Bilingualism: Language and Cognition

cambridge.org/bil

The effect of proficiency on phonological encoding in L2 speech production

Man Wang[1], Shuai Liu[1], Jiahuan Zhang[1] and Niels O. Schiller[2,3]

[1]School of Foreign Languages, Qingdao University, Qingdao, China; [2]Department of Linguistics and Translation, City University of Hong Kong, Hong Kong SAR, China and [3]Center for Linguistics, Leiden University, Leiden, The Netherlands

Research Article

Cite this article: Wang, M., Liu, S., Zhang, J. and Schiller, N.O. (2025). The effect of proficiency on phonological encoding in L2 speech production. *Bilingualism: Language and Cognition* 1–10. https://doi.org/10.1017/S1366728925000276

Received: 13 August 2024
Revised: 16 February 2025
Accepted: 18 February 2025

Keywords:
phonological encoding; Mandarin Chinese-English bilinguals; speech production; picture-word interference paradigm

Corresponding author:
Niels O. Schiller;
Email: nschille@cityu.edu.hk

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

Abstract

During speech production, bilinguals need to encode target words phonologically before articulation, and the encoding units differ across languages. It remains an open question whether bilinguals employ the encoding unit in their L1 or L2 for phonological encoding. The present study examined the primary unit of phonological encoding in L2 speech production by Mandarin Chinese-English bilinguals with high and low L2 proficiency using the picture-word interference paradigm. Results revealed segmental priming effects with one or two segments and syllabic overlap at varied stimulus onset asynchronies (SOAs), for both groups in their L2 speech production. Additionally, the results demonstrated increasing effects with more overlapping segments for both groups, and the facilitation effects decreased as SOA increased. These results indicate that bilinguals encode English words with the segment as a primary planning unit regardless of their L2 proficiency. The time course of segmental encoding in L2 production is also discussed.

Highlights

- Mandarin-English bilinguals use segments as the primary unit in L2 phonological encoding.
- The encoding unit is the same for high- and low- proficient bilinguals.
- The segmental effects increase with more overlapping segments.
- The segmental effects decrease as stimulus onset asynchronies increase.

1. Introduction

Speech production is a skilled cognitive action to convey thoughts via audible sounds. During speech production, speakers need to go through different stages, that is, conceptual preparation, lexical selection, phonological encoding, and articulation (e.g., Dell, 1986; Levelt et al., 1999). Abstract lexical information is transcoded into physical speech sounds during phonological encoding. Dysfunction at this stage is one of the main reasons that cause anomia in aphasic patients (e.g., Calabria et al., 2020; Schwartz, 2014) and tip-of-the-tongue instances in healthy speakers (e.g., Sadat et al., 2014).

It is generally agreed upon that segments are the primary phonological encoding units of spoken word production in Indo-European languages (e.g., Damian & Dumay, 2007, 2009; O'Seaghdha et al., 2010 for English; Roelofs, 1999 for Dutch). For instance, if a speaker plans to say the word "monkey," the segments /m/, /ʌ/, /ŋ/, /k/, /i/ will be retrieved, respectively, as well as its metrical framework (i.e., a disyllabic structure with lexical stress on the first syllable). After accessing the set of segments and the corresponding metrical frame, the segmental information is inserted into the metrical frame in a rightward incremental fashion to construct the syllables [ˈmʌŋ.ki] (syllable boundaries indicated by dots; e.g., Cholin et al., 2004; Meyer & Schriefers, 1991; Roelofs, 2015; Wheeldon & Levelt, 1995; see Figure 1).

In the form preparation paradigm, speakers generally respond faster in a segment-homogeneous condition compared to a heterogeneous condition (Alario et al., 2007; Damian & Bowers, 2003; Jacobs & Dell, 2014; Meyer, 1991). This suggests that speakers can prepare overlapping segments. Further evidence for the segment as the encoding unit has also been reported in other speech production paradigms, such as in the picture-word interference paradigm (e.g., Damian & Martin, 1999; Meyer & Schriefers, 1991) and the masked priming paradigm (e.g., Forster & Davis, 1991; Malouf & Kinoshita, 2007; Schiller, 1998, 2000).

However, for Mandarin Chinese, studies found that the primary phonological encoding units in speech production are more likely to be syllables instead of segments. Studies using various paradigms have demonstrated that syllabic overlap (e.g., "鼻 /bi^2/" and "笔 /bi^3/") instead of segmental overlap (e.g., "鼻 /bi^2/" and "布 /bu^4/") significantly affects speech production in Mandarin Chinese (e.g., masked priming paradigm, Cai et al., 2020; Chen et al., 2003; Chen et al., 2016; Zhang & Damian, 2019; picture-word interference paradigm, Zhang & Yang, 2005; picture naming paradigm, You et al., 2012). Please note that phonemic effects were observed

CAMBRIDGE UNIVERSITY PRESS

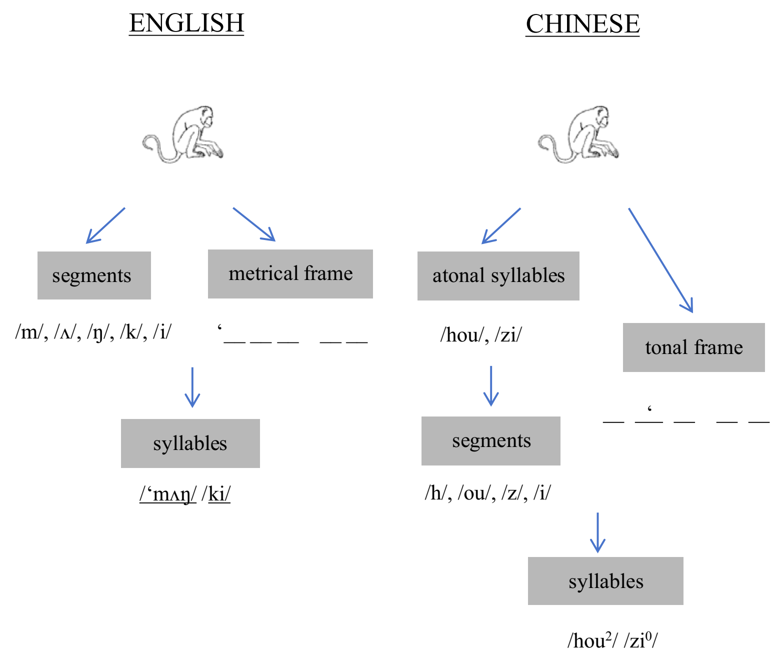

Figure 1. Model of phonological encoding for English and Mandarin Chinese (adapted from Schiller, 2006, and Zhang et al., 2018). The apostrophe marks the stress position in English and the number marks the lexical tone in Mandarin Chinese, with "2" indicating a rising tone.

in ERPs (see Cai et al., 2020; Qu et al., 2012), which were suggested to reflect a phonemic encoding stage after syllabic encoding (Cai et al., 2020). O'Seaghdha et al. (2010) proposed the *proximate units principle* to explain differences in phonological encoding units across languages. With this principle, O'Seaghdha et al. (2010) refer to the proximate units as the primary phonological encoding units, that is, the first explicitly selectable phonological production units. According to this principle, the primary phonological encoding units have cross-linguistic variations. Specifically, segments are claimed to be the primary phonological encoding units in Indo-European languages (e.g., O'Seaghdha et al., 2010; Roelofs, 1999) but syllables in Chinese (e.g., Cai et al., 2020; Zhang & Damian, 2009; Zhang & Yang, 2005; see Figure 1).

With such cross-linguistic differences, researchers have been drawn to the mechanisms of phonological encoding in bilinguals. It is believed that bilinguals have shared lexical representations across languages (e.g., Macizo, 2016), although there are disputes over whether a non-target language's phonological form is activated in speech production of bilinguals (see, e.g., Costa et al., 1999; De Bot, 1992; Green, 1998; Poulisse & Bongaerts, 1994 for the *Language-Specific Phonological Activation* account, see Costa, 2005 for a review; and see, e.g., Macizo, 2016; Nakayama et al., 2014; Spalek et al., 2014; Thierry & Wu, 2004; Xu et al., 2021; Zhang et al., 2021 for the *Language Non-specific Phonological Activation* account). In second language (L2) speech production, bilinguals may recruit the processing mechanisms of their native language (i.e., L1) to produce L2, leading to the assimilation hypothesis (e.g., Liu et al., 2023; Xin et al., 2020) or recruit addition neural networks to accommodate L2 processing, leading to the accommodation hypothesis (e.g., Cao et al., 2013), respectively.

In the phonological encoding stage of L2 speech production, it remains unresolved whether Mandarin Chinese-English bilinguals are influenced by their native language (i.e., syllables as primary units) or conform to L2 (i.e., segments as primary units). Previous studies have shown discrepancies in terms of the primary phonological encoding units in L2 speech production (e.g., Li et al., 2017; Timmer & Chen, 2017; Verdonschot et al., 2013; Wang et al., 2021; Xin et al., 2020). For instance, using a colored picture-naming task where participants produced noun phrases (e.g., 藍駱駝, /laam^4/ /lok^{3}to^4/, "blue camel"), Timmer and Chen (2017) reported a (onset) segment priming effect for Dutch-Cantonese bilinguals in their L2 (i.e., Cantonese), whose phonological encoding units are believed to be larger than the phoneme (e.g., Wong et al., 2012). Their results indicate that Dutch-Cantonese bilinguals employed the L1 (i.e., Dutch) phonological encoding units to encode their L2. However, Xin et al. (2020) reported syllabic priming effects for English-Mandarin Chinese bilinguals when they named pictures in L1 or L2 in the picture-word interference paradigm, suggesting that they relied on the same phonological encoding units as Mandarin Chinese native speakers. Xin et al. (2020) explained this inconsistency was caused by the language environment in which the experiments were carried out (see, Li & Wang, 2017 for the influence of language environment on L1 phonological encoding). Specifically, participants whose daily language environment is Mandarin Chinese use the same phonological encoding units as native Mandarin Chinese speakers when they produce L2-Mandarin Chinese.

The study by Li et al. (2015) suggests that tasks that explicitly require orthographic information processing, such as associative naming cued by visually presented prompt words, encourage participants to employ different phonological encoding units in their L1 production. Nevertheless, in the two studies above (i.e., Timmer & Chen, 2017; Xin et al., 2020), although orthographic information processing is not required in the picture naming tasks in both studies, participants use different phonological encoding units in L2 production. Therefore, the cross-task differences cannot completely explain the discrepant findings in Mandarin Chinese and Indo-European languages.

Furthermore, differences in L2 proficiency may contribute to different processing mechanisms of phonological encoding during

https://doi.org/10.1017/S1366728925000276 Published online by Cambridge University Press

L2 speech production. It is suggested that the degree to which bilinguals inhibit the non-response language is dependent on their L2 proficiency (e.g., Costa et al., 2003; Costa & Santesteban, 2004; Guo & Peng, 2006; Nakayama et al., 2016; see Jiao et al., 2020 for a review of executive control to manage bilingual processing), and thus high and low proficiency bilinguals may demonstrate differences in response times in speech production (e.g., Dash & Kar, 2020; De Bot, 2004; Macizo, 2016). For instance, Nakayama et al. (2016) recruited Japanese-English bilinguals with high or low L2 (i.e., English) proficiency and asked them to read aloud English words preceded by masked primes that overlapped in just the onset segment (e.g., bark-BENCH) or the onset segment plus the following vowel corresponding to the mora-sized units CV (consonant + vowel; e.g., bell-BENCH). Participants demonstrated different phonological encoding units in L2 (i.e., English) spoken word production, that is, high proficiency Japanese-English bilinguals showed a significant onset segment priming effect while the low proficiency group showed CV priming, indicating that high proficiency bilinguals used segments as the primary phonological encoding units while low proficiency bilinguals used the mora-sized units CV (Nakayama et al., 2016).

Similar findings were also reported by Verdonschot et al. (2013) who used a masked priming-naming task to investigate Mandarin Chinese-English bilinguals' L2 speech production. They found that bilinguals with high L2 proficiency showed a significant masked onset segment priming effect in L2 production, employing the same phonological encoding units (i.e., segments) as English native speakers did. The results of Nakayama et al. (2016) and Verdonschot et al. (2013) suggest that the primary phonological encoding units produced by high-proficiency bilinguals were accommodated to their L2, even when their language environment is not L2. This finding also contradicts that of Timmer and Chen (2017) who found that bilinguals' L2 phonological encoding units were assimilated to their L1. However, the study of Verdonschot et al. (2013) did not include bilinguals with low L2 proficiency, but the results of such participants are necessary to resolve the discrepancy.

Given the cross-linguistic differences in primary phonological encoding units as well as the influence of L2, it is necessary to resolve the discrepancies over the primary phonological encoding units in L2, especially with varied L2 proficiency. Therefore, we aim to investigate the primary phonological encoding units in the L2 production of Mandarin Chinese-English bilinguals with high and low L2 proficiency who are not immersed in L2, to avoid possible influence from the language environment (see, e.g., Li & Wang, 2017; Xin et al., 2020). The present study addresses the following research questions: (1) What are the primary phonological encoding units of Mandarin Chinese-English bilinguals when they utter L2? More specifically, will the bilinguals encode L2 words using L1 units or L2 units? (2) Are there any differences between high and low-proficiency Mandarin Chinese-English bilinguals in terms of the primary phonological encoding units? Based on previous research in Japanese (Nakayama et al., 2016), we hypothesize that Mandarin Chinese-English bilinguals with high L2 proficiency use segments as the primary phonological encoding units in L2 speech production, whereas Mandarin Chinese-English bilinguals with low L2 proficiency use syllables as the primary phonological encoding units.

2. Methods

2.1. Participants

Two groups of native Mandarin Chinese speakers differing in their English proficiency participated in this study. They were recruited from a university in Northern China. All participants were right-handed, with normal or corrected-to-normal vision. The students were paid for their participation and signed an informed consent letter. The high L2 proficiency group (Group 1) consisted of 30 students majoring in English (2 males; average age = 22 years; *SD* = 1.91 years). All of them passed the Test for English Majors-Band 4 (TEM-4) and/or the TEM-8 when applicable. TEM-4 and TEM-8 are authoritative tests to judge the English proficiency of university undergraduate English majors in China (Chen, 2022). Participants who are able to pass these two tests are generally considered to have a relatively high proficiency in English. The low L2 proficiency group (Group 2) consisted of another 30 students (6 males; average age = 19.54 years; *SD* = 0.88 years), who had studied English for less than four semesters at the university according to a systematic curriculum. These participants had passed the College English Test Band 4 (CET-4), which is a large-scale test used to test the English proficiency of Chinese non-English majors (Wu et al., 2022), indicating that they were equipped with general knowledge of English, but less L2 experience and lower L2 proficiency than Group 1. Before the experiments, all participants were asked to fill out the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007), and their self-assessment scores were listed in Table 1. The differences between scores of high and low-proficiency bilinguals were significant ($ps < .0001$).

2.2. Design

The present study employed the picture-word interference paradigm, which is sensitive to the phonological relationship between the target picture and the distractor word (e.g., Levelt et al., 1991; Meyer & Schriefers, 1991; Starreveld, 2000). The picture-word interference paradigm is a widely used paradigm to investigate the process of speech production (e.g., Cai et al., 2020; Wong et al., 2012; Xin et al., 2020; Zhang & Yang, 2005). In this paradigm, participants are required to name the target picture while trying to ignore the distractor word, which is superimposed on the line drawing portraying concrete objects (Glaser & Düngelhoff, 1984) and shares certain properties with the target picture name. The target picture and the distractor may appear at pre-determined stimulus onset asynchronies (SOAs, the time duration between the distractor and the target) to reveal the time course of any potential effect. The studies of both Mandarin Chinese (e.g., Bi et al., 2009; Wang et al., 2021; Zhang & Yang, 2005; 2006; Zhao et al., 2012) and English (e.g., Damian & Martin, 1999; Jescheniak & Schriefers, 2001) manifested relatively stable phonological effects at positive SOAs (i.e., the target picture appears prior to the distractor word). The phonological forms of both the target picture and the distractor word will be activated as soon as they are retrieved, and the

Table 1. Self-assessment scores for the L2 English language skills from high and low proficiency bilinguals; the level was marked from 1 to 10, with 10 being the highest

	High proficiency	Low proficiency	*t* value	*p* value
Speaking	6.7 (1.43)	4.4 (1.69)	5.382	< .0001
Understanding spoken language	7.5 (1.17)	5.2 (1.49)	6.239	< .0001
Reading	7.6 (1.26)	6.1 (1.24)	4.321	< .0001

https://doi.org/10.1017/S1366728925000276 Published online by Cambridge University Press

phonological relatedness will facilitate the naming process (see Bürki, 2017 for a review). Thus, the current study chose three positive SOAs where phonological relatedness has been reported to facilitate picture naming (0 ms, 75 ms, and 150 ms, see also e.g., Wang et al., 2021; Zhang & Weekes, 2009) to investigate the primary phonological encoding units in L2 (i.e., English) spoken word production by Mandarin Chinese-English bilinguals with varied L2 proficiency.

Meanwhile, the degree of phonological relatedness between the target word and the distractor word was manipulated. There were four distractor types for each target, according to the extent of overlap in their phonological forms, that is, (1) syllabic overlap (S+), (2) two-segment overlap (P2+), (3) one-segment overlap (P1+), and (4) unrelated (U). The experimental design included two factors: Distractor Type (4 conditions: S+, P2+, P1+, U) and SOA (3 levels: 0 ms, 75 ms, 150 ms). There were 480 trials in total (40 pictures × 4 conditions × 3 SOAs), blocked by SOA. All trials were presented pseudo-randomly to make sure the same condition would not appear in two consecutive trials. The sequence of trials was counterbalanced across participants. There were self-paced rests between blocks. The materials and design were identical for the two groups.

2.3. Materials

Twenty-five target pictures were selected from CRL-IPNP (CRL International Picture Naming Project; Bates et al., 2000) and the standardized Snodgrass and Vanderwart picture databases (Snodgrass & Vanderwart, 1980) or drawn similarly. Target picture names were all monosyllabic. There were four distractor types for each target, according to the extent of overlap in their phonological forms: syllabic overlap (S+), two-segment overlap (P2+), one-segment overlap (P1+), unrelated (U). For instance, one target picture was a line drawing of a *nest*, and its distractor words were: *nest* (S+), *neck* (P2+), *nap* (P1+), and *salt* (U). Distractor words and target pictures were matched in terms of word frequency, $t = -.658$, $p = .512$, based on the log frequency in the SUBTLEX-UK database (Van Heuven et al., 2014), and visual complexity (number of letters), $t = -.473$, $p = .638$. Each pair of distractor and target pictures was semantically unrelated. They were also considered phonologically unrelated in their Chinese translations, except for one or two instances of onset or rhyme overlap between the target and one of the distractor conditions. Nevertheless, since the Chinese translations of English words are not a one-to-one correspondence, the rare instances of onset or rhyme overlap should not affect our results. Another 15 picture names were selected as fillers from the same database.

2.4. Procedure and analysis

Participants were seated in a comfortable chair in a quiet room facing a computer screen, approximately 60 cm away from the screen. Before starting the experiment, the participants filled out the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007) and signed an agreement to participate in the experiment voluntarily.

The experiment consisted of a familiarization, a practice session, and a formal experimental session. Participants were first presented with the line drawings on the screen with the target names underneath. After being familiarized with all the target pictures, they were asked to name the pictures in English without the names presented. Mistakes that occurred were reported to the participants and corrected by the experimenter.

The formal experiment started with a fixation cross "+" appearing in the middle of the screen for 300 ms. After the fixation cross disappeared, a blank screen appeared and lasted for 20 ms. Then, a target picture was presented with a distractor word superimposed at different SOAs. At last, the picture-word combination disappeared by the vocal trigger or after 2 s if the participants failed to name the targets. The whole experiment lasted about 25 minutes. The procedure was identical for the two groups. The whole procedure of the experiment is illustrated in Figure 2.

The experiment was conducted with PsychoPy2 Version 2021.2 (Peirce et al., 2019) with stimuli presented on a 15-inch computer screen 60 cm away from the participant. The reaction times (RTs, i.e., the naming latencies) were measured online by an HP laptop microphone. RTs were collected and manually checked using the program CheckVocal (Protopapas, 2007) based on the participants' vocal responses. R Version 3.1.0 (R Core Team, 2014) was used to analyze participants' picture naming RTs. The initial model was built employing the "lmer4" package (Bates et al., 2014) with two predictors: distractor type and SOA, the interaction between distractor type and SOA, and two random intercepts: participants and target pictures. The naming latencies showed a skewed distribution and were therefore log-transformed. The log-transformed naming latencies were submitted to the mixed-effects modeling in R as the dependent variable. The data analysis procedure was identical for both groups. There was a significant interaction between distractor type and SOA for both groups of participants ($ps < .001$). Therefore, the data were then divided into three subsets per SOA. Separate

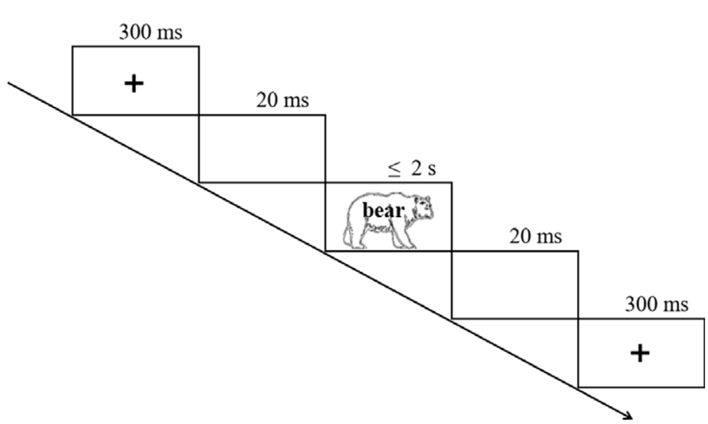

Figure 2. Procedure of the experiment.

https://doi.org/10.1017/S1366728925000276 Published online by Cambridge University Press

models were built with the distractor type and SOA levels as the fixed predictor and random intercepts for participants and target pictures.

3. Results

3.1. Group 1 – high L2 proficiency

3.1% of 9,000 data points, including incorrect naming and false voice triggering (2.46%) and outliers (i.e., data points that exceed a participant's mean RTs by 3 *SD*s, 0.64%), were excluded from further analysis. A total of 8,725 data points were submitted to R. The error rates were relatively low, and thus not included in further statistical analysis. Descriptive statistics are provided in Table 2.

For high proficiency bilinguals, at SOA = 0 ms, the model showed significant differences between the unrelated condition and other phonologically related conditions, suggesting that phonological relatedness facilitated picture naming. However, at SOA = 75 ms, the significant phonological facilitation effects were only obtained for the P2+ and S+ conditions, and at SOA = 150 ms, only for the S+ condition. See Table 3 for the results summary for high-proficiency bilinguals.

As shown in Figure 3, Tukey's multiple comparison tests showed that the differences between S+ and P2+ conditions reached significance over the range of SOA from 0 ms to 150 ms, βs < .0308, *p*s < .001, which revealed that facilitation increased as the amount of overlap increased. Moreover, for P2+ and P1+ conditions, the average time differences of 31 ms reached significance when SOA was 0 ms, $\beta = -.017$, $p < .001$, and when SOA was 75 ms with an average 20 ms difference, $\beta = -.012$, $p = .018$. However, the average 7 ms difference at SOA = 150 ms did not reach significance, $\beta = .003$, $p = .849$.

Table 2. Mean reaction times (RTs) in ms and standard deviation (SD) for high proficiency bilinguals

	Phonological relatedness							
	S+		P2+		P1+		U	
SOA (ms)	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
0	791	138	855	163	886	164	906	176
75	705	138	760	173	780	172	790	183
150	638	140	672	163	665	164	671	180

3.2. Group 2 – low L2 proficiency

6.04% of 9,000 data points were discarded (4.73% errors and 1.31% outliers). A total of 8,456 data points were submitted to R. The error rates were relatively low and thus were not included in further statistical analysis. Descriptive statistics are provided in Table 4 and detailed results are provided in Table 5.

For low proficiency bilinguals, at SOA = 0 ms, 75 ms, and 150 ms, the model showed significant differences between the unrelated condition and other phonologically related conditions, suggesting that phonological relatedness facilitated picture naming at all the predefined SOAs. See Table 5 for the summary of results for low-proficiency bilinguals.

As shown in Figure 4, Tukey's multiple comparison test was carried out and showed that the differences between the S+ and P2+ conditions reached significance at all SOA conditions, βs < .031, *p*s < .001. For P2+ and P1+ conditions, the 40 ms difference at SOA = 0 ms was significant, $\beta = -.021$, $p < .001$, and the 26 ms difference at SOA = 75 ms was also significant with $\beta = -.012$, $p = .023$. However, the effect at SOA = 150 ms was not significant, $\beta = -.002$, $p = .986$.

4. Discussion

Using the picture-word interference paradigm, the present study examined the primary phonological encoding units of L2 spoken word production in Mandarin Chinese-English bilinguals with high and low L2 proficiency. In both groups of participants, phonological facilitation effects were observed with segmental overlap (one or two segments) and syllabic overlap, suggesting Mandarin Chinese-English bilinguals employed segments as the primary

Table 3. Results for coefficient estimates, standard errors (SE), t values, and p values for the effect of distractor type in each SOA condition for high proficiency bilinguals

SOA (ms)	Distractor type	Coefficient estimate	SE	*t* value	*p* value
0	Intercept	2.95100	.008413	350.766	< .001
	S+	−.05712	.003481	−16.407	< .001
	P2+	−.02629	.003439	−7.644	< .001
	P1+	−.00908	.003469	−2.618	.00889
75	Intercept	2.88800	.010160	284.393	< .001
	S+	−.04710	0.003974	−11.851	< .001
	P2+	−.01635	.003989	−4.099	< .001
	P1+	−.00477	.003989	−1.196	.23
150	Intercept	2.81500	.011770	239.079	< .001
	S+	−0.01654	.004130	−4.003	< .001
	P2+	.00004	.004061	.011	.99
	P1+	−.00323	.004083	−.791	.43

https://doi.org/10.1017/S1366728925000276 Published online by Cambridge University Press

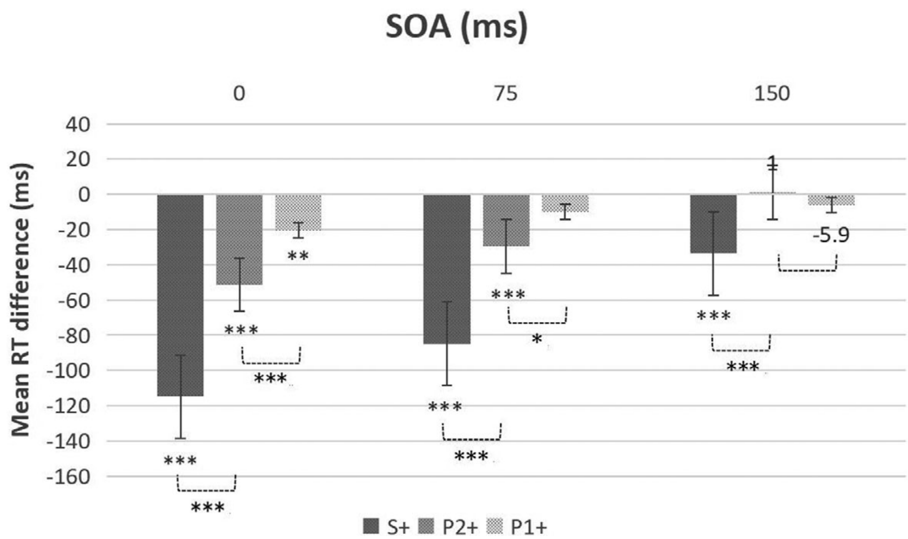

Figure 3. RT differences between the unrelated and phonologically related conditions for high proficiency bilinguals in Group 1. The dashed lines below the RT bars represent pairwise comparison results between adjacent levels in the chart (* $p < .05$; ** $p < .01$; *** $p < 0.001$).

Table 4. Mean reaction times (RTs) in ms and standard deviation (SD) for low proficiency bilinguals

	Phonological relatedness							
	S+		P2+		P1+		U	
SOA (ms)	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
0	865	177	942	198	982	209	997	206
75	776	171	863	222	889	224	909	230
150	715	177	778	218	778	218	800	232

phonological encoding units during spoken word production in their L2, resembling the units employed by native English speakers.

In both groups, overlap in the onset segment produced significant phonological facilitation effects, suggesting that Mandarin Chinese-English bilinguals use segments as the primary phonological encoding units in L2 spoken word production regardless of their L2 proficiency. The onset priming effect is consistent with the one reported by Schiller (2000) in an English monolingual picture naming task, which investigated the functional role of segments in English phonological encoding. As syllables were assumed to be the primary phonological encoding units in Mandarin Chinese (e.g., Cai et al., 2020; Chen et al., 2002; O'Seaghdha et al., 2010; Zhang & Yang, 2005), it seemed that Mandarin Chinese-English bilinguals employed language-specific units, that is, segments, to perform phonological encoding when producing their L2. This finding indicates that Mandarin Chinese-English bilinguals adopt an additional system for L2 phonological processing, supporting the accommodation hypothesis.

In addition to the onset priming effect observed with the masked priming paradigm (Verdonschot et al., 2013), the onset priming effect was reinforced with picture naming. Apart from the study

Table 5. Results for coefficient estimates, standard errors (SE), *t* values and *p* values for the effect of distractor type in each SOA condition for low proficiency bilinguals

SOA (ms)	Distractor type	Coefficient estimate	SE	*t* value	*p* value
0	Intercept	2.99400	.010450	286.523	< .001
	S+	−.06314	.003684	−17.141	< .001
	P2+	−.02854	.003654	−7.809	< .001
	P1+	−0.00769	.003685	−2.086	.0371
75	Intercept	2.94900	.012240	240.920	< .001
	S+	−.06799	.004242	−16.027	< .001
	P2+	−.02296	.004283	−5.361	< .001
	P1+	−.01082	.004290	−2.523	.0117
150	Intercept	2.89000	.014050	205.730	< .001
	S+	−.04353	.004589	−9.486	< .001
	P2+	−.01289	.004553	−2.832	.00466
	P1+	−.01135	.004563	−2.487	.01294

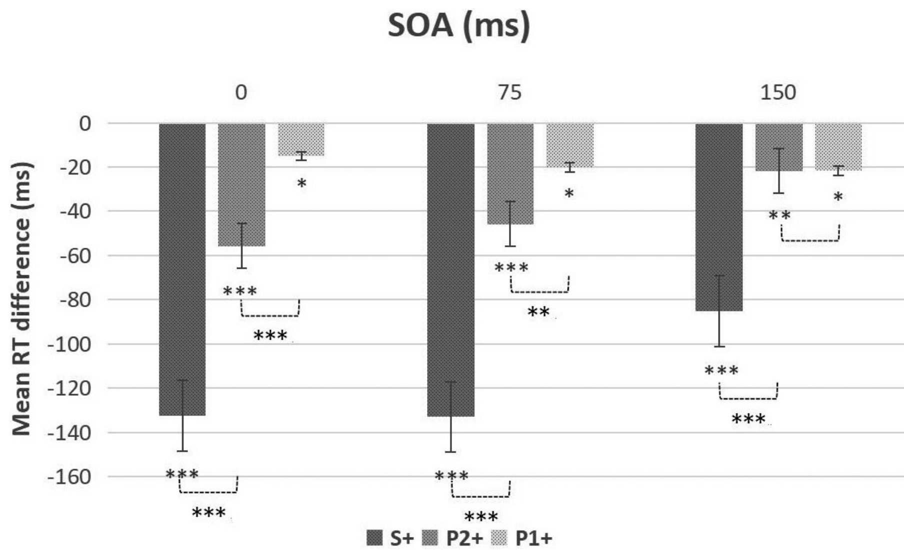

Figure 4. RT differences between the unrelated and phonologically related conditions for low proficiency bilinguals in Group 2. The dashed lines below the RT bars represent pairwise comparison results between adjacent levels in the chart (* $p < .05$; ** $p < .01$; *** $p < 0.001$).

with only highly proficient bilinguals (Verdonschot et al., 2013), our study further revealed that even when the participants' L2 proficiency was relatively low, segments were still employed as the primary phonological encoding units.

However, the finding of the low-proficiency group using segments as the primary phonological encoding units is inconsistent with that of Nakayama et al. (2016), where low-proficiency Japanese-English bilinguals showed CV priming but not segmental onset priming. One possible reason for the discrepancy is that most Mandarin Chinese speakers use Pinyin, an alphabetic transcription system to represent the sounds of the language, as the input method in typing, whereas Japanese speakers tend to use kana that usually represents a CV structure in typing. However, Japanese speakers may use "romaji," similar to Pinyin, when typing on a computer keyboard. The other possible reason is that the former study employed a reading-aloud task with prime words, but we used the picture naming task with visual distractors, which could contribute to the different results of Nakayama et al. (2016) and our study[1]. Further research is needed to examine these possibilities.

In addition, we observed increasing effects with more overlapping segments during L2 phonological encoding, which was consistent with the results in Dutch (Schiller, 1998) and English (Schiller, 1999, 2000) native speakers. Specifically, in both groups, we observed the time difference reached significance between the S + and P2+ conditions as well as the P2+ and P1+ conditions with varied SOAs (except for SOA = 150 ms) in both groups. The increasing effects of segmental overlap, with more overlapping segments producing larger facilitation effects in L2 production (see Figures 3 and 4), were consistent with the predictions that the overlapping segments increased the activation level of the target's phonemes and thus facilitated the syllabification at the phonological word (Levelt et al., 1999; Meyer & Schriefers, 1991; Wheeldon, 2003).

Although both groups of participants showed the segmental priming effect in phonological encoding, these two groups' performances were different in terms of distractor type and SOA. Specifically, the high-proficiency bilinguals seemed to have a naming advantage in L2 over low-proficiency bilinguals, based on a post-hoc *t*-test between the mean reaction times of the two groups in all the conditions ($t = -18.332$, $p < .0001$). One of the probable reasons for the naming speed difference could be lexical competition between L1 and L2, with L1 causing stronger interference in the low proficiency group (Colomé, 2001; Costa et al., 2000; Guo & Peng, 2006; Hoshino & Thierry, 2011; Macizo, 2016; Sullivan et al., 2018). However, it could also be that the ability of lexical access of the high proficiency group becomes better with increased L2 proficiency. Still, another possibility is that the prolonged naming could be caused by the delay at the L2 phonetic encoding stratum. Previous studies suggested that the disadvantages in the speed of speech production originated from the phonetic encoding level, which prolonged verbal action manner (e.g., Broos et al., 2018). Future research is needed to explore these different possibilities directly.

Additionally, there was a significant interaction between distractor type and SOA in both groups. Specifically, the priming effect was smaller at larger positive SOAs, and it was even absent in the P1 + condition at SOA = 75 ms, as well as the P1+ and P2+ conditions at SOA = 150 ms for high proficiency bilinguals. One possible reason is that the process of phonological encoding is (nearly) finished at these later points in time, especially for the high-proficiency group who tends to have faster word production. Specifically, based on the temporal signature of word production components proposed by Indefrey and Levelt (2004), lexical access starts within the time window of 250 ms after stimulus onset in spoken word encoding, followed by phonological encoding, which starts from phonological code retrieval at around 330 ms, online syllabification at around 455 ms, ending with phonetic encoding at approximately 600 ms. Crucially, the encoding takes about 25 ms per phonemic segment for native Dutch speakers (Van Turennout et al., 1997), while the speed may be slower for L2 learners in processing their weaker language (e.g., Dash & Kar, 2020; De Bot, 2004; Macizo, 2016). The mean number of phonemic segments of the target picture name was around four in our experiment. Thus, the segmental encoding cost would be around 100 ms for four phonemic segments, and the recognition of a distractor takes about

[1]We thank our reviewer for this suggestion.

https://doi.org/10.1017/S1366728925000276 Published online by Cambridge University Press

100 ms (e.g., Hauk et al., 2006). Therefore, the distractor might be presented too late to affect the production process. In other words, the segmental encoding process might be finished by high-proficiency bilinguals after the effective recognition of a distractor at SOAs of 75 ms and 150 ms in the P1+ and P2+ conditions. Comparatively, under the S+ and P2+ conditions, when SOA = 0 ms and SOA = 75 ms, the facilitation effects were obtained with enough processing time for both distractor word and target picture. Speakers benefit from the activated segments which primed the shared phonological codes and produced the segmental priming effect. Furthermore, the segmental priming at larger SOAs is more likely to be absent in the P1+ condition than the P2+ condition, compared to the robust priming in the syllabic overlap condition (i.e., lexical overlap) at all the specified SOAs, suggesting that the first segment is encoded first and then the second. Nevertheless, more fine-grained research is needed to make further conclusions.

One caveat of the current study is that all target words were monosyllabic. One consequence is that distractor words in the syllabic overlap condition are identical to the target words. This also explains why the syllabic priming effects are the most prominent across all the SOAs. It has been shown that in the form preparation paradigm, Mandarin Chinese-English bilinguals and Japanese-English bilinguals manifest only syllabic preparation effects but not phonemic effects in disyllabic word production (Li et al., 2020). Future cross-paradigm studies with polysyllabic words are necessary to further investigate the syllabic priming effects. Nevertheless, the finding of the syllabic priming does not compromise the findings of the segmental priming effects.

To interpret our results within the framework of the WEAVER++ model (Levelt et al., 1999; Roelofs & Meyer, 1998) and the schematic representation of the lexical system of bilinguals (Costa et al., 2006), we assume that in the process of L2 picture naming for Mandarin Chinese-English bilinguals, after the selection of lexical concepts, the activation spreads to corresponding lemma nodes in both L1 and L2 (see also Costa & Caramazza, 1999). Following lexical selection, the respective phonological forms are activated followed by the phonological encoding of the target words. Although our study did not directly investigate L1 activation in L2 production, our results are compatible with this account in terms of the suggested possibility of L1 interference causing lexical competition in L2 production. Nevertheless, in terms of the phonological encoding units in L2 production, we did not observe any apparent influence from L1.

Finally, the findings of the current study may have some pedagogical implications for L2 speech learning and segmental acquisition, as well as pronunciation instruction. Since this study demonstrated the significant role of segmental encoding in L2 production in Mandarin Chinese-English bilinguals regardless of their L2 proficiency, teachers should make students aware of the importance of segments. Studies examining the impact of segmental-based pronunciation instruction on intelligibility have demonstrated instructional gains (e.g. Saito, 2011; Saito & Lyster, 2012). Teachers may help students analyze their pronunciation features and help them identify and deal with features they find difficult to pronounce or discriminate (Wang, 2022).

In conclusion, we have investigated the primary phonological encoding units of L2 speech production for both high- and low-proficiency bilinguals. We found that Mandarin Chinese-English bilinguals, regardless of their L2 proficiency, employed segments as the primary phonological encoding units to process L2, demonstrating that they use the accommodation mechanism. In addition, we observed the decrease or even absence of facilitation with fewer overlapping segments at later SOAs. Our results shed light on the detailed underlying mechanism of L2 phonological encoding and may provide implications for L2 segmental acquisition and pronunciation instruction.

Data availability statement. The data that support the findings of this study are openly available in OSF at https://osf.io/tb7pq/?view_only=e5a69b04a8fa45fa8c9ff338aaf9d5e1.

Acknowledgements. We thank our participants for their participation.

Funding statement. This research was supported by a grant from the National Social Science Fund of China (Grant No. 24CYY098) awarded to M.W. N.O.S. is supported by grant no. 9380177 from CityUHK.

Competing interest. The authors declare none.

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. https://doi.org/10.1016/j.cognition.2006.02.002

Bates, E., **Federmeier, K.**, **Dan, H.**, **Iyer, G.**, & **Pechmann, T**. (2000). Introducing the CRL international picture-naming project (CRL-IPNP). *Center for Research in Language Newsletter*, **12**(1), 1–14.

Bates, D., **Maechler, M.**, **Bolker, B.**, & **Walker, S**. (2014). *lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7. This is a computer program.* R package. http://CRAN.R-project.org/package= lme4

Bi, Y., **Xu, Y.**, & **Caramazza, A**. (2009). Orthographic and phonological effects in the picture-word interference paradigm: Evidence from a logographic language. *Applied Psycholinguistics*, **30**(4), 637–658. https://doi.org/10.1017/S0142716409990051

Broos, W. P. J., **Duyck, W.**, & **Hartsuiker, R. J**. (2018). Are higher-level processes delayed in second language word production? Evidence from picture naming and phoneme monitoring. *Language, Cognition and Neuroscience*, **33**(10), 1219–1234. https://doi.org/10.1080/23273798.2018.1457168

Bürki, A. (2017). Differences in processing times for distractors and pictures modulate the influence of distractors in picture-word interference tasks. *Language, Cognition and Neuroscience*, **32**(6), 709–723. https://doi.org/10.1080/23273798.2016.1267783

Cai, X., **Yin, Y.**, & **Zhang, Q**. (2020). The roles of syllables and phonemes during phonological encoding in Chinese spoken word production: A topographic ERP study. *Neuropsychologia*, **140**, 107382. https://doi.org/10.1016/j.neuropsychologia.2020.107382

Calabria, M., **Grunden, N.**, **Iaia, F.**, & **García-Sánchez, C**. (2020). Interference and facilitation in phonological encoding: Two sides of the same coin? Evidence from bilingual aphasia. *Journal of Neurolinguistics*, **56**, 100935. https://doi.org/10.1016/j.jneuroling.2020.100935

Cao, F., **Tao, R.**, **Liu, L.**, **Perfetti, C. A.**, & **Booth, J. R**. (2013). High proficiency in a second language is characterized by greater involvement of the first language network: Evidence from Chinese learners of English. *Journal of Cognitive Neuroscience*, **25**(10), 1649–1663. https://doi.org/10.1162/jocn_a_00414

Chen, X. (2022). A study on content validity of reading comprehension in TEM-8 (2016–2021). *International Journal of New Developments in Education*, **4**(4), 1–6. https://doi.org/10.25236/IJNDE.2022.040401

Chen, J., **Chen, T.**, & **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. https://doi.org/10.1006/jmla.2001.2825

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

Chen, J. Y., **O'Séaghdha, P. G.**, & **Chen, T. M**. (2016). The primacy of abstract syllables in Chinese word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **42**(5), 825–836. https://doi.org/10.1037/a0039911

Cholin, J., **Schiller, N. O.**, & **Levelt, W. J. M**. (2004). The preparation of syllables in speech production. *Journal of Memory and Language*, **50**(1), 47–61. https://doi.org/10.1016/j.jml.2003.08.003

Costa, A. (2005). Lexical access in bilingual production. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 289–307). Oxford University Press.

Costa, A., **La Heij, W.**, & **Navarrete, E**. (2006). The dynamics of bilingual lexical access. *Bilingualism: Language and Cognition*, **9**(2), 137–151. https://doi.org/10.1017/S1366728906002495

Costa, A., **Colomé, À.**, & **Caramazza, A**. (2000). Lexical access in speech production: The bilingual case. *Psicológica*, **21**(2), 403–437.

Costa, A., **Colomé, À.**, **Gómez, O.**, & **Sebastián-Gallés, N**. (2003). Another look at cross language competition in bilingual speech production: Lexical and phonological factors. *Bilingualism: Language and Cognition*, **6**(3), 167–179. https://doi.org/10.1017/S1366728903001111

Costa, A., & **Caramazza, A**. (1999). Is lexical selection in bilingual speech production language- specific? Further evidence from Spanish-English and English-Spanish bilinguals. *Bilingualism: Language and Cognition*, **2**(3), 231–244. https://doi.org/10.1017/S1366728999000334

Costa, A., **Miozzo, M.**, & **Caramazza, A**. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language*, **41**(3), 365–397. https://doi.org/10.1006/jmla.1999.2651

Costa, A., & **Santesteban, M**. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, **50**(4), 491–511. https://doi.org/10.1016/j.jml.2004.02.002

Colomé, À. (2001). Lexical activation in bilinguals' speech production: Language-specific or language-independent? *Journal of Memory and Language*, **45**(4), 721–736. https://doi.org/10.1006/jmla.2001.2793

Dash, T., & **Kar, B. R**. (2020). Behavioural and ERP correlates of bilingual language control and general-purpose inhibitory control predicted by L1 and L2 proficiency. *Journal of Neurolinguistics*, **56**, 100914. https://doi.org/10.1016/j.jneuroling.2020.100914

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, **93**(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

De Bot, K. (1992). A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics*, **13**(1), 1–24. https://doi.org/10.1093/applin/13.1.1

De Bot, K. (2004). The multilingual lexicon: Modelling selection and control. *International Journal of Multilingualism*, **1**(1), 17–32. https://doi.org/10.1080/14790710408668176

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. https://doi.org/10.1016/S0749-596X(03)00008-1

Damian, M. F., & **Dumay, N**. (2007). Time pressure and phonological advance planning in spoken production. *Journal of Memory and Language*, **57**(2), 195–209. https://doi.org/10.1016/j.jml.2006.11.001

Damian, M. F., & **Dumay, N**. (2009). Exploring phonological encoding through repeated segments. *Language and Cognitive Processes*, **24**(5), 685–712. https://doi.org/10.1080/01690960802351260

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. https://doi.org/10.1037/0278-7393.25.2.345

Forster, K. I., & **Davis, C**. (1991). The density constraint on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language*, **30**(1), 1–25. https://doi.org/10.1016/0749-596X(91)90008-8

Guo, T., & **Peng, D**. (2006). Event-related potential evidence for parallel activation of two languages in bilingual speech production. *Cognitive Neuroscience and Neuropsychology*, **17**(17), 1757–1760. https://doi.org/10.1097/01.wnr.0000246327.89308.a5

Glaser, W. R., & **Düngelhoff, F. J**. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, **10**(5), 640–654. https://doi.org/10.1037/0096-1523.10.5.640

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, **1**(2), 67–81. https://doi.org/10.1017/S1366728998000133

Hauk, O., **Davis, M. H.**, **Ford, M.**, **Pulvermüller, F.**, & **Marslen-Wilson, W. D**. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *NeuroImage*, **30**(4), 1383–1400. https://doi.org/10.1016/j.neuroimage.2005.11.048

Hoshino, N., & **Thierry, G**. (2011). Language selection in bilingual word production: Electrophysiological evidence for cross-language competition. *Brain Research*, **1371**, 100–109. https://doi.org/10.1016/j.brainres.2010.11.053

Indefrey, P., & **Levelt, W. J. M**. (2004). The spatial and temporal signatures of word production components. *Cognition*, **92**(1), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

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. https://doi.org/10.1080/23273798.2014.892144

Jescheniak, J. D., & **Schriefers, H**. (2001). Priming effects from phonologically related distractors in picture-word interference. *The Quarterly Journal of Experimental Psychology*, **54**(2), 371–382. https://doi.org/10.1080/713755981

Jiao, L., **Grundy, J. G.**, **Liu, C.**, & **Chen, B**. (2020). Language context modulates executive control in bilinguals: Evidence from language production. *Neuropsychologia*, **142**, 107441. https://doi.org/10.1016/j.neuropsychologia.2020.107441

Levelt, W. J. M., **Roelofs, A.**, & **Meyer, A. S**. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, **22**(1), 1–38. https://doi.org/10.1017/S0140525X99001776

Levelt, W. J. M., **Schriefers, H.**, **Vorberg, D.**, **Meyer, A. S.**, **Pechmann, T.**, & **Havinga, J**. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, **98**(1), 122–142. https://doi.org/10.1037/0033-295X.98.1.122

Li, C., **Kronrod, Y.**, & **Wang, M**. (2020). The influence of the native language on phonological preparation in spoken word production in a second language. *Linguistic Approaches to Bilingualism*, **10**(1), 109–151. https://doi.org/10.1075/lab.16027.li

Li, C. & **Wang, M**. (2017). The influence of orthographic experience on the development of phonological preparation in spoken word production. *Memory & Cognition*, **45**, 956–973. https://doi.org/10.3758/s13421-017-0712-5

Li, C., **Wang, M.**, & **Davis, J. A**. (2017). The phonological preparation unit in spoken word production in a second language. *Bilingualism: Language and Cognition*, **20**(2), 351–366. https://doi.org/10.1017/S1366728915000711

Li, C., **Wang, M.**, & **Idsardi, W**. (2015). The effect of orthographic form-cuing on the phonological preparation unit in spoken word production. *Memory & Cognition*, **43**, 563–578. https://doi.org/10.3758/s13421-014-0484-0

Liu, X., **Hu, L.**, **Qu, J.**, **Zhang, S.**, **Su, X.**, **Li, A.**, & **Mei, L**. (2023). Neural similarities and differences between native and second languages in the bilateral fusiform cortex in Chinese-English bilinguals. *Neuropsychologia*, **179**, 108464. https://doi.org/10.1016/j.neuropsychologia.2022.108464

Macizo, P. (2016). Phonological coactivation in the bilinguals' two languages: Evidence from the color naming task. *Bilingualism: Language and Cognition*, **19**(2), 361–375. https://doi.org/10.1017/S136672891500005X

Malouf, T., & **Kinoshita, S**. (2007). Masked onset priming effect for high-frequency words: Further support for the speech-planning account. *Quarterly Journal of Experimental Psychology*, **60**(8), 1155–1167. https://doi.org/10.1080/17470210600964035

Marian, V., **Blumenfeld, H. K.**, & **Kaushanskaya, M**. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, **50**(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

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. https://doi.org/10.1016/0749-596X(91)90011-8

Meyer, A. S., & **Schriefers, H**. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **17**(6), 1146–1160. https://doi.org/10.1037/0278-7393.17.6.1146

Nakayama, M., **Kinoshita, S.**, & **Verdonschot, R. G**. (2016). The emergence of a phoneme-sized unit in L2 speech production: Evidence from Japanese-English bilinguals. *Frontiers in Psychology*, **7**, 175. https://doi.org/10.3389/fpsyg.2016.00175

Nakayama, M., **Verdonschot, R. G.**, **Sears, C. R.**, & **Lupker, S. J**. (2014). The masked cognate translation priming effect for different-script bilinguals is modulated by the phonological similarity of cognate words: Further support for the phonological account. *Journal of Cognitive Psychology*, **26**(7), 714–724. https://doi.org/10.1080/20445911.2014.953167

O'Seaghdha, P. G., **Chen, J.** & **Chen, T**. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition*, **115**(2), 282–302. https://doi.org/10.1016/j.cognition.2010.01.001

Peirce, J., **Gray, J. R.**, **Simpson, S.**, **MacAskill, M.**, **Höchenberger, R.**, **Sogo, H.**, **Kastman, E.**, & **Lindeløv, J. K**. (2019). Psychopy2: Experiments in behavior made easy. *Behavior Research Methods*, **51**(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Poulisse, N. & **Bongaerts, T**. (1994). First language use in second language production. *Applied Linguistics*, **15**(1), 36–57. https://doi.org/10.1093/applin/15.1.36

Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, **39**(4), 859–862. https://doi.org/10.3758/BF03192979

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), 14265–14270. https://doi.org/10.1073/pnas.1200632109

R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria. http://www. R-project. org

Roelofs, A. (1999). Phonological segments and features as planning units in speech production. *Language and Cognitive Processes*, **14**(2), 173–200. https://doi.org/10.1080/016909699386338

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. https://doi.org/10.1111/jpr.12050

Roelofs, A., & **Meyer, A. S**. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24**(4), 922–939. https://doi.org/10.1037/0278-7393.24.4.922

Sadat, J., **Martin, C. D.**, **Costa, A.**, & **Alario, F. X**. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology*, **68**(1), 33–58. https://doi.org/10.1016/j.cogpsych.2013.10.001

Saito, K. (2011). Examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: An instructed SLA approach to L2 phonology. *Language Awareness*, **20**(1), 45–59. https://doi.org/10.1080/09658416.2010.540326

Saito, K., & **Lyster, R**. (2012). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /J/ by Japanese learners of English. *Language Learning*, **62**(2), 595–633. https://doi.org/10.1111/j.1467-9922.2011.00639.x

Schiller, N. O. (1998). The effect of visually masked syllable primes on the naming latencies of words and pictures. *Journal of Memory and Language*, **39**(3) 484–507. https://doi.org/10.1006/jmla.1998.2577

Schiller, N. O. (1999). Masked syllable priming of English nouns. *Brain and Language*, **68**(1–2), 300–305. https://doi.org/10.1006/brln.1999.2109

Schiller, N. O. (2000). Single word production in English: The role of subsyllabic units during phonological encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **26**(2), 512–528. https://doi.org/10.1037/0278-7393.26.2.512

Schiller, N. O. (2006). Lexical stress encoding in single word production estimated by event-related brain potentials. *Brain Research*, **1112**(1), 201–212. https://doi.org/10.1016/j.brainres.2006.07.027

Schwartz, M. F. (2014). Theoretical analysis of word production deficits in adult aphasia. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **369**(1634), 20120390. https://doi.org/10.1098/rstb.2012.0390

Snodgrass, J. G., & **Vanderwart, M**. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, **6**(2), 174–215. https://doi.org/10.1037/0278-7393.6.2.174

Spalek, K., **Hoshino, N.**, **Wu, Y. J.**, **Damian, M.**, & **Thierry, G**. (2014). Speaking two languages at once: Unconscious native word form access in second language production. *Cognition*, **133**(1), 226–231. https://doi.org/10.1016/j.cognition.2014.06.016

Starreveld, P. A. (2000). On the interpretation of auditory context effects in word production. *Journal of Memory and Language*, **42**(4), 497–525. https://doi.org/10.1006/jmla.1999.2693

Sullivan, M. D., **Poarch, G. J.**, & **Bialystok, E**. (2018). Why is lexical retrieval slower for bilinguals? Evidence from picture naming. *Bilingualism: Language and Cognition*, **21**(3), 479–488. https://doi.org/10.1017/S1366728917000694

Thierry, G., & **Wu, Y**. (2004). Electrophysiological evidence for language interference in late bilinguals. *Cognitive Neuroscience and Neuropsychology*, **15**(10), 1555–1558. https://doi.org/10.1097/01.wnr.0000134214.57469.c2

Timmer, K., & **Chen, Y**. (2017). Dutch-Cantonese bilinguals show segmental processing during Sinitic language production. *Frontiers in Psychology*, **8**, 1133. https://doi.org/10.3389/fpsyg.2017.01133

Van Heuven, W. J. B., **Mandera, P.**, **Keuleers, E.**, & **Brysbaert, M**. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, **67**, 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Van Turennout, M., **Hagoort, P.**, & **Brown, C. M**. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **23**(4), 787–806. https://doi.org/10.1037/0278-7393.23.4.787

Verdonschot, R. G., **Nakayama, M.**, **Zhang, Q.**, **Tamaoka, K.**, & **Schiller, N. O**. (2013). The proximate phonological unit of Chinese-English bilinguals: Proficiency matters. *PLoS ONE*, **8**(4), e61454. https://doi.org/10.1371/journal.pone.0061454

Wang, X. (2022). Segmental versus suprasegmental: Which one is more important to teach? *RELC Journal*, **53**(1), 194–202. https://doi.org/10.1177/0033688220925926

Wang, J., **Wong, A. W. K.**, & **Chen, H**. (2021). Second language experience influences salience of phonological units in spoken word production in the first language. *International Journal of Bilingualism*, **25**(6), 1–15. https://doi.org/10.1177/13670069211031001

Wheeldon, L. R. (2003). Inhibitory form priming of spoken word production. *Language and Cognitive Processes*, **18**(1), 81–109. https://doi.org/10.1080/\01690960143000470

Wheeldon, L. R., & **Levelt, W. J. M**. (1995). Monitoring the time-course of phonological encoding. *Journal of Memory and Language*, **34**(3), 311–334. https://doi.org/10.1006/jmla.1995.1014

Wong, A. W. K., **Huang, J.**, & **Chen, H. C**. (2012). Phonological units in spoken word production: Insights from Cantonese. *PLoS ONE*, **7**(11), e48776. https://doi.org/10.1371/journal.pone.0048776

Wu, F., **Chen, X.**, & **Zheng, H**. (2022). CET-4 listening test effect on listening learning based on machine learning. *Wireless Communications and Mobile Computing*, **2022**(1), 5604032. https://doi.org/10.1155/2022/5604032

Xin, X., **Lan, T.**, & **Zhang, Q**. (2020). Assimilation mechanisms of phonological encoding in second language spoken production for English-Chinese bilinguals. *Acta Psychologica Sinica*, **52**(12), 1377–1392.

Xu, G., **Lin, J.**, & **Dong, Y**. (2021). Cross-script phonological activation in Chinese-English bilinguals: The effect of SOA from masked priming. *Canadian Journal of Experimental Psychology*, **75**(4), 374–386. https://doi.org/10.1037/cep0000262

You, W., **Zhang, Q.**, & **Verdonschot, R. G**. (2012). Masked syllable priming effects in word and picture naming in Chinese. *PLoS ONE*, **7**, e46595. https://doi.org/10.1371/journal.pone.0046595

Zhang, Q., & **Damian, M. F**. (2009). The time course of segment and tone encoding in Chinese spoken production: An event-related potential study. *Neuroscience*, **163**, 252–265. https://doi.org/10.1016/j.neuroscience.2009.06.015

Zhang, Q., & **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., **Qian, Z.**, & **Zhu, X**. (2021). The multiple phonological activation in Chinese spoken word production: An ERP study in a word translation task. *Acta Psychologica Sinica*, **53**(1), 1–14.

Zhang, Q., & **Weekes, B. S**. (2009). Orthographic facilitation effects on spoken word production: Evidence from Chinese. *Language and Cognitive Processes*, **24**(7–8), 10821096. https://doi.org/10.1080/01690960802042133

Zhang, Q., & **Yang, Y**. (2005). The phonological planning unit in Chinese monosyllabic word production. *Psychological Science*, **28**(2), 374–378.

Zhang, Q., & **Yang, Y**. (2006). The interaction of lexical selection and phonological encoding in Chinese word production. *Acta Psychologica Sinica*, **38**(4), 480–488.

Zhang, Q., **Zhu, X.**, & **Damian, M. F**. (2018). Phonological activation of category coordinates in spoken word production: Evidence for cascaded processing in English but not in Mandarin. *Applied Psycholinguistics*, **39**, 835–860. https://doi.org/10.1017/S0142716418000024

Zhao, H., **La Heij, W.**, & **Schiller, N. O**. (2012). Orthographic and phonological facilitation in speech production: New evidence from picture naming in Chinese. *Acta Psychologica*, **139**(2), 272–280. https://doi.org/10.1016/j.actpsy.2011.12.001