in the phoneme-unrelated condition. Time-frequency results showed that 270–460 ms after stimulus onset, the power of theta band neural oscillations was significantly lower in the syllable-related condition compared to the syllable-unrelated condition, with no significant difference between phoneme-related and phoneme-unrelated conditions. In summary, we propose that theta band activity during spoken Chinese production reflects syllable processing, providing neural oscillatory evidence that the syllable serves as the unit of phonological encoding in spoken Chinese word production.

Full Text

Theta Band (4–8 Hz) Oscillations Reflect Syllable Processing in Chinese Spoken Word Production

Jiang Yuchen, Cai Xiao, Zhang Qingfang

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

Abstract

Neural oscillations in the brain often reflect various human cognitive activities. While research on language comprehension has found that theta band activity is closely related to syllable processing, no studies have yet investigated the relationship between specific frequency band activity and syllable processing during language production. Using EEG time-frequency analysis and a masked priming paradigm, we examined the relationship between theta band activity and syllable processing in 23 healthy participants during picture naming. Behavioral results showed that naming reaction times were faster in the syllable-related condition compared to the syllable-unrelated condition when prime words and target picture names shared syllabic information, whereas naming times were slower in the phoneme-related condition compared to the phoneme-unrelated condition. Time-frequency results revealed that between 270–460 ms after stimulus onset, theta band oscillation power was significantly lower in the syllable-related condition than in the syllable-unrelated condition, with no significant differences between phoneme-related and phoneme-unrelated conditions. We conclude that theta band activity in Chinese spoken production reflects syllable processing, providing neural oscillatory evidence that syllables constitute the unit of phonological encoding in Chinese spoken word production.

Keywords: speech production; syllable; theta band; time-frequency analysis

Spoken word production refers to the process of expressing thoughts through articulatory organs (Zhang & Yang, 2003b), typically involving three stages: conceptualization, formulation, and articulation. First, speakers must clarify what they intend to express at the conceptual level. Second, they organize specific conceptual information and establish corresponding articulatory motor programs, including lexical selection, morphophonological encoding, phonological encoding, and phonetic encoding. Finally, articulatory targets are output as sound through vocal tract movements (Levelt, Roelofs, & Meyer, 1999; Roelofs, 1997). The processing unit during the phonological encoding stage represents a central focus of debate in spoken production research.

1.1 Cross-Linguistic Differences in Phonological Encoding Units in Speech Production

Research on Indo-European languages indicates that phonemes are the first units extracted during phonological encoding. Dell (1986) found that speech errors in alphabetic languages primarily involve phoneme omissions or substitutions. Using the implicit priming paradigm, researchers have demonstrated that naming sequences of target words sharing initial phonemes (e.g., *cible-cintre-cerf*) yields faster reaction times than sequences with different initial phonemes, producing a first-phoneme priming effect (Alario, Perre, Castel, & Ziegler, 2007; Damian & Bowers, 2003; Jacobs & Dell, 2014; Meyer, 1991). In adjective-noun naming tasks, Damian and Dumay (2007) found faster naming for phoneme-overlapping pairs (e.g., *green goat*) compared to non-overlapping pairs (e.g., *green rug*). Sim-

ilar first-phoneme priming effects have been observed using picture-word interference (Damian & Martin, 1999) and masked priming paradigms (Forster & Davis, 1991; Schiller, 2008).

However, research suggests that syllables rather than phonemes serve as the phonological encoding unit in Chinese spoken production. Speech errors in Chinese mainly involve syllable exchanges (Chen, 2000). Chen, Chen, and Dell (2002) employed the implicit priming paradigm and found shorter naming latencies in syllable-homogeneous conditions compared to syllable-heterogeneous conditions, but no significant differences between phoneme-homogeneous and phoneme-heterogeneous conditions. Similar patterns have been found using masked priming (Chen, O' Seaghdha, & Chen, 2016) and picture-word interference paradigms (Zhang & Yang, 2005). Yue and Zhang (2015) combined picture-word interference with different experimental tasks (immediate naming, delayed naming, delayed naming with articulatory suppression) and found that syllable priming effects occurred during phonological encoding, representing a stable and reliable effect (Cohen's $d = 0.85 > 0.8$) (Cohen, 1988). These convergent findings indicate that syllables are the first units extracted during phonological encoding in Chinese spoken word production.

Based on these cross-linguistic findings, O' Seaghdha, Chen, and Chen (2010) proposed the Proximate Units Principle. Proximate units refer to the phonological encoding units that are first activated after lexical morpheme activation, and the initially selected unit varies across languages: phonemes in Indo-European languages such as English or Dutch, and syllables in Chinese. In Indo-European languages, speakers select phonemes first, then combine metrical information during a syllabification process to extract syllables from a mental syllabary for articulatory programming. In Chinese spoken production, speakers select syllables first, then decompose them into phonemic or segmental information (phonological encoding) to prepare articulatory motor programs (phonetic encoding) before final articulation (see also Roelofs, 2015).

These cross-linguistic differences in phonological encoding units are closely related to the structural characteristics of different languages. In Chinese, syllables represent the minimal natural articulatory units constrained by semantic and structural properties of the language, making them crucial for spoken production (Zhang, 2005). Chinese has far fewer syllables than alphabetic languages (Zhang & Yang, 2005), and syllable boundaries are relatively clear without resyllabification phenomena found in alphabetic languages. Therefore, a more economical and efficient processing strategy for Chinese is to store syllable information of lexical items in long-term memory for direct extraction during early phonological encoding. In Indo-European languages, the vast number of syllables and extensive resyllabification during articulation (Levelt et al., 1999) lead speakers to first extract phonemes during phonological encoding.

Researchers have investigated this issue using event-related potential (ERP) technology. Qu, Damian, and Kazanina (2012) employed a first-phoneme repetition paradigm (e.g., "yellow box" with repeated first phoneme versus "green

box” without repetition) and found phoneme repetition effects between 200–300 ms after picture presentation, indicating that phonemic information is also activated during phonological encoding (see also Yu, Mo, & Mo, 2014). Using picture-picture interference with delayed picture naming, Wang, Wong, Wang, and Chen (2017) found syllable effects during 200–400 ms (phonological encoding) and 400–600 ms (phonetic encoding). Using masked priming, Zhang and Damian (2019) found that syllable-related conditions elicited smaller ERP amplitudes during 300–400 ms after target picture presentation (phonological encoding). Although ERP can distinguish the temporal course of different experimental effects, these studies have overlooked neural oscillation activity during spoken production. Traditional ERP analysis involves averaging neural signals across identical experimental conditions (Rugg & Coles, 1995). However, after multiple averaging, EEG signals can attenuate or even eliminate non-phase-locked neural oscillatory activity (Bidelman, 2015). The present study employs EEG time-frequency analysis to explore the characteristics of neural oscillations during phonological encoding in Chinese native speakers, focusing specifically on neural oscillations corresponding to syllable and phoneme effects.

1.2 The Relationship Between Theta Band (4–8 Hz) Neural Oscillations and Syllable Processing

Neural oscillations are considered rhythmic responses of brain neurons, including delta (δ , < 4 Hz), theta (θ , 4–8 Hz), alpha (α , 8–13 Hz), beta (β , 13–30 Hz), and gamma (γ , > 30 Hz) bands (Ward, 2003). Research indicates that neural oscillations are closely linked to cognitive processes such as attention, memory, and decision-making (Fell & Axmacher, 2011; Jensen, Kaiser, & Lachaux, 2007; Klimesch, 2012; Siegel, Donner, & Engel, 2012). Human language activities also elicit specific frequency band oscillations (Giraud & Poeppel, 2012; Lewis, Wang, & Bastiaansen, 2015). For example, when processing syntactically erroneous materials, beta band power decreases significantly (Bastiaansen, Magyari, & Hagoort, 2009).

Regarding the cognitive neural mechanisms of phonological information processing, the Asymmetric Sampling in Time (AST) theory proposes that the auditory ventral pathway (superior temporal gyrus–superior temporal sulcus–inferior temporal sulcus) differentiates into two parallel processing pathways based on acoustic properties of stimuli, responsible for extracting syllable-level and phoneme-level information respectively (Poeppel, 2003). Syllable information in continuous speech streams changes at a relatively slower temporal rate compared to phonemes, and theta band activity precisely represents syllable processing (Doelling, Arnal, Ghitza, & Poeppel, 2014; Howard & Poeppel, 2012; Luo & Poeppel, 2007; Peelle, Gross, & Davis, 2013). Although neuronal firing frequencies and syllable change frequencies in speech signals are not in simple direct correspondence, this suggests that theta band neural oscillatory activity is closely related to syllable processing.

First, the phase information of theta activity represents syllable tracking (Gh-

inst et al., 2016; Gross et al., 2013; Molinaro, Lizarazu, Lallier, Bourguignon, & Carreiras, 2016). In Pefkou, Arnal, Fontolan, and Giraud's (2017) experiment, researchers first performed envelope analysis on naturally produced sentence audio to obtain the total number of syllables in the acoustic signal, then divided this by total duration to calculate syllable rate, and finally created experimental stimuli with different syllable rates through temporal compression. During the formal experiment, EEG signals were recorded while participants listened to sentences with different syllable rates. Regression analysis revealed that individuals' theta phase coherence decreased as syllable rate increased; that is, when sentences were played faster and participants had greater difficulty tracking syllable information, theta phase coherence became poorer. Additionally, research shows that theta oscillation phase coherence increases when participants are exposed to continuous syllable sequences (James, Natasha, Lisa, & Usha, 2012).

Second, the power of theta activity reflects syllable recognition. In a cross-linguistic study, Peña and Melloni (2012) used spoken sentences in Japanese, Spanish, and Italian to examine language comprehension dynamics in Spanish and Italian native speakers. Results showed that regardless of native language, theta band power was significantly higher when listening to forward-played sentences compared to reversed sentences. Moreover, when listening to non-native materials, theta band activity showed the same pattern as when processing native language materials. Although forward and reversed materials maintained highly consistent basic acoustic properties, reversed materials caused severe phonological distortions that prevented comprehension (Binder et al., 2000; Gross et al., 2013; Saur et al., 2010). Peña and Melloni (2012) noted that temporal reversal of experimental materials destroyed the original phonological structure of words, making syllable segmentation difficult and resulting in weaker theta band neural oscillatory activity. In a study on Chinese, researchers similarly found stronger theta band power for forward-played compared to reversed syllable stimuli (Ding, Melloni, Zhang, Tian, & Poeppel, 2015).

In summary, behavioral and ERP studies demonstrate that syllables play an important role in Chinese spoken production, while language perception and comprehension research shows that theta band power is closely related to syllable processing. No studies have yet examined whether syllable priming effects in Chinese spoken production are associated with specific frequency band neural oscillatory activity, and EEG time-frequency analysis can help us better understand the mechanisms of syllable processing during Chinese spoken production. In this study, we used a masked priming paradigm, manipulating phonological relationships between prime words and target picture names (syllable-related vs. syllable-unrelated, phoneme-related vs. phoneme-unrelated), requiring participants to complete a picture naming task while recording EEG signals. We then performed time-frequency analysis on the EEG data to examine the neural oscillations underlying syllable and phoneme effects. In the masked priming paradigm, prime words are presented very briefly (approximately 50 ms), and prime processing is typically subliminal, excluding the influence of irrelevant fac-

tors such as naming strategies on experimental results (Chen et al., 2016; You, Zhang, & Verdonschot, 2012). Given the close relationship between theta band and syllable processing (Ghinst et al., 2016; Gross et al., 2013; James et al., 2012; Peña & Melloni, 2012), we expected to find significant syllable priming effects but not phoneme priming effects, with correspondingly significant differences in theta band power only between syllable-related and unrelated conditions.

2. Method

2.1 Participants

Twenty-three university students and graduate students (11 males, mean age 22 years) participated in the study. All participants were right-handed, had no history of psychiatric disorders, were native Chinese speakers who spoke standard Mandarin, and had normal or corrected-to-normal vision. Participants read and signed informed consent forms before the experiment and received monetary compensation afterward.

2.2 Materials

Sixty-four black-and-white line drawings were selected from the Chinese picture database established by Zhang and Yang (2003a), with 60 pictures used in the formal experiment and 4 for practice trials. Each picture name was a disyllabic noun, with the first character matched to four experimental conditions. For example, for the picture name “鼻子” (/bi2zi5/), the syllable-related prime was “彼” (/bi3/), which shared the identical syllable but differed in tone; the phoneme-related prime was “柏” (/bai3/), which shared the initial phoneme but differed in tone. Word frequencies for syllable-related and phoneme-related primes did not differ significantly (Cai & Brysbaert, 2010; $t = 0.027$, $p = 0.978$). Primes for syllable-unrelated and phoneme-unrelated conditions were created by randomly reassigning the primes from the related conditions to different pictures. Checks confirmed no phonological relationship between primes and target picture names in the unrelated conditions. No semantic or orthographic relationships existed between primes and picture names across all four conditions (see Appendix for all materials).

2.3 Design

The study employed a $2 \times 2 \times 2$ within-subjects design with independent variables including relatedness type (syllable, phoneme), relatedness condition (related, unrelated), and repetition (first, second). Each picture was matched with four types of prime words, resulting in 240 trials per test block. Each test block was repeated twice, yielding 480 trials per participant. Trial presentation order differed for each participant and was pseudo-randomized to ensure at least five trials between presentations of the same picture and prevent consecutive trials with pictures sharing the same initial phoneme.

2.4 Apparatus

The experiment was programmed using E-Prime 2.0. A PST SRBOX response box, microphone, and computer were used. All pictures were presented centrally on a computer screen, with participant responses recorded via a microphone connected to the response box. Stimulus presentation and response time collection were controlled by the computer, and the experimenter recorded whether participants responded correctly. EEG signals were recorded using a NeuroScan system with a 64-channel cap based on the international 10-20 system.

2.5 Procedure

The experiment consisted of three phases: a learning phase, a test phase, and a formal experimental phase. In the learning phase, each picture and its corresponding name were presented centrally for 2 s. Participants were informed that these pictures would appear in the formal experiment and were instructed to memorize the content and names of the pictures. In the test phase, pictures were presented and participants were required to name them. Participants proceeded to the formal experiment only after naming all pictures correctly. All materials were common everyday pictures with high naming consistency, and all participants successfully learned the picture names.

The trial procedure in the formal experiment was as follows (see Figure 1 [Figure 1: see original paper]): First, a fixation point (“+”) appeared for 500 ms, followed by a 500 ms forward mask (@@). The prime word was then presented for 49 ms, followed by a 20 ms backward mask (@@). After mask offset, the target picture appeared centrally. Participants were required to name the picture as quickly and accurately as possible within 2000 ms, after which the picture disappeared immediately. The next trial began after a 1000 ms interval. Four practice trials preceded the formal experiment to familiarize participants with the task. All picture stimuli were standardized to uniform size, prime words were presented in 28-point Song font, and mask stimuli (@@) in 36-point Song font. Completing all trials required approximately 100 minutes.

2.6 EEG Recording and Analysis

EEG was recorded online with the left mastoid as reference and re-referenced offline to bilateral mastoids. Horizontal electrooculogram (HEOG) and vertical electrooculogram (VEOG) from the left eye were simultaneously recorded. Impedance between electrodes and scalp was maintained below 5 k Ω , with a filter bandpass of 0.05-70 Hz and sampling rate of 500 Hz.

Preprocessing was conducted using the EEGLAB toolbox (Delorme & Makeig, 2004). First, for electrodes with poor contact or damage during recording (fewer than 5% of total electrodes), bad channel replacement was performed using EEGLAB’s built-in spherical interpolation function, which estimates signals using data from surrounding electrodes (Perrin, Pernier, Bertrand, & Echallier, 1989; Pivik et al., 1993). Second, data were filtered at 0.1-30 Hz and subjected

to Independent Component Analysis (ICA). ICA is based on blind source separation, decomposing multi-channel observed signals into independent components (ICs) according to statistical independence principles. ICA shows high sensitivity in identifying artifacts caused by blinking and muscle movements (Makeig, Bell, Jung, & Sejnowski, 1995). Ocular artifact criteria included: top-ranked components, concentrated frontal scalp activity, slowly decaying power with increasing frequency, and large single-trial power. Muscle artifact criteria included: diffuse lateral scalp activity, power increasing with frequency, and large single-trial power. Each component was inspected according to these criteria, combined with the EEGLAB plugin ADJUST for artifact identification and rejection. Third, EEG signals were segmented from 1000 ms before to 1500 ms after picture onset, excluding trials with amplitudes exceeding ± 100 μ V.

Time-frequency analysis was performed using the Fieldtrip toolbox with wavelet transform methods (Oostenveld, Fries, Maris, & Schoffelen, 2011). Single-trial total duration was 2500 ms with temporal resolution of 10 ms; frequency analysis ranged from 3–30 Hz with 1 Hz resolution. Wavelet cycles increased linearly, with a cycle of 3 at the lowest frequency and 8 at the highest frequency (see Li, Shao, Xia, & Xu, 2019 for similar parameters). Sinusoidal wavelets of different cycles were convolved with EEG time-domain signals to obtain neural oscillation power values across different frequency ranges and time points (Goupillaud, Grossmann, & Morlet, 1984). Analysis first estimated power for single trials, then averaged across multiple trials. A baseline of 300–100 ms before stimulus onset was used for correction on a decibel scale: $\text{dB} = 10 \cdot \log_{10}(\text{post-baseline power} / \text{mean baseline power})$. For simplicity, Event-Related Spectral Perturbation (ERSP) figures display frequencies from 4–20 Hz.

Based on electrode spatial distribution, six Regions of Interest (ROIs) were selected and power was averaged within each: left anterior (F3, FC3, FC5), mid anterior (Fz, FCz, Cz), right anterior (F4, FC4, FC6), left posterior (P5, P3, PO3), mid posterior (CPz, Pz, POz), and right posterior (P6, P4, PO4). A 2 (relatedness type: syllable, phoneme) $\times 2$ (relatedness condition: related, unrelated) $\times 2$ (repetition: first, second) $\times 6$ (ROI: left anterior, mid anterior, right anterior, left posterior, mid posterior, right posterior) repeated-measures ANOVA was conducted, with p-values corrected using Greenhouse-Geisser when sphericity was violated.

To examine significant differences between syllable-related vs. unrelated and phoneme-related vs. unrelated conditions, cluster-based permutation tests were used for statistical analysis, which effectively corrects for multiple comparisons (Maris & Oostenveld, 2007). The permutation test time window was 600 ms after stimulus onset with 10 ms steps across six ROIs. Repeated-measures t-tests were performed on data (time \times frequency \times electrode) for each pair of conditions of interest. Data points with $p < 0.05$ that were adjacent in time and space were merged into clusters. The sum of t-values within each cluster was calculated to determine cluster-level statistics, with significance tested using Monte Carlo methods with 1000 random permutations.

3. Results

3.1 Behavioral Results

Trials from two participants with error rates $>10\%$ and two items with mean reaction times >3 standard deviations were excluded. Data not recorded by the equipment (2.5%) and incorrect naming responses (1.8%) were removed. Reaction times <200 ms or >1500 ms, including those >3 standard deviations from the mean, accounted for 2.3% of trials. Figure 2 [Figure 2: see original paper] shows mean naming reaction times and 95% confidence intervals across experimental conditions.

A 2 (relatedness type: syllable, phoneme) \times 2 (relatedness condition: related, unrelated) \times 2 (repetition: first, second) repeated-measures ANOVA on reaction times showed that the main effect of relatedness type was significant in participant analysis and marginally significant in item analysis: participants named faster in syllable conditions, $F(1, 22) = 7.67$, $p = 0.01$, $p^2 = 0.26$; $F(1, 55) = 3.46$, $p = 0.07$, $p^2 = 0.06$. The main effect of repetition was significant: naming was faster in the second repetition, $F(1, 22) = 48.24$, $p < 0.001$, $p^2 = 0.69$; $F(1, 55) = 282.59$, $p < 0.001$, $p^2 = 0.84$. The interaction between relatedness type and condition was significant: $F(1, 22) = 5.51$, $p = 0.03$, $p^2 = 0.20$; $F(1, 55) = 7.04$, $p = 0.01$, $p^2 = 0.11$. Simple effects analysis revealed significantly faster naming in syllable-related vs. syllable-unrelated conditions, $F(1, 22) = 6.12$, $p = 0.02$, $p^2 = 0.22$; $F(1, 55) = 5.87$, $p = 0.02$, $p^2 = 0.10$. No significant difference emerged between phoneme-related and phoneme-unrelated conditions, $F(1, 22) = 0.85$, $p = 0.37$; $F(1, 55) = 1.45$, $p = 0.23$. The three-way interaction among relatedness type, condition, and repetition was significant in participant analysis and marginally significant in item analysis: $F(1, 22) = 7.88$, $p = 0.01$, $p^2 = 0.26$; $F(1, 55) = 3.75$, $p = 0.06$, $p^2 = 0.06$. In the first repetition, the interaction between relatedness type and condition was significant, $F(1, 22) = 13.38$, $p = 0.001$, $p^2 = 0.38$; $F(1, 55) = 8.53$, $p = 0.005$, $p^2 = 0.13$. Specifically, naming was significantly faster in syllable-related vs. syllable-unrelated conditions, $F(1, 22) = 5.88$, $p = 0.02$, $p^2 = 0.21$; $F(1, 55) = 6.01$, $p = 0.02$, $p^2 = 0.10$, while naming was significantly slower in phoneme-related vs. phoneme-unrelated conditions, $F(1, 22) = 5.04$, $p = 0.04$, $p^2 = 0.19$; $F(1, 55) = 4.53$, $p = 0.04$, $p^2 = 0.08$. In the second repetition, the interaction between relatedness type and condition was not significant, $F < 1$, $F < 1$. No other main effects or interactions were significant.

3.2 Time-Frequency Analysis Results

One participant was excluded from subsequent analysis due to excessive signal artifacts. Trials with reaction times <500 ms (5.1%), >1500 ms (1.0%), and those with large artifacts (5.3%) were removed. Based on relevant research, six consecutive time windows were defined: 0–100 ms, 100–200 ms, 200–300 ms, 300–400 ms, 400–500 ms, and 500–600 ms (see Qu et al., 2012; Zhang & Damian,

2019 for similar windows). ANOVA results for theta power across these windows are presented in Table 1 .

In the 0-100 ms, 100-200 ms, and 200-300 ms windows, the main effect of relatedness type was significant, with syllable conditions eliciting greater power: F0-100 ms(1, 21) = 8.08, $p = 0.01$, $p^2 = 0.28$; F100-200 ms(1, 21) = 11.03, $p = 0.003$, $p^2 = 0.35$; F200-300 ms(1, 21) = 5.18, $p = 0.03$, $p^2 = 0.20$. In the 300-400 ms window, the interaction between relatedness type and condition was significant, $F(1, 21) = 5.13$, $p = 0.03$, $p^2 = 0.20$. Mean power was lower in syllable-related vs. syllable-unrelated conditions, $F(1, 21) = 4.43$, $p = 0.047$, $p^2 = 0.17$, with no significant difference between phoneme-related and phoneme-unrelated conditions, $p = 0.14$. In the 400-500 ms window, the interaction between relatedness type and condition was significant, $F(1, 21) = 6.41$, $p = 0.02$, $p^2 = 0.23$. Mean power was lower in syllable-related vs. syllable-unrelated conditions, $F(1, 21) = 5.81$, $p = 0.03$, $p^2 = 0.22$, with no significant difference between phoneme-related and phoneme-unrelated conditions, $p = 0.38$. In the 500-600 ms window, the interaction between relatedness type and condition was significant, $F(1, 21) = 6.00$, $p = 0.02$, $p^2 = 0.22$. Mean power was marginally lower in syllable-related vs. syllable-unrelated conditions, $F(1, 21) = 3.57$, $p = 0.07$, $p^2 = 0.15$, with no significant difference between phoneme-related and phoneme-unrelated conditions, $p = 0.16$. No other significant interactions were found for power in other frequency bands across time windows.

Cluster-based permutation tests for syllable and phoneme effects showed that in the first repetition, syllable-related conditions elicited lower theta power than syllable-unrelated conditions in the right anterior ROI (F4, FC4, FC6) between 270-460 ms after stimulus onset ($p = 0.01$, 4-8 Hz). Phoneme-related conditions produced marginally higher theta power than phoneme-unrelated conditions in the left anterior ROI (F3, FC3, FC5) between 340-390 ms ($p = 0.052$, 4-8 Hz). In the second repetition, no significant syllable or phoneme effects were found in any ROI. No significant differences between conditions of interest were found for power in other frequency bands (δ , α , β) between 3-30 Hz (see Table 2). Figures 3a [Figure 3: see original paper] and 3b present ERSPPs for syllable and phoneme effects across repetitions.

4. Discussion

This study used a masked priming paradigm to investigate the relationship between syllable and phoneme processing and theta band power in Chinese spoken production. Behavioral results showed faster picture naming in syllable-related vs. unrelated conditions, but slower naming in phoneme-related vs. unrelated conditions. EEG time-frequency analysis revealed that between 270-460 ms after picture onset, theta band power differed significantly between syllable-related and unrelated conditions, being lower in syllable-related conditions, with no significant differences between phoneme-related and unrelated conditions. These results indicate that theta band activity during Chinese spoken production reflects syllable information processing.

At the behavioral level, consistent with Zhang and Wang (2014), practice effects from trial repetition not only reduced overall naming latencies in the second session but also eliminated the experimental effects of interest. Critically, in the first naming session, we successfully replicated syllable priming effects found previously using implicit priming (Chen et al., 2002), picture-word interference (Yue & Zhang, 2015), and masked priming paradigms (Chen et al., 2016). Additionally, we observed longer naming latencies in phoneme-related conditions, demonstrating a phoneme inhibition effect (Chen et al., 2016). Notably, research on alphabetic languages has found that presenting phoneme-related primes reduces naming latencies for target stimuli (Alario et al., 2007; Damian & Bowers, 2003; Damian & Martin, 1999; Forster & Davis, 1991; Schiller, 2008). This suggests different cognitive mechanisms for spoken production in alphabetic languages versus Chinese. Specifically, alphabetic languages require processing stress and phoneme information to determine syllable structure during phonological encoding (Levelt et al., 1999), whereas Chinese has fewer syllables, clear syllable boundaries, and no resyllabification, allowing Chinese speakers to directly extract syllable information during phonological encoding (O' Seaghdha et al., 2010). Our reaction time results suggest that syllables function as independent information representation units in Chinese spoken production, with advance processing of target syllable information reducing naming latencies in syllable-related conditions. The phoneme inhibition effect likely results from competition due to co-activation of syllables sharing the same initial phoneme. In phoneme-related conditions, overlapping initial phonemes (prime: 柏/bai3/, target: 鼻子/bi2zi5/) activated all syllables beginning with that phoneme (e.g., /ba/, /bang/), creating interference during target syllable retrieval and prolonging naming latencies (discussed further below in relation to theta oscillations). Similarly, Sereno and Lee (2015) found phoneme inhibition effects in a Chinese auditory lexical decision task. However, we consider our phoneme inhibition effect relatively weak: neither syllable nor phoneme effects were significant in the second test session, though the two-way interaction between relatedness type and condition remained significant, with faster naming in syllable-related conditions and no significant difference between phoneme conditions, indicating at least a trend toward syllable priming rather than phoneme inhibition (see Figure 2).

At the neural level, we found an interaction between relatedness type and condition for theta band oscillations: lower theta power in syllable-related vs. unrelated conditions. Importantly, syllable effects occurred approximately 270–460 ms after stimulus onset, consistent with time windows for syllable effects in previous picture naming ERP studies (Cai, Yin, & Zhang, 2020; Dell'acqua et al., 2010; Wang et al., 2017; Yu et al., 2014; Zhu, Damian, & Zhang, 2015). According to meta-analysis results (Indefrey & Levelt, 2004), speakers typically enter phonological encoding around 250 ms after picture onset, when Chinese speakers need to extract target syllable information for subsequent processing. Two possible explanations exist for lower theta power in syllable-related vs. unrelated conditions. First, syllable repetition may cause power reduction

due to repeated identical syllables. Grill-Spector, Henson, and Martin (2006) noted that neuronal activity decreases when processing repeated stimuli, manifesting as reduced oscillatory power. Gruber and Müller (2002) found lower gamma band power for repeated vs. non-repeated stimuli during object recognition between 220–350 ms. Using task-switching paradigms, researchers found significant decreases in beta and gamma power when consecutive tasks were consistent (repetition condition) vs. inconsistent (non-repetition condition) (Gruber, Giabbiconi, Trujillo-Barreto, & Müller, 2006). This repetition-induced power reduction has been replicated in face recognition studies (Engell & McCarthy, 2014). Brookes et al. (2005) used fMRI to show that repetition-induced oscillatory power decreases correlate with attenuated cortical BOLD signals. In our study, syllable-related primes shared identical syllables with target picture names, likely causing a “repetition effect” and reduced theta power. Second, differences in cognitive processing load between conditions may cause power changes. Working memory research shows that tasks with heavier cognitive load elicit relatively higher neuronal power, while lighter-load tasks show lower power (Roux & Uhlhaas, 2014). In syllable-related conditions, advance processing of target syllable information reduces cognitive load during phonological encoding when viewing targets. Conversely, syllable-unrelated conditions prevent direct syllable extraction from primes, resulting in higher cognitive load. Therefore, theta power is weaker in syllable-related vs. unrelated conditions, consistent with behavioral reaction time patterns: lower cognitive load enables faster naming in syllable-related conditions.

It should be noted that speech perception studies comparing natural vs. reversed speech found higher theta power for natural speech, inconsistent with our pattern. This discrepancy arises from different comparison conditions: Peña and Melloni (2012) and Ding et al. (2015) found lower theta power for reversed vs. forward speech, attributing this to reversed materials destroying normal phonological structure and preventing syllable recognition, leaving corresponding cortical neurons in an inhibited state with relatively low oscillatory power. In forward speech conditions, successful syllable processing places cortical neurons in an excited state with relatively high power. Our study compared syllable-related vs. unrelated conditions—both excitatory states—whereas speech perception studies compared excitatory vs. inhibitory states, producing different patterns.

Unlike syllable effects, we found marginally significant power differences between phoneme-related and unrelated conditions between 340–390 ms, with higher theta power in phoneme-related vs. unrelated conditions during the first repetition, showing a different pattern from syllable-related conditions. When primes and targets shared phoneme-relatedness, overlapping initial phonemes activated all syllables beginning with that phoneme (e.g., /ba/, /bang/), creating competition for target syllable production and resulting in stronger theta power for phoneme-related vs. unrelated conditions. This aligns with Chen et al. (2016), who used identical experimental design and tasks and found weak inhibition effects for phoneme-related vs. unrelated conditions.

In Chinese spoken word production, theta band power changes for syllable effects occurred earlier than for phoneme effects, consistent with established temporal patterns and providing evidence from power changes that syllables are proximate units of phonological encoding. For example, Feng, Yue, and Zhang (2019) used picture-word interference, manipulating relationships between distractors and target names, finding syllable effects around 320 ms and phoneme effects around 368 ms after picture onset, both during phonological encoding. In Zhang and Damian (2019), besides syllable effects at 300–400 ms, phoneme effects were detected: phoneme-related conditions elicited smaller ERP amplitudes than unrelated conditions at 500–600 ms.

Although no studies have directly compared the time course of phoneme vs. syllable effects in Indo-European speakers using EEG, research on syllable frequency effects may provide partial evidence. Syllable frequency effects refer to faster processing and shorter reaction times for higher-frequency syllables during language production, closely related to syllable-level processing (Levelt et al., 1999). Using ERP response-locked analysis, Burki, Cheneval, and Laganaro (2015) found syllable frequency effects between 180–150 ms before articulation, corresponding to the phonetic encoding stage. Dell’Acqua et al. (2010) compared semantic and phonological effects in Italian speakers, finding that first-phoneme overlap produced facilitation effects between 250–400 ms after stimulus onset—during phonological encoding (Indefrey & Levelt, 2004). These studies provide indirect EEG temporal evidence that alphabetic languages process phonemes before syllables. In contrast, our findings show the opposite pattern for Chinese: syllable effects precede phoneme effects in terms of power changes. Chinese and Indo-European languages show completely opposite temporal patterns for syllable and phoneme extraction, consistent with cross-linguistic predictions of the Proximate Units Principle (O’ Seaghdha et al., 2010).

In summary, syllable-related conditions shortened picture naming times and reduced theta power, while phoneme-related conditions tended to prolong naming times and increase theta power. Both language comprehension and production studies indicate that theta band power is related to syllable processing. Nevertheless, our weak, marginally significant phoneme effect suggests theta power may be related to phonological processing more broadly. The disappearance of both syllable and phoneme effects in reaction times during the second repetition is consistent with previous research (Chen et al., 2016), and our time-frequency results align with behavioral findings, indicating rapid neural adaptation. Future research should examine English speakers to test whether significant phoneme effects behaviorally correspond to theta power differences between related and unrelated conditions, enabling cross-linguistic comparisons and deeper investigation of the cognitive significance of theta power.

It is worth noting that although traditional ERP analysis can distinguish the temporal sequence of syllable and phoneme processing, ERP amplitudes show opposite patterns across studies: unlike Zhang and Damian (2019), Dell’Acqua et al. (2010) found larger ERP amplitudes for phonologically related vs. un-

related conditions. Similar to contradictory N400 effects in semantic violation studies (Petten & Luka, 2012), inconsistent ERP direction and amplitude across studies make it difficult to interpret language processing mechanisms. Even when ERP results show high cross-experiment consistency, difference waves do not correspond to specific cognitive processes, preventing complete attribution of different ERP patterns to distinct processing mechanisms in spoken production. For example, in Wang et al. (2017), syllable-related conditions had relatively greater stimulus overlap (剪刀/jian3dao1/—键盘/jian4pan2/) compared to phoneme-related conditions (西瓜/xi1gua1/—信封/xin4feng1/), making it possible that participants subliminally detected differences in overlap magnitude, producing larger differences between syllable-related vs. unrelated than phoneme-related vs. unrelated conditions. Although our experimental conditions were similar to previous studies, time-frequency analysis calculates power across different time points and frequency bands, directly reflecting activation/inhibition of neurons with different firing rates (Cohen, 2017). Our finding of no significant differences in other frequency bands (δ , α , β) between 3-30 Hz indicates that differences in overlap magnitude between syllable- and phoneme-related conditions did not confound our main results. Moreover, theta power was lower in syllable-related vs. unrelated conditions, but showed a trend toward being higher in phoneme-related vs. unrelated conditions. These opposing oscillatory patterns reflect different cognitive-neural mechanisms for syllable vs. phoneme processing during spoken production.

In conclusion, theta band power in Chinese spoken production reflects syllable processing, providing frequency-band evidence that syllables constitute the unit of phonological encoding in Chinese spoken word production. Our study also demonstrates that language comprehension and production elicit similar power changes in neural oscillations. Future research should combine different experimental paradigms to further explore the cognitive significance of power changes across frequency bands in Chinese spoken production.

References

- Alario, F. X., Perre, L., Castel, C., & Ziegler, J. C. (2007). The role of orthography in speech production revisited. *Cognition*, *102*(3), 464-475.
- Bastiaansen, M., Magyari, L., & Hagoort, P. (2009). Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *Journal of Cognitive Neuroscience*, *22*(7), 1333-1347.
- Bidelman, G. M. (2015). Induced neural beta oscillations predict categorical speech perception abilities. *Brain and Language*, *141*, 62-69.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., & Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, *10*(5), 512-528.
- Brookes, M. J., Gibson, A. M., Hall, S. D., Furlong, P. L., Barnes, G. R.,

- Hillebrand, A., & Morris, P. G. (2005). GLM-beamformer method demonstrates stationary field, alpha ERD and gamma ERS co-localisation with fMRI BOLD response in visual cortex. *NeuroImage*, *26*(1), 302-308.
- Burki, A., Cheneval, P. P., & Laganaro, M. (2015). Do speakers have access to a mental syllabary? ERP comparison of high frequency and novel syllable production. *Brain and Language*, *150*, 90-102.
- Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLOS ONE*, *5*(6), e10729.
- Cai, X., Yin, Y. L., & Zhang, Q. F. (2020). The roles of syllables and phonemes during phonological encoding in Chinese spoken word production: A topographic ERP study. *Neuropsychologia*, *144*, 1-10.
- 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*(1), 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*(4), 751-781.
- Chen, J.-Y., O' Seaghdha, 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), 825-836.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cohen, M. X. (2017). Where Does EEG Come From and What Does It Mean. *Trends in Neurosciences*, *40*(4), 208-218.
- 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*(2), 345-361.
- Damian, M. F., & Bowers, J. S. (2003). Effects of orthography on speech production in a form-preparation paradigm. *Journal of Memory and Language*, *49*(1), 119-132.
- Damian, M. F., & Dumay, N. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language*, *57*(2), 195-209.
- Dell'acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., & Grainger, J. (2010). ERP evidence for ultra-fast semantic processing in the picture-word interference paradigm. *Frontiers in Psychology*, *1*, 177.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283-321.

- Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.
- Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2015). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158–164.
- Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, *85*, 761–768.
- Engell, A. D., & McCarthy, G. (2014). Repetition suppression of face-selective evoked and induced EEG recorded from human cortex. *Human Brain Mapping*, *35*(8), 4155–4162.
- Fell, J., & Axmacher, N. (2011). The role of phase synchronization in memory processes. *Nature Reviews Neuroscience*, *12*(2), 105–118.
- Feng, C., Yue, Y., & Zhang, Q. F. (2019). Syllables are Retrieved before Segments in the Spoken Production of Mandarin Chinese: An ERP Study. *Scientific Reports*, *9*(1), 1–9.
- 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.
- Ghinst, M. V., Bourguignon, M., De Beeck, M. O., Wens, V., Marty, B., Hassid, S., ...De Tieghe, X. (2016). Left superior temporal gyrus is coupled to attended speech in a cocktail-party auditory scene. *The Journal of Neuroscience*, *36*(5), 1596–1606.
- Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517.
- Goupillaud, P. L., Grossmann, A., & Morlet, J. (1984). Cycle-octave and related transforms in seismic signal analysis. *Geoexploration*, *23*(1), 85–102.
- Grill-Spector, K., Henson, R. N., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*(1), 14–23.
- Gross, J., Hoogenboom, N., Thut, G., Schyns, P. G., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLOS Biology*, *11*(12), e1001752.
- Gruber, T., & Müller, M. M. (2002). Effects of picture repetition on induced gamma band responses, evoked potentials, and phase synchrony in the human EEG. *Cognitive Brain Research*, *13*(3), 377–392.

- Gruber, T., Giabbiconi, C. M., Trujillo-Barreto, N. J., & Müller, M. M. (2006). Repetition suppression of induced gamma band responses is eliminated by task switching. *European Journal of Neuroscience*, *24*(9), 2654-2660.
- Howard, M. F., & Poeppel, D. (2012). The neuromagnetic response to spoken sentences: co-modulation of theta band amplitude and phase. *Neuroimage*, *60*(4), 2118-2127.
- Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1-2), 101-144.
- Jacobs, C. L., & Dell, G. S. (2014). 'Hotdog', not 'Hot' 'dog': the phonological planning of compound words. *Language, Cognition and Neuroscience*, *29*(4), 512-523.
- James, P. A., Natasha, M., Lisa, B., & Usha, G. (2012). Neural entrainment to rhythmically presented auditory, visual, and audio-visual speech in children. *Frontiers in Psychology*, *3*, 216.
- Jensen, O., Kaiser, J., & Lachaux, J. P. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends in Neurosciences*, *30*(7), 317-324.
- Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606-617.
- Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1-75.
- Lewis, A. G., Wang, L., & Bastiaansen, M. (2015). Fast oscillatory dynamics during language comprehension: unification versus maintenance and prediction? *Brain and Language*, *148*, 51-63.
- Li, X., Shao, X., Xia, J., & Xu, X. (2019). The cognitive and neural oscillatory mechanisms underlying the facilitating effect of rhythm regularity on speech comprehension. *Journal of Neurolinguistics*, *49*, 155-167.
- Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*(6), 1001-1010.
- Makeig, S., Bell, A. J., Jung, T., & Sejnowski, T. J. (1995). Independent component analysis of electroencephalographic data. *Neural Information Processing Systems*, *8*(8), 145-151.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177-190.
- 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.

- Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., & Carreiras, M. (2016). Out-of-synchrony speech entrainment in developmental dyslexia. *Human Brain Mapping, 37*(8), 2767-2783.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*, 156869.
- 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.
- Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex, 23*(6), 1378-1387.
- Pefkou, M., Arnal, L. H., Fontolan, L., & Giraud, A. L. (2017). γ -band and β -band neural activity reflects independent syllable tracking and comprehension of time-compressed speech. *Journal of Neuroscience, 37*(33), 7889-7896.
- Peña, M., & Melloni, L. (2012). Brain oscillations during spoken sentence processing. *Journal of Cognitive Neuroscience, 24*(5), 1149-1164.
- Perrin, F., Pernier, J., Bertrand, O. F., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology, 72*(2), 184-187.
- Petten, C. V., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176-190.
- Pivik, R. T., Broughton, R., Coppola, R., Davidson, R. J., Fox, N. A., & Nuwer, M. R. (1993). Guidelines for the recording and quantitative analysis of electroencephalographic activity in research contexts. *Psychophysiology, 30*(6), 547-558.
- Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication, 41*(1), 245-255.
- 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*(35), 14265-14269.
- Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition, 64*(3), 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.

- Roux, F., & Uhlhaas, P. J. (2014). Working memory and neural oscillations: alpha-gamma versus theta-gamma codes for distinct WM information? *Trends in Cognitive Sciences*, 18(1), 16-25.
- Rugg, M. D., & Coles, M. G. H. (1995). *Electrophysiology of mind: Event-related brain potentials and cognition* (Oxford Psychology Series, No. 25). New York, NY: Oxford University Press.
- Saur, D., Schelter, B., Schnell, S., Kratochvil, D., Kupper, H., Kellmeyer, P., ...Weiller, C. (2010). Combining functional and anatomical connectivity reveals brain networks for auditory language comprehension. *NeuroImage*, 49(4), 3187-3197.
- Schiller, N. O. (2008). The masked onset priming effect in picture naming. *Cognition*, 106, 952-962.
- Sereno, J. A., & Lee, H. (2015). The Contribution of Segmental and Tonal Information in Mandarin Spoken Word Processing. *Language and Speech*, 58(2), 131-151.
- Siegel, M., Donner, T. H., & Engel, A. K. (2012). Spectral fingerprints of large-scale neuronal interactions. *Nature Reviews Neuroscience*, 13(2), 121-134.
- 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(1), 5815.
- Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences*, 7(12), 553-559.
- You, W. P., Zhang, Q. F., & 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. F. (2003a). The determiners of picture-naming latency. *Acta Psychologica Sinica*, 35(4), 447-454.
- Zhang, Q. F., & Yang, Y. F. (2003b). The lexical access theory in speech production. *Journal of Developments in Psychology*, 11(1), 6-11.
- Zhang, Q. F., & Yang, Y. F. (2005). The phonological planning unit in Chinese monosyllabic word production. *Psychological Science*, 28(2), 374-378.
- Zhang, Q. F. (2005). The syllable' s role in language production. *Advances in Psychological Science*, 13(6), 752-759.
- Zhang, Q. F., & Wang, C. (2014). Syllable frequency and word frequency effects in spoken and written word production in a non-alphabetic script. *Frontiers in*

Psychology, 5, 120.

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

Zhu, X., Damian, M. F., & Zhang, Q. (2015). Seriality of semantic and phonological processes during overt speech in Mandarin as revealed by event-related brain potentials. *Brain and Language*, 144, 16-25.

Appendix: Experimental Materials

The appendix lists experimental materials including prime characters with phonetic transcriptions and corresponding target picture names. For example: 彼/bi3/, 冈/gang1/, 柏/bai3/, 估/gu1/ paired with 鼻子/bi2.zi5/; 胖/pang4/, 负/fu4/, 票/piao4/, 份/fen4/ paired with 螃蟹/pang2.xie4/; and similar pairings for all experimental items.

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

Source: ChinaXiv – Machine translation. Verify with original.