Cognitive Mechanisms of Syllable Frequency Effects in Speech Production: A Cross-Linguistic Perspective Postprint

Authors: Pan Jiabing, Zhang Qingfang

Date: 2023-10-09T00:00:00+00:00

Abstract

In spoken word production, the syllable frequency effect refers to the processing advantage of high-frequency syllables over low-frequency syllables. This article reviews the manifestations and theoretical foundations of the syllable frequency effect in speech production, elaborating on the cross-linguistic differences in this effect between Indo-European languages and Chinese from the perspectives of influencing factors, stage of occurrence, and neural mechanisms. Based on theoretical models of spoken production and the appropriate unit hypothesis, combined with the inherent characteristics of different languages, it analyzes the reasons for the existence of cross-linguistic differences in the syllable frequency effect, proposes a model concerning the mechanism of syllable action in Chinese spoken word production, and provides a new perspective on the role of syllables in the spoken production of both Chinese and alphabetic languages. Future research should further explore the specific mechanisms by which the two measurement methods of syllable frequency influence the syllable frequency effect, strengthen the investigation of the syllable frequency effect in Chinese spoken word production, and utilize multiple experimental techniques and paradigms to deeply examine the cross-linguistic differences in syllable processing during speech production.

Full Text

Preamble

ChinaXiv Cooperative Journal
Advances in Psychological Science 2023, Vol. 31, No. 9, 1676–1687

https://doi.org/10.3724/SP.J.1042.2023.01676

Cognitive Mechanisms of Syllable Frequency Effects in Speech Production: A Cross-Language Perspective

PAN Jiabing, ZHANG Qingfang
(Department of Psychology, Renmin University of China, Beijing 100872, China)

Abstract

In spoken word production, the syllable frequency effect refers to the processing advantage of high-frequency syllables over low-frequency syllables. This review summarizes the manifestations and theoretical foundations of syllable frequency effects in speech production, and elaborates on cross-linguistic differences between Indo-European languages and Chinese from the perspectives of influencing factors, processing stages, and neural mechanisms. Based on speech production models and the proximate units principle, combined with the inherent properties of different languages, we analyze the reasons for cross-linguistic differences in syllable frequency effects and propose a model concerning the mechanism of syllable processing in Chinese spoken word production, providing a new perspective on the role of syllables in speech production across Chinese and alphabetic languages. Future research should further investigate the specific mechanisms through which the two measurement methods of syllable frequency influence syllable frequency effects, strengthen the examination of syllable frequency effects in Chinese spoken word production, and deeply explore cross-linguistic differences in syllable processing during speech production using multiple experimental techniques and paradigms.

Keywords: speech production, syllable frequency effect, cross-language, mental syllabary, syllabic neighbours

Classification Code: B842

Speech production plays a crucial role in human communication, referring to the psychological process of transforming thoughts into speech and expressing them through articulatory organs. It comprises three cognitive processing stages: conceptual preparation, where speakers clarify the concepts and content they wish to express; formulation, where concepts are organized and converted into corresponding linguistic forms; and articulation, where vocal muscle movements generate speech output [**?**, **?**].

According to the WEAVER++ (Word-form Encoding by Activation and Verification++) model of speech production [**?**, **?**], formulation corresponds to the lexical access stage, which includes lexical selection, morphological encoding, and phonological encoding. During phonological encoding, speakers extract segmental and metrical information from morphemes and insert segments into metrical frames through a syllabification process in a strict order. Subsequently, phonetic encoding occurs, where articulatory motor programs for syllables are retrieved and prepared for articulatory output. Previous research has debated the role of syllables in spoken word production.

Syllables are natural articulatory units constrained by specific language structures and semantics [?, ?]. The cognitive mechanisms and temporal dynamics of syllable processing exhibit cross-linguistic differences during speech production. Studies on Chinese spoken word production have consistently observed a stable syllable priming effect, where naming latencies are shorter when prime and target stimuli share syllable overlap. Researchers argue that this effect indicates syllables are extracted first after lexical selection and serve as important functional units in phonological encoding [?, ?, ?, ?, ?, ?, ?, ?]. For Indo-European languages such as English and Dutch, syllable access and processing primarily occur during phonetic encoding, manifesting as syllable frequency effects [?, ?, ?, ?], where naming latencies are shorter for words with high-frequency syllables during picture naming or word reading tasks. Although syllable frequency effects in speech production have been extensively validated across languages [?, ?, ?, ?, ?], systematic reviews and cross-linguistic comparisons of research findings in this domain are lacking. This paper reviews and compares the cognitive mechanisms of syllable frequency effects in Indo-European languages and Chinese to deepen our understanding of cross-linguistic differences in syllable frequency effects during speech production.

2. Measurement of Syllable Frequency: Type Frequency and Token Frequency

Syllable frequency measurement is closely related to syllabic neighbours. In alphabetic languages, syllabic neighbours broadly refer to all words sharing a syllable at the same position regardless of which syllable position [?, ?], while narrowly defined neighbours refer to all words sharing the same initial syllable [?, ?]. For example, Spanish words *casa* (“house”), *caro* (“expensive”), *caja* (“box”), and *cama* (“bed”) are syllabic neighbours [?, ?]. As a tonal language with numerous homophones and near-homophones [?, ?, ?, ?], Chinese syllabic neighbours refer to all monosyllabic characters sharing the same syllable (tone not considered). Two measurement methods exist: syllabic neighbour count and cumulative word frequency of syllabic neighbours, termed type frequency and token frequency, respectively [?, ?]. Type frequency refers to the total number of syllabic neighbours for a target syllable, while token frequency represents the arithmetic sum of word frequencies of all syllabic neighbours. For instance, Chinese syllable */shua/* has syllabic neighbours “刷”, “唰”, and “耍”. Based on word frequencies from the Chinese word frequency database SUBTLEX-CH [?, ?], their frequencies are 16.4, 0.21, and 54.37 per million, respectively. Therefore, the type frequency of syllable */shua/* is 3, and its token frequency is 70.98 per million.

Most previous studies have only considered token frequency [?, ?, ?, ?, ?, ?, ?, ?, ?, ?]. Few studies have measured both types of frequency, and those that have failed to truly differentiate their effects. Researchers have simultaneously measured both frequency types but only to ensure selected syllables had comparable values, such that high-frequency syllables had high token and type frequencies while low-frequency syllables had low values, without examining whether the two frequency types produced different effects [?, ?, ?, ?]. Word token and type frequencies are closely related, making independent manipulation difficult [?, ?, ?], which complicates clear attribution of syllable frequency effects. Conrad et al. (2008) proposed that these two frequency types operate at different processing levels: token frequency of syllabic neighbours is easily influenced by extremely high-frequency neighbours, reflecting lexical-level competition among syllabic neighbours and the target word, and tends to produce inhibitory effects during lexical selection. In contrast, type frequency better reflects sublexical-level usage frequency and typicality of syllable representations themselves. Kwon (2014) tested this hypothesis using Korean word reading tasks. As a phonographic script similar to Indo-European languages with shallow orthographic depth, Korean word reading relies more on grapheme-to-phoneme conversion rules to access syllable representations directly and less on lexical-level information. Therefore, type frequency, which is sensitive to syllable representations, was expected to show facilitatory effects, while token frequency should show no effect. However, results showed facilitatory effects for both type and token frequencies, leading the researcher to conclude that both reflect speakers' processing of syllable representations at the sublexical level. The different directional effects of the two measurement methods suggest they influence spoken word production through different mechanisms.

3. Cognitive Mechanisms of Syllable Frequency Effects in Indo-European Languages

Using picture naming, word reading, or symbol-word association tasks, researchers have found shorter naming latencies for high-frequency syllable words compared to low-frequency syllable words, demonstrating the syllable frequency effect [?, ?, ?, ?]. Syllable frequency effects in speech production have been repeatedly validated across Indo-European languages including English [?, ?, ?, ?], French [?, ?, ?], Dutch [?, ?, ?, ?, ?], Spanish [?, ?, ?, ?], and German [?, ?, ?].

3.1 Manifestations and Processing Stages of Syllable Frequency Effects

Most studies examining Indo-European speech production have found facilitatory effects of token frequency [?, ?, ?, ?, ?, ?, ?, ?], though some have reported inhibitory effects [?, ?, ?, ?, ?] or null effects [?, ?, ?]. The different directional effects of syllable frequency on speech production may stem from its differential impact on two processing stages [?, ?, ?, ?]. During lexical selection in lexical access, high-frequency syllable words activate larger initial candidate sets because they have more syllabic neighbours, and these sets are more likely to contain high-frequency words. The activated lexical nodes compete with each other and interfere with target word access, producing inhibitory effects. During phonetic encoding, speakers retrieve articulatory motor programs from a mental syllabary. The mental syllabary is a repository of syllables containing abstract phonological syllables and context-dependent phonetic syllables (i.e., articulatory motor programs). During phonetic encoding, phonological syllables serve as input signals to the mental syllabary, mapping onto stored articulatory motor programs that are then output to the articulatory system [?, ?, ?, ?, ?]. Since articulatory motor programs for high-frequency syllables can be retrieved faster from the mental syllabary, high-frequency syllable words are named more quickly, showing facilitatory effects (see Figure 1a [Figure 1: see original paper]).

Researchers have found that syllable frequency typically produces inhibitory effects in lexical decision tasks [?, ?, ?, ?, ?, ?, ?] and facilitatory effects in pseudoword naming tasks [?, ?, ?, ?, ?, ?], providing support for the view that inhibitory and facilitatory effects occur during lexical selection and post-lexical processing stages (including phonological and phonetic encoding), respectively. Lexical decision tasks require judging whether stimuli are real words or pseudowords, a process that does not involve speech output, so syllable frequency primarily exerts inhibitory effects at the lexical selection stage. Compared to low-frequency syllables, high-frequency syllabic neighbours produce stronger activation that competes with the target word, interfering with lexical decisions and producing inhibitory effects. Pseudoword naming tasks, however, do not involve lexical selection, so syllable frequency primarily operates during phonological and phonetic encoding stages, where high-frequency syllables have retrieval advantages that produce facilitatory effects. Hutzler et al. (2005) did not observe syllable frequency effects in real word naming tasks, suggesting that the inhibitory effect on lexical selection and facilitatory effect on speech output may have equal strength, resulting in mutual cancellation in reaction time measures.

Regarding the locus of facilitatory syllable frequency effects, debate exists over whether they occur during phonological or phonetic encoding. Two main research approaches have been employed. The first uses the additive-factor method [?, ?] to test for interactions between syllable frequency and other linguistic factors. Cholin and Levelt (2009) observed an interaction between syllable preparation effects and syllable frequency effects in Dutch using the implicit priming paradigm. Since syllable preparation effects encompass all word-form encoding stages (morphological encoding, phonological encoding, and phonetic encoding), an interaction between two factors suggests they may operate at the same stage in a serial processing sequence. The researchers therefore concluded that the mental syllabary likely resides at the phonetic encoding stage following whole-word access. The second approach uses tasks involving different speech production stages and compares whether syllable frequency effects appear across these tasks to infer their processing locus. Laganaro and Alario (2006) used immediate naming, delayed naming, and delayed naming with articulatory suppression tasks to examine the processing stage of syllable frequency effects in French. Immediate naming includes all speech production processes from conceptualization to articulation. Delayed naming does not require participants to name pictures immediately upon presentation; instead, they complete phonological and phonetic encoding before a cue appears, after which they execute the prepared articulatory motor program. Therefore, naming latencies in this task reflect only articulatory processing. Delayed naming with articulatory suppression requires participants to repeatedly articulate a specific sound (e.g., /da/) after picture onset and stop immediately upon seeing the cue to name the picture. Since repeated articulation occupies the articulatory loop [?, ?] and blocks phonetic encoding of the target word, naming latencies measured after the cue reflect both phonetic encoding and articulation stages. Syllable frequency effects were observed in immediate naming and delayed naming with articulatory suppression but not in delayed naming alone. Since both immediate naming and delayed naming with articulatory suppression include phonetic encoding processes while delayed naming does not, the contrast between tasks suggests syllable frequency effects occur during phonetic encoding. However, Croot et al. (2017) using a similar design in English found syllable frequency effects only in immediate naming. Based on the task comparisons, this suggests the effect does not occur during phonetic encoding or articulation but possibly during phonological encoding. Croot et al. (2017) noted that syllable frequency effects might also appear during phonetic encoding but were too weak to detect in the delayed naming with articulatory suppression task.

High temporal resolution techniques can precisely examine the time course of syllable frequency effects. Bürki et al. (2015) used event-related potentials to examine French pseudoword naming with high-frequency, low-frequency, and novel syllables (frequency = 0). They found differences in ERP waveforms and topographic patterns between high-frequency and low-frequency/novel syllables approximately 180–160 ms and 115–100 ms before articulation onset. According to meta-analyses of spoken word production time courses, this time window corresponds to phonetic encoding [?, ?, ?]. The ERP findings thus indicate that syllable frequency effects occur during phonetic encoding. In summary, most evidence supports that syllable frequency effects in Indo-European languages occur during phonetic encoding, though the possibility that they occur during phonological encoding cannot be ruled out. The facilitatory effects of syllable frequency may appear during phonological and/or phonetic encoding. Although English and French are both Indo-European languages, differences in their lexical-phonological relationships may influence the cognitive mechanisms of syllable frequency effects in spoken word production.

Researchers have also examined how syllable position characteristics affect syllable frequency effects in Indo-European languages. For disyllabic and polysyllabic words, positional syllable frequency refers to the number of times a syllable appears at a specific position in words (weighted by word frequency, [?, ?, ?]) or the cumulative word frequency of all words containing that syllable at the same position [?, ?, ?]. After controlling for potential effects of stress assignment patterns in disyllabic words, studies have found that syllable position influences the presence of syllable frequency effects, though results remain controversial. Levelt and Wheeldon (1994) observed syllable frequency effects only for second syllables in Dutch disyllabic real word naming, arguing that articulation is not output until all syllables in a word complete phonetic encoding, making processing speed advantages for high-frequency first syllables less apparent. However, other studies found syllable frequency effects only for first syllables in Spanish [?, ?] and Dutch [?, ?] disyllabic pseudoword naming, suggesting that articulation can begin once first syllable phonetic encoding is complete—that is, speakers retrieve the second syllable while articulating the first, making second syllables insensitive to frequency manipulation in rapid naming tasks. Cholin et al. (2011) proposed that these different speech retrieval strategies between studies may stem from lexicality of experimental materials. In real word production, speakers complete phonetic encoding for the entire word before articulation to convey meaning correctly, whereas for pseudowords without meaning requirements, speakers simply articulate them. Some English disyllabic word naming studies have found syllable frequency effects for both first and second syllables (pseudoword naming: [?, ?]; real word naming: [?, ?]), reflecting that English speakers may have larger planning scope at the phonetic encoding-articulation interface. Cholin et al. (2011) suggested that syllable boundary clarity affects articulatory planning scope across languages: compared to English, Dutch and Spanish have relatively clear syllable boundaries, allowing speakers to detect syllable boundaries earlier and access the mental syllabary more quickly. In rapid naming tasks, Dutch and Spanish speakers tend to adopt syllable-sized planning scope, while English speakers tend to adopt multi-syllable planning scope that crosses syllable boundaries due to relatively ambiguous syllable boundaries. Whether syllables at different positions in polysyllabic word production show frequency effects depends not only on different speech retrieval strategies for real words versus pseudowords but also on cross-linguistic differences in articulatory planning scope.

3.2 Theoretical Accounts

Researchers have debated the mechanisms underlying facilitatory syllable frequency effects in speech production, proposing mental syllabary theory [?, ?], mixed models [?, ?], and dual-route models [?, ?, ?].

Levelt (1993) proposed mental syllabary theory to explain facilitatory syllable frequency effects in alphabetic language speech production. According to this theory, syllable frequency effects arise from direct retrieval of articulatory motor programs stored in the mental syllabary. Since high-frequency syllables access the mental syllabary faster than low-frequency syllables, high-frequency syllable words are named more quickly [?, ?, ?, ?, ?]. Mental syllabary theory emphasizes retrieval efficiency: since most syllables are frequently used articulatory units, recalculating complete articulatory motor programs for each use would be computationally wasteful [?, ?]. Direct retrieval of stored programs from the mental syllabary substantially reduces computational load and enables rapid, fluent speech [?, ?, ?, ?].

Levelt and Wheeldon (1994) validated mental syllabary theory through syllable frequency effects in Dutch speech production, while noting that speakers can correctly pronounce pseudowords composed of non-existent but phonotactically legal syllables (e.g., *fliltirp*). Since the mental syllabary does not store such non-existent syllables (i.e., novel syllables), speakers cannot access them directly and must rely on alternative mechanisms. They therefore proposed a mixed model, assuming two processing routes in the mental syllabary: direct retrieval and online computation. Novel and extremely low-frequency syllables are computed online, while frequently used high-frequency syllables are stored in the mental syllabary for direct retrieval (see also [**?**, **?**]). Facilitatory syllable frequency effects arise because direct retrieval of high-frequency syllables is faster than online computation of low-frequency syllables. Ferrand et al. (1996), using French syllable counts as an example, noted that not all syllables are stored in the mental syllabary. Potential novel syllables in French could be assembled online: while approximately 6,000 syllables actually exist, up to 33,600 potential syllables could be generated by including syllables that violate French word formation rules but follow French phonotactic rules (e.g., /km lu/). Speakers need not store such numerous syllables in the mental syllabary and can generate correct pronunciations for phonotactically legal novel syllables through analogy. As in early speech acquisition where all syllables are effectively novel, speakers gradually develop retrieval mechanisms from the mental syllabary based on language use experience.

The dual-route model of phonetic encoding [**?**, **?**, **?**] similarly posits that high-frequency and low-frequency syllables rely on different processing mechanisms: high-frequency syllables are retrieved directly from the mental syllabary via a direct route, while low-frequency syllables are assembled online from segments via an indirect route (see also [**?**, **?**, **?**, **?**, **?**]). Evidence from initial segment exchange errors supports the view that segments are assembled online into syllables: in continuous speech, initial segments of two words exchange positions (e.g., *car park* misarticulated as /pa: ka:k/), indicating that segments can be independent representational units during speech assembly [**?**, **?**]. The dual-route model further hypothesizes that syllable retrieval is related to its resting activation level, with high-frequency syllables having higher resting activation and lower selection thresholds from the mental syllabary [**?**, **?**], making direct retrieval easier than online assembly of low-frequency syllables. As extensions and developments of mental syllabary theory, both mixed and dual-route models predict that speakers name high-frequency syllable words faster than low-frequency syllable words, showing facilitatory effects, though the underlying cognitive mechanisms differ.

3.3 Neural Mechanisms of Syllable Frequency Effects

Recent neuroimaging research has examined neural mechanisms of syllable frequency effects in Indo-European languages using various imaging techniques. Based on mixed and dual-route models, high-frequency and low-frequency syllables are associated with different cognitive processing mechanisms, suggesting different brain activation patterns when producing syllables of different frequencies. Hagoort et al. (1999) used PET to examine silent and overt reading of real words (containing more high-frequency syllables) and pseudowords (containing more low-frequency syllables) by German speakers. Since this study confounded syllable frequency and lexicality, researchers could only speculate that the supplementary motor area might be involved in accessing the mental syllabary for high-frequency syllables, while the left medial premotor cortex might participate in online assembly of low-frequency syllables.

fMRI studies have found that low-frequency syllables elicit stronger brain activation than high-frequency syllables, while the reverse contrast shows no significant differences, indicating that producing low-frequency syllables requires greater processing costs [**?**, **?**]. For example, Carreiras et al. (2006) found that low-frequency syllable words in Spanish word reading tasks significantly activated the left anterior insula compared to high-frequency syllable words, while high-frequency syllable words did not show specific brain activation compared to low-frequency syllables. Researchers suggested that stronger activation for low-frequency syllables might reflect slower access to stored motor programs, while high-frequency syllables facilitate articulatory motor planning, supporting mental syllabary theory. Similarly, Papoutsi et al. (2009) asked English speakers to listen to pseudowords with different syllable frequencies and repeat them after a 6-second delay, finding that low-frequency syllable words significantly activated dorsal premotor cortex, bilateral inferior frontal gyrus, and supplementary motor area, while no brain regions showed significant activation for high-frequency syllables compared to low-frequency syllables. However, differences in activation levels between high-frequency and low-frequency syllables cannot prove that their processing is controlled by different brain regions. Although high-frequency syllables did not elicit specific brain activation compared to low-frequency syllables, they may still differ in neural networks, but the time window of these differences may be too narrow for detection by low temporal resolution imaging techniques like fMRI.

Furthermore, Bürki et al. (2015) used ERP technology to examine French pseudoword naming with high-frequency, low-frequency, and novel syllables. They found differences in ERP waveforms and overall topographic patterns between high-frequency and low-frequency/novel syllables approximately 170–100 ms before articulation onset, while the latter two syllable types showed identical and stable electrophysiological patterns. This indicates that individuals rely on different neural mechanisms when producing high-frequency versus low-frequency/novel syllables, but similar mechanisms for low-frequency and novel syllables. Researchers interpreted these results as reflecting that the processing advantage for high-frequency syllables may stem from storage of their articulatory motor programs in the mental syllabary, while low-frequency syllable processing mechanisms are similar to those for novel syllables, with articulatory motor programs generated through online assembly. This provides support for the dual-route model of syllable frequency effects [**?**, **?**].

4. Cognitive Mechanisms of Syllable Frequency Effects in Chinese

4.1 The Role of Syllables in Chinese Spoken Word Production

As a tonal language with non-alphabetic writing, Chinese differs substantially from alphabetic languages. Alphabetic languages not only have numerous syllables (e.g., over 12,000 in Dutch and English) but also exhibit ambisyllabicity and resyllabification, where word-final consonants can cross word boundaries and recombine with word-initial vowels in continuous speech [?, ?]. In contrast, Chinese has few syllables—approximately 400 without considering tone and only about 1,200 with tone—and has clear syllable boundaries without resyllabification [?, ?]. Therefore, Chinese speakers can directly retrieve whole syllables from the mental lexicon during phonological encoding \cite{O' Seaghdha et al., 2010; Jiang et al., 2020; Zhang & Wang, 2020}.

Cross-linguistic differences exist in the role of syllables in spoken word production. For Indo-European languages, the processing unit first extracted during phonological encoding is the phoneme [?, ?, ?, ?, ?, ?, ?]. Phonemes are the smallest speech units that distinguish meaning, comprising vowels and consonants [?, ?]. According to the WEAVER++ model, speakers then construct syllables through a syllabification process to complete phonetic encoding. In Chinese, numerous studies using the implicit priming paradigm [?, ?], masked priming paradigm [?, ?, ?], and picture-word interference paradigm [?, ?, ?] have found that syllables are phonological encoding units in Chinese speech production. ERP studies comparing the time courses of syllable and phoneme effects have found that syllable extraction precedes phoneme extraction in Chinese speech production, with syllables functioning during phonological encoding while phoneme effects appear mainly in late phonological encoding or phonetic encoding stages [?, ?, ?, ?, ?], though some studies have observed syllable effects in both phonological and phonetic encoding stages [?, ?]. Notably, Jiang et al. (2020) used time-frequency analysis to find theta-band (4–8 Hz) activity changes reflecting syllable effects rather than phoneme effects during phonological encoding, confirming from neural oscillation perspectives that syllables are phonological encoding units in Chinese speech production.

Although some studies have found phoneme activation during Chinese speech production [?, ?, ?, ?], evidence for phoneme processing in Chinese remains limited and controversial [?, ?]. Current research primarily confirms the important role of syllables in Chinese phonological encoding.

Based on different findings from Indo-European languages and Chinese, researchers proposed the proximate units principle to explain cross-linguistic differences in phonological encoding units during speech production \cite{O'Seaghdha et al., 2010}. Proximate units refer to phonological processing units selected first after morpheme activation: phonemes in Indo-European languages and syllables in Chinese. For Indo-European languages, speakers first extract phonemes, then combine them with metrical frames through syllabification, and subsequently retrieve articulatory motor programs from the mental syllabary during phonetic encoding. In contrast, Chinese phonological encoding involves first extracting syllables, then decomposing them into segmental information (phonemes or subsyllabic units) and suprasegmental information (tone) before completing phonetic encoding and articulation (see Figure 2 [Figure 2: see original paper]). Roelofs (2015) validated this hypothesis through computational simulation.

4.2 Manifestations and Processing Stages of Syllable Frequency Effects in Chinese

Existing studies on Chinese syllable frequency effects have primarily used picture naming tasks requiring speakers to name pictures with monosyllabic words [?, ?, ?]. Most studies have observed facilitatory effects of token frequency [?, ?, ?, ?] while neglecting type frequency. Only one study [?, ?] using picture naming and single-character naming tasks found that when manipulating both token and type frequencies simultaneously, picture naming showed facilitatory effects while single-character naming showed inhibitory effects. When type frequency was controlled while manipulating token frequency, facilitatory effects persisted in picture naming but no effects were found in single-character naming. As a typical spoken word production task, the facilitatory syllable frequency effect in picture naming is consistent with Indo-European language findings, and research further indicates that token frequency changes primarily drive this facilitatory effect. Inhibitory effects in single-character naming may arise because high-frequency syllable words activate more syllabic neighbours that interact through parallel distributed processing [?, ?] and interfere with the target word, resulting in longer naming latencies for high-frequency syllable words [?, ?]. Additionally, single-character naming tasks involve both visual word recognition and speech production processes including phonological encoding, phonetic encoding, and articulation. The inhibitory effects observed may result from cancellation between inhibitory effects from visual word access and facilitatory effects from phonological and phonetic encoding.

This finding indicates that syllable type frequency and token frequency play different roles in Chinese spoken word production, though current research cannot clearly articulate their specific functions.

As previously mentioned, based on the proximate units principle and research on Chinese speech production, Chinese speakers extract syllable information during phonological encoding \cite{Cai et al., 2020; Feng et al., 2019; O' Seaghdha et al., 2010; Zhang & Damian, 2019}. Therefore, Chinese syllable frequency effects likely occur during phonological encoding. At the behavioral level, Zhang and Wang (2014) using Chinese picture naming tasks found no interaction between word frequency and syllable frequency, indicating that these two factors independently affect naming processes, consistent with findings in Indo-European languages [?, ?]. Some studies using the picture-word interference paradigm observed interactions between syllable frequency and phonological relatedness [?, ?, ?]. Since phonological relatedness-induced syllable facilitation effects occur during phonological encoding in Chinese speech production [?, ?], this result suggests that Chinese syllable frequency effects likely also occur during phonological encoding. However, other studies using the same paradigm failed to replicate these results, suggesting that Chinese syllable frequency effects may not occur during phonological encoding but possibly during phonetic encoding [?, ?].

Regarding temporal dynamics, Yang (2017) using ERP technology found that syllable frequency effects in young adults appeared primarily between 250–350 ms. According to meta-analyses of spoken word production time courses, this time window corresponds to phonological encoding [?, ?, ?], further confirming that Chinese syllable frequency effects occur during phonological encoding. In summary, although the possibility that syllable frequency affects Chinese phonetic encoding cannot be ruled out, most evidence supports that Chinese syllable frequency effects occur during phonological encoding (see Figure 1b), providing support for the important role of syllables in this stage of Chinese speech production.

Based on existing findings, we propose a model of syllable processing mechanisms in Chinese spoken word production (Figure 1b) to provide a theoretical foundation for future research. During lexical selection, the target word competes with syllabic neighbours. Compared to low-frequency syllables, high-frequency syllables have more highly activated neighbours that produce lexical-level inhibitory effects. During phonological encoding, speakers extract syllables for phonological processing, with high-frequency syllables retrieved faster than low-frequency syllables, producing facilitatory effects. During phonetic encoding, speakers encode articulatory motor programs for selected target words, a process that must consider syllable-tone combinations. Articulatory motor programs for high-frequency syllables are also retrieved faster than those for low-frequency syllables due to repeated use, potentially producing facilitatory effects at this stage as well. Whether the final articulatory output shows facilitatory or inhibitory syllable frequency effects depends on the magnitude of inhibitory and facilitatory effects at each processing stage and the relationships between stages (independent or interactive). Current research on Chinese syllable frequency effects is relatively limited, and future studies should employ multiple experimental tasks and techniques to investigate characteristics and cognitive mechanisms of Chinese syllable frequency effects at behavioral, electrophysiological, and neural levels, with cross-linguistic comparisons.

5. Syllables from a Cross-Linguistic Perspective: Summary and Outlook

Syllable frequency effects are widely present in human language processing. Researchers have systematically explored cognitive mechanisms of syllable frequency effects in Indo-European languages and Chinese during speech production using picture naming, word reading, pseudoword naming, and picture-word interference paradigms. Cross-linguistic differences in syllable frequency effects primarily manifest in two aspects: expression patterns and processing stages. Regarding expression patterns, Indo-European languages show facilitatory, inhibitory, or null syllable frequency effects, while Chinese shows primarily facilitatory effects. The direction of syllable frequency effects is closely related to the frequency index manipulated (type or token frequency) and the task employed. Regarding processing stages, although both Indo-European and Chinese studies debate the locus of syllable frequency effects, theoretical and empirical evidence primarily supports that syllable frequency effects in Indo-European spoken word production occur during phonetic encoding, while Chinese syllable frequency effects likely occur during phonological encoding. Cross-linguistic differences in syllable frequency effects may relate to inherent properties of different language systems, providing new perspectives on the role of syllables at different processing stages in alphabetic versus Chinese speech production.

According to the proximate units principle, the phonological encoding unit in Chinese is the syllable [?, ?, ?, ?, ?, ?, ?, ?]. Chinese has few syllables, so speakers can directly retrieve syllables during phonological encoding, with high-frequency syllables retrieved faster than low-frequency syllables, producing facilitatory effects. Interactions between syllable frequency and phonological relatedness in picture-word interference paradigms provide supporting evidence that syllable frequency effects occur during phonological encoding [?, ?, ?, ?]. Based on characteristics of phonetic encoding stages, high-frequency syllables are retrieved faster during phonological encoding, and their articulatory motor program encoding and retrieval are also faster than for low-frequency syllables. We therefore propose that syllable frequency effects may also occur during phonetic encoding. Given the current state of research, Chinese syllable frequency effects require systematic and in-depth investigation.

Specifically, future research can explore the following directions: First, further investigation is needed on the roles of type frequency and token frequency. Whether these two measurement indices play different roles across word production tasks and whether their effects show cross-linguistic differences remain open questions. Second, the cognitive mechanisms of facilitatory syllable frequency effects require further exploration. Current mainstream views suggest that syllable processing operates through the mental syllabary [?, ?], with mixed and dual-route models further developing this theory and proposing separate mechanisms for high-frequency and low-frequency syllables. However, both models predict faster naming for high-frequency than low-frequency syllable words, making it impossible to determine whether low-frequency syllables are stored in the mental syllabary or assembled online based solely on syllable frequency effects. Apart from Bürki et al. (2015), who supported the dual-route model by demonstrating dissociated electrophysiological patterns for high-frequency and low-frequency syllables, no other studies have directly tested mixed and dual-route models. Given that Chinese phonological system characteristics differ from Indo-European languages, the mechanisms of Chinese syllable frequency effects may be entirely different and require more systematic exploration. Third, current research on Chinese syllable frequency effects is relatively scarce, focusing only on monosyllabic words while neglecting disyllabic or polysyllabic words. Consequently, evidence is lacking on how syllable position affects Chinese syllable frequency effects. Moreover, studies examining processing stages of Chinese syllable frequency effects have used relatively single paradigms (primarily picture-word interference). Future research should employ diverse speech production tasks to clarify processing stages and neural mechanisms of Chinese syllable frequency effects at behavioral, electrophysiological, and neural levels, providing deeper understanding of syllable roles in Chinese speech production and cross-linguistic differences in cognitive and neural mechanisms of syllable frequency effects. Fourth, consider Chinese language acquisition from the perspective of syllable frequency. Due to numerous homophones and near-homophones in Chinese, particularly for syllables with high type frequency, hearing a single syllable often makes it difficult to access a specific character' s semantics without contextual support. Therefore, Chinese learners typically need to memorize characters' phonological forms holistically at the syllable level rather than relying on grapheme-to-phoneme conversion rules as in alphabetic languages. This suggests that Chinese beginners could use strategies such as writing or overlearning to distinguish confusing syllabic neighbours among homophones, gradually reducing interference from confusing neighbours during lexical selection and achieving facilitatory effects in phonological construction for high-frequency syllable words.

References

Duanmu, S. (2021). Analysis and quantitative comparison of English and Chinese syllables. *Linguistic Sciences*, 20(6), 561—588.

Jiang, Y. (2021). The cognitive mechanisms of how cumulative syllable frequency and number of neighbors influence Chinese spoken word production (Master' s thesis). Renmin University of China, Beijing.

Jiang, Y., Cai, X., & Zhang, Q. (2020). Theta-band (4-8 Hz) activity reflects syllable information processing in Chinese spoken word production. *Acta Psychologica Sinica*, 52(10), 1199—1211.

Ouyang, M. (2020). The cognitive and neural mechanisms of semantic and phonological retrieval aging in Chinese word production (Doctoral dissertation). Renmin University of China, Beijing.

Qu, Q., Liu, W., & Li, X. (2018). Phonological processing units in Chinese speech production: A phoneme-based study. *Advances in Psychological Science*, 26(9), 1535—1545.

Shi, Y. (2019). Toned syllables: A starting point for explaining Chinese. *Journal of East China Normal University (Philosophy and Social Sciences Edition)*, 51(3), 1—17.

Yang, Q. (2017). The cognitive aging mechanisms of lexical access in Chinese spoken word production (Doctoral dissertation). Renmin University of China, Beijing.

Yang, Q., & Zhang, Q. (2015). Cognitive aging of word frequency effect, syllable frequency effect, and phonological facilitation effect in spoken word production. *Journal of Psychological Science*, 38(6), 1303–1310.

Yue, Y., & Zhang, Q. (2015). Facilitatory and inhibitory effects of syllables and segments in Chinese spoken word production. *Acta Psychologica Sinica*, 47(3), 319–328.

Zhang, Q. (2005). The role of syllables in language production. *Advances in Psychological Science*, 13(6), 752–759.

Zhang, Q., & Wang, X. (2020). Phonological encoding units in Chinese spoken word production: An ERP study using the implicit priming paradigm. *Acta Psychologica Sinica*, 52(4), 414–425.

Zhang, Q., & Yang, Y. (2005). Phonological encoding units in Chinese monosyllabic word production. *Journal of Psychological Science*, 28(2), 374–378.

Álvarez, C. J., Carreiras, M., & Taft, M. (2001). Syllables and morphemes: Contrasting frequency effects in Spanish. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(2), 545–555.

Baddeley, A., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology Section A*, 36(2), 233–252.

Bürki, A. (2018). Variation in the speech signal as a window into the cognitive architecture of language production. *Psychonomic Bulletin & Review*, 25, 1973–2004.

Bürki, 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), Article e10729. https://doi.org/10.1371/journal.pone.0010729

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*, 140, Article 107382. https://doi.org/10.1016/j.neuropsychologia.2020.107382

Carreiras, M., Álvarez, C. J., & De Vega, M. (1993). Syllable frequency and visual word recognition in Spanish. *Journal of Memory and Language*, 32(6), 766–780.

Carreiras, M., Mechelli, A., & Price, C. J. (2006). Effect of word and syllable frequency on activation during lexical decision and reading aloud. *Human Brain Mapping*, 27(12), 963–972.

Carreiras, M., & Perea, M. (2004). Naming pseudowords in Spanish: Effects of syllable frequency. *Brain and Language*, 90(1−3), 393−400.

Chetail, F., & Mathey, S. (2011). Effect of syllabic neighbourhood as a function of syllabic neighbour length. *Psychonomic Bulletin & Review*, 18(5), 951−957.

Cholin, J. (2008). The mental syllabary in speech production: An integration of different approaches and domains. *Aphasiology*, 22(11), 1127−1141.

Cholin, J., Dell, G. S., & Levelt, W. J. (2011). Planning and articulation in incremental word production: Syllable-frequency effects in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1), 109−122.

Cholin, J., & Levelt, W. J. (2009). Effects of syllable preparation and syllable frequency in speech production: Further evidence for syllabic units at a post-lexical level. *Language and Cognitive Processes*, 24(5), 662−684.

Cholin, J., Levelt, W. J., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition*, 99(2), 205−235.

Clark, J., & Yallop, C. (1995). *An Introduction to Phonetics and Phonology* (2nd ed.). Oxford: Blackwell.

Conrad, M., Carreiras, M., & Jacobs, A. M. (2008). Contrasting effects of token and type syllable frequency in lexical decision. *Language and Cognitive Processes*, 23(2), 296−326.

Conrad, M., & Jacobs, A. (2004). Replicating syllable frequency effects in Spanish in German: One more challenge to computational models of visual word recognition. *Language and Cognitive Processes*, 19(3), 369−390.

Conrad, M., Grainger, J., & Jacobs, A. M. (2007). Phonology as the source of syllable frequency effects in visual word recognition: Evidence from French. *Memory & Cognition*, 35(5), 974−983.

Conrad, M., Stenneken, P., & Jacobs, A. M. (2006). Associated or dissociated effects of syllable frequency in lexical decision and naming. *Psychonomic Bulletin & Review*, 13(2), 339−345.

Croot, K., Lalas, G., Biedermann, B., Rastle, K., Jones, K., & Cholin, J. (2017). Syllable frequency effects in immediate but not delayed syllable naming in English. *Language, Cognition and Neuroscience*, 32(9), 1119−1132.

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.

den Hollander, J., Jonkers, R., Mariën, P., & Bastiaanse, R. (2019). Identifying the speech production stages in early and late adulthood by using electroencephalography. *Frontiers in Human Neuroscience*, 13, Article 298. https://doi.org/10.3389/fnhum.2019.00298

Farrell, M. T., & Abrams, L. (2014). Picture−word interference reveals inhibitory effects of syllable frequency on lexical selection. *Quarterly Journal of Experimental Psychology*, 67(3), 525−541.

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), Article 11773. https://doi.org/10.1038/s41598-019-48033-3

Ferrand, L., Segui, J., & Grainger, J. (1996). Masked priming of word and picture naming: The role of syllabic units. *Journal of Memory and Language*, 35(5), 708−723.

Hagoort, P., Indefrey, P., Brown, C., Herzog, H., Steinmetz, H., & Seitz, R. J. (1999). The neural circuitry involved in the reading of German words and pseudowords: A PET study. *Journal of Cognitive Neuroscience*, 11(4), 383−398.

Hutzler, F., Conrad, M., & Jacobs, A. M. (2005). Effects of syllable-frequency in lexical decision and naming: An eye-movement study. *Brain and Language*, 92(2), 138−152.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, 2, Article 255. https://doi.org/10.3389/fpsyg.2011.00255

Indefrey, P. & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92, 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.

Kwon, Y. (2014). The syllable type and token frequency effect in naming task. *Korean Journal of Cognitive Science*, 25(2), 91−107.

Laganaro, M. (2019). Phonetic encoding in utterance production: A review of open issues from 1989 to 2018. *Language, Cognition and Neuroscience*, 34(9), 1193−1201.

Laganaro, M., & Alario, F. X. (2006). On the locus of the syllable frequency effect in speech production. *Journal of Memory and Language*, 55(2), 178−196.

Laganaro, M., Python, G., & Toepel, U. (2013). Dynamics of phonological−phonetic encoding in word production: Evidence from diverging ERPs between stroke patients and controls. *Brain and Language*, 126(2), 123−132.

Levelt, W. J. M. (1993). Timing in speech production with special reference to word form encoding. *Annals of the New York Academy of Sciences*, 682, 283−295.

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

Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, 50(1—3), 239—269.

Macizo, P., & Van Petten, C. (2007). Syllable frequency in lexical decision and naming of English words. *Reading and Writing*, 20(4), 295—331.

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.

Miguel-Abella, R. S., Pérez-Sánchez, M. Á., Cuetos, F., Marín, J., & González-Nosti, M. (2022). SpaVerb-WN—A megastudy of naming times for 4562 Spanish verbs: Effects of psycholinguistic and motor content variables. *Behavior Research Methods*, 54(6), 2640—2664.

Onochie-Quintanilla, E., Defior, S. A., & Simpson, I. C. (2019). RAN and orthographic processing: What can syllable frequency tell us about this relationship?. *Journal of Experimental Child Psychology*, 182, 1—17.

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.

Papoutsi, M., de Zwart, J. A., Jansma, J. M., Pickering, M. J., Bednar, J. A., & Horwitz, B. (2009). From phonemes to articulatory codes: An fMRI study of the role of Broca' s area in speech production. *Cerebral Cortex*, 19(9), 2156—2165.

Perea, M., & Carreiras, M. (1998). Effects of syllable frequency and syllable neighbourhood frequency in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 24(1), 134—144.

Perret, C., Schneider, L., Dayer, G., & Laganaro, M. (2014). Convergences and divergences between neurolinguistic and psycholinguistic data in the study of phonological and phonetic encoding: A parallel investigation of syllable frequency effects in brain-damaged and healthy speakers. *Language, Cognition and Neuroscience*, 29(6), 714—727.

Qu, Q. 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. Q., Feng, C., Hou, F. Y., & Damian, M. F. (2020). Syllables and phonemes as planning units in Mandarin Chinese spoken word production: Evidence from ERPs. *Neuropsychologia*, 146, Article 107559. https://doi.org/10.1016/j.neuropsychologia.2020.107559

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

Schiller, N. O. (2008). The masked onset priming effect in picture naming. *Cognition*, 106(2), 952—962.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96(4), 523—568.

Stenneken, P., Conrad, M., & Jacobs, A. M. (2007). Processing of syllables in production and recognition tasks. *Journal of Psycholinguistic Research*, 36(1), 65—78.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 30, 276—315.

Tremblay, P., Deschamps, I., Bédard, P., Tessier, M. H., Carrier, M., & Thibeault, M. (2018). Aging of speech production, from articulatory accuracy to motor timing. *Psychology and Aging*, 33(7), 1022—1034.

Varley, R., & Whiteside, S. P. (2001). What is the underlying impairment in acquired apraxia of speech?. *Aphasiology*, 15(1), 39—49.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, 9(4), 325—329.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40(3), 374—408.

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—5822.

Whiteside, S. P., & Varley, R. A. (1998). A reconceptualisation of apraxia of speech: A synthesis of evidence. *Cortex*, 34(2), 221—231.

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), Article e46595. https://doi.org/10.1371/journal.pone.0046595

Yu, M. X., Mo, C., & Mo, L. (2014). The role of phoneme in Mandarin Chinese production: Evidence from ERPs. *PLoS ONE*, 9(9), Article e106486. https://doi.org/10.1371/journal.pone.0106486

Zhang, Q. F., & Damian, M. F. (2019). Syllables constitute proximate units for Mandarin speakers: Electrophysiological evidence from a masked priming task. *Psychophysiology*, 56(4), Article e13317. https://doi.org/10.1111/psyp.13317

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, Article 120. https://doi.org/10.3389/fpsyg.2014.00120

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