

Time course of syllabic and sub-syllabic processing in Mandarin word production: Evidence from the picture-word interference paradigm

Jie Wang¹ · Andus Wing-Kuen Wong² · Hsuan-Chih Chen¹

Published online: 5 June 2017
© Psychonomic Society, Inc. 2017

Abstract The time course of phonological encoding in Mandarin monosyllabic word production was investigated by using the picture-word interference paradigm. Participants were asked to name pictures in Mandarin while visual distractor words were presented before, at, or after picture onset (i.e., stimulus-onset asynchrony/SOA = -100, 0, or +100 ms, respectively). Compared with the unrelated control, the distractors sharing atonal syllables with the picture names significantly facilitated the naming responses at -100- and 0-ms SOAs. In addition, the facilitation effect of sharing word-initial segments only appeared at 0-ms SOA, and null effects were found for sharing word-final segments. These results indicate that both syllables and subsyllabic units play important roles in Mandarin spoken word production and more critically that syllabic processing precedes subsyllabic processing. The current results lend strong support to the proximate units principle (O'Seaghdha, Chen, & Chen, 2010), which holds that the phonological structure of spoken word production is language-specific and that atonal syllables are the proximate phonological units in Mandarin Chinese. On the other hand, the significance of word-initial segments over word-final segments suggests that serial processing of segmental information seems to be universal across Germanic languages and Chinese, which remains to be verified in future studies.

Electronic supplementary material The online version of this article (doi:10.3758/s13423-017-1325-5) contains supplementary material, which is available to authorized users.

✉ Hsuan-Chih Chen
ikechen@cuhk.edu.hk

¹ Department of Psychology, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, S.A.R., China

² Nam Shan Psychology Laboratory, Department of Applied Social Sciences, City University of Hong Kong, Hong Kong, S.A.R., China

Keywords Phonological encoding · Mandarin spoken word production · Picture-word interference

In spoken word production, the phonological form of the target word needs to be constructed before the preparation of articulatory gestures, and this process is called phonological encoding (Caramazza, 1997; Dell, 1986, 1988; Levelt, 1989, 1999). It has been heatedly discussed whether the planning units at this stage are language-specific or universal. Previous studies in alphabetic languages, especially in Germanic languages, have shown that phonemic segments are effective planning units in spoken word production (Dutch: Meyer, 1991; Meyer & Schriefers, 1991; Roelofs, 2006; Schiller, 2008; English: Damian & Bowers, 2003; Damian & Dumay, 2009; French: Alario, Perre, Castel, & Ziegler, 2007; Russian: Timmer, Ganushchak, Mitlina, & Schiller, 2014; Spanish: Santiago, 2000; Persian: Timmer, Vahid-Gharavi, & Schiller, 2012). Participants' naming performance was significantly improved if the onset segment of the target word was primed or repeated in certain ways. In contrast, researchers failed to see the onset effect in Chinese or Japanese spoken word production but consistently found the effects of atonal syllables (Chinese: Chen & Chen, 2013; Chen, Chen, & Dell, 2002; Chen, O'Seaghdha, & Chen, 2016; O'Seaghdha et al., 2010; You, Zhang, & Verdonschot, 2012; Zhang, 2008; Zhang & Yang, 2005; see Li, Wang, & Idsardi, 2015; Verdonschot, Nakayama, Zhang, Tamaoka, & Schiller, 2013, for exceptions) and moras (Japanese: Kureta, Fushimi, & Tatsumi, 2006; Verdonschot et al., 2011) respectively.

To account for these various findings, a proximate units principle (O'Seaghdha et al., 2010) was proposed that in spoken word production the phonological units directly connected to morpheme nodes (or lexeme nodes) are language-specific (e.g., segments and metrical frames in English, moras

and tonal frames in Japanese, atonal syllables and tonal frames in Chinese). This principle has been incorporated into one influential model of speech production (i.e., WEAVER++: Levelt, Roelofs, & Meyer, 1999; Roelofs, 2015), as shown in Figure 1a and b. In the form network of English spoken word production (Figure 1a), morpheme nodes are directly connected to segments and metrical frames, and segments are associated with the metrical frame serially from the onset to its end (i.e., syllabification), followed by the activation of syllable motor programs. In the Chinese version of WEAVER++ (Roelofs, 2015; Figure 1b), atonal syllables are associated to the tonal frame serially as the first selectable units below the word level. Without the need of syllabification, segments within the atonal syllable are subsequently specified in parallel, but not in serial, resulting in the null onset effect.

Nevertheless, one may argue that the syllabification process also occurs in Chinese spoken word production but

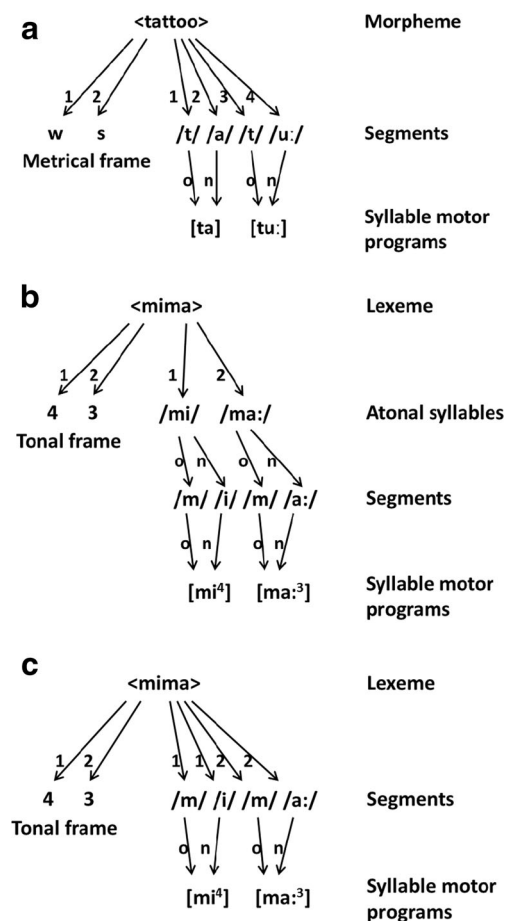


Fig. 1 Word form representation of the English word *tattoo* (a) and the Chinese word *mima* (“密码”; b and c) in spoken word production. (a) and (b) Adapted from the WEAVER++ model (Roelofs, 2015); (c) an alternative proposal with segments as the proximate phonological units in Chinese. In the Figure, o = onset; n = nucleus; w = weak; s = strong. The numbers 4 and 3 in the tonal frame denote the type of tones. The numbers attached to the arrows indicate the order of encoding.

proceeds more efficiently due to the simpler syllable structure in Chinese. Unlike Germanic languages, consonant clusters are generally lacking in spoken Chinese, and no resyllabification is needed due to relatively clear syllable boundaries (each syllable corresponds to a logogram in the Chinese orthography). Consequently, segments may be syllabified relatively more easily and rapidly in Chinese, resulting in a too subtle onset effect to be observed. When more segments are manipulated, behavioral differences become relatively larger and observable (see Meyer, 1991, for evidence in Dutch). Consistent with this alternative account, researchers have found subsyllabic effects in Chinese word production when more than one segment was manipulated (Verdonschot, Lai, Chen, Tamaoka, & Schiller, 2015; Wong & Chen, 2008, 2009, 2015; Wong, Huang, & Chen, 2012). For example, using a masked priming paradigm, Verdonschot et al. (2015) found that native Mandarin speakers’ naming responses were facilitated when prime words and target words shared the same onset and vowel (e.g., 啤 /pi²/ primes 贫 /pin²/, 班 /ban¹/ primes 八 /ba¹/). Therefore, the assumption of segments as primary phonological units in many models of language production (Caramazza, 1997; Dell, 1986, 1988; Levelt, 1989, 1999) may apply to Chinese word production as well. Figure 1c shows this alternative proposal for the phonological encoding process in Chinese spoken word production, which assumes that segments are retrieved as the proximate phonological units and are associated with the tonal frame in a highly efficient manner (i.e., rapidly in serial, or even in parallel within a syllable).

Thus, it is still debatable whether atonal syllables or segments are the proximate phonological units in Chinese spoken word production. One way to distinguish the two accounts mentioned above is to look at the time course of syllabic and subsyllabic processing. The model of Roelofs (2015; see also O’Seaghdha et al., 2010) assumes that atonal syllables are the proximate units and that syllable selection precedes segmental specification. Contrarily, the alternative proposal assumes that segments are the proximate units and are integrated into a syllable-size output. However, little evidence has been reported explicitly comparing the time course of syllabic and subsyllabic processing in Chinese word production. In a picture-word interference (PWI) paradigm, Wong and Chen (2008) asked native speakers of Cantonese (i.e., a Chinese dialect) to name aloud pictures while ignoring visual word distractors that were presented before, at, or after picture onset (i.e., SOA = -200, -100, 0, or +100 ms, respectively) for 200 ms. When the picture name and the distractor shared the tonal syllable, the atonal syllable, or the rhyme + tone, participants’ responses were facilitated as compared with unrelated distractors, and no reliable interactions between phonological variables and SOA were found. Note, however, that different sets of distractors were used in the related and unrelated conditions of Wong and Chen (2008), which could possibly

confound their results. Furthermore, their participants were proficient Cantonese-English bilinguals in Hong Kong. Their relatively high English proficiency might influence phonological encoding in Cantonese and moderate the results (Verdonschot et al., 2013).

The current study aims to use an improved PWI design to investigate the time course of syllabic and subsyllabic processing in Chinese spoken word production. Pictures with monosyllabic Mandarin names were used as targets, whereas visual distractors were presented at three SOA durations (i.e., -100, 0, and +100 ms) for a constant duration (i.e., 200 ms) and replaced by a cross sign during the remainder of the trial, following Wong and Chen (2008). The choice for the SOAs also was based on the estimation of Indefrey and Levelt (2004; see also Indefrey, 2011) that the phonological encoding process lasts approximately 180 ms (with an average naming latency of 600 ms). SOAs ranging from -100 to +100 ms should be broad and sensitive enough to cover the major processes involved in phonological encoding. Importantly, distractors related to atonal syllables or subsyllabic units of the targets were recombined with the targets to generate their corresponding unrelated controls so that the same distractors were used in the related and unrelated conditions (Meyer & Schriefers, 1991; Wong, Wang, Ng, & Chen, 2016). If atonal syllables are the proximate phonological units for Chinese, it is expected that the effect of syllabic relatedness would occur earlier than that of subsyllabic relatedness.

Another issue that the current study aims to explore is whether or not segmental information is processed in parallel in Chinese spoken word production. While Roelofs (2015) assumes parallel selection of segments within an atonal syllable, it is not implied by the proximate units principle itself (O'Seaghdha et al., 2010). Also in the alternative proposal, Chinese speakers are highly efficient in the syllabification process, so either serial or parallel selection of segments could hide the onset effect in Chinese. To distinguish these two manners of segmental processing (i.e., parallel vs. serial), two types of subsyllabic relatedness were included in the current study: distractors consisting of word-initial or word-final segments of the target (e.g., target 刀 /dao1/ and its distractors 达 /da2/ and 奥 /ao4/). Segments not contained by the target were excluded in its corresponding distractors so as to maximize the possibility of observing subsyllabic effects. If segments are selected in parallel, similar effects should be obtained for these two types of subsyllabic relatedness, with no difference in time course. Otherwise a larger and/or earlier effect of word-initial relatedness should be observed as compared with that of word-final relatedness.

Method

Participants

Thirty-four students (8 males) from the Chinese University of Hong Kong, with an average age of 22 years ($SD = 3.6$ years), participated in the current experiment with monetary rewards (50 HKD per participant). They are native Mandarin speakers from Mainland China and had stayed in Hong Kong for 1.2 years on average ($SD = 1.1$ years). Their self-reported English proficiency was 4.9 scores ($SD = 0.6$) on a 7-point scale with 7 denoting highest proficiency. Some participants also speak other Chinese dialects, such as Cantonese. All participants were neurologically healthy and had normal or corrected-to-normal vision. Informed consent was obtained from each participant at the beginning of the experiment. Data of two female participants were discarded due to excessive naming latencies (exceeding 2.5 SD of group mean).

Stimuli and apparatus

Twenty-four white-on-black line drawings of common objects were used as targets, each with a monosyllabic Mandarin name (e.g., 刀 /dao1/, meaning “knife”). Three sets of single Chinese characters were selected as distractors to pair with the target pictures in following ways (see [Supplementary Material](#)): (a) syllable-related, the character sharing the whole atonal syllable with the picture name but with a different tone (e.g., 导 /dao3/, “guide”, for 刀 /dao1/); (b) body-related (i.e., word-initial related), the character consisting of the initial segments of the picture name but with a different tone (e.g., 达 /da2/, “reach”, for 刀 /dao1/); and (c) rhyme-related (i.e., word-final related), the character consisting of the rhyme of the picture name but with a different tone (e.g., 奥 /ao4/, “abstruse”, for 刀 /dao1/). They were closely matched in visual character frequency and number of strokes ($F_s < 0.1$), based on the corpus developed by Da (2004). Each set of characters were also recombined with the pictures to produce their unrelated control conditions. The number of shared segments between each target-distractor pair was on average 3.25 ($SD = 0.44$), 2.00 ($SD = 0$), 2.25 ($SD = 0.44$) for the syllable-related, body-related, rhyme-related conditions respectively, and 0.67 ($SD = 0.56$), 0.13 ($SD = 0.34$), 0.46 ($SD = 0.51$) for their corresponding control conditions. Other than phonological overlap, the target-distractor pairs were not related in written form or semantic meaning in any obvious way. Characters were presented in simplified Chinese.

Presented with E-Prime 2.0 software, the pictures expanded 6 cm × 6 cm (approximately 5° × 5° in visual angle) on the computer screen while the characters 1.2 cm × 1.2 cm (approximately 1° × 1°). The screen refresh rate was 60 Hz. The onset of naming response was detected by a serial response-box with a microphone.

Design and procedure

The experiment adopted a three-factor within-participants design (target-distractor relatedness: phonologically related vs. unrelated; type of relatedness: syllable, body, rhyme; SOA: -100, 0, +100 ms). All pictures appeared once in each of the 18 conditions ($2 \times 3 \times 3$), resulting in 432 trials in total. Conditions were intermixed in 18 blocks so that in each block different levels of each factor were equiprobable and all pictures appeared once. The order of blocks and trials were randomized.

In each experimental trial, a white fixation was first presented at the center of the black screen for 500 ms, followed by a blank of 500 ms. The distractor (at -100-ms SOA) or the mask “x” (at 0- or +100-ms SOA) was then presented at the center for 100 ms before the picture onset. The picture stayed on the screen for 2,000 ms or until a naming response was detected by the voicekey. At each SOA duration, the distractor stayed for 200 ms and was replaced by the mask. Participants were required to ignore the distractor and name the picture aloud as accurately and quickly as possible. The trial ended with a blank of 1,500 to 2,500 ms. Short breaks were allowed between blocks. Before the experimental trials, participants were asked to learn the names of the 24 pictures and perform 24 practice trials upon successful learning. Naming errors were corrected at this stage. The whole procedure lasted for approximately 45 minutes.

Results

Table 1 shows the mean naming latencies and error rates in each condition. Trials with incorrect or no naming responses (1.1%), voicekey mistriggers (4.6%), or extreme latencies (exceeding 2.5 SD of individual mean, 2.5%) were excluded from naming latency analyses. Given that the error rates were very low in each condition (all approximately 1%),

subsequent analyses focused on the naming latencies, which were inverse transformed ($-1,000/\text{RT}$) and submitted to linear mixed-effect modeling (LMEM; Baayen, Davidson, & Bates, 2008; Bates, Maechler, Bolker, Walker, 2015) implemented in R Version 3.3.1 (R development core team, 2016). *lmerTest* package (Kuznetsova, Brockhoff, Christensen, 2016) was used to calculate p values with Satterthwaite approximation. Details of the LMEM analyses including the formulas for models can be found in the [Supplementary Material](#).

SOA, type of relatedness, target-distractor relatedness, block, and all possible interactions were fixed effects, whereas participants and items were random effects. By-participant and by-item random intercepts, and by-participant random slopes for SOA, type of relatedness, target-distractor relatedness and block were included. This model (m1) was compared with a more specific model (m2) that excluded the interactions involving block through a likelihood ratio test. The result showed that these two models produced similar fit for the naming latencies ($\chi^2_{(17)} = 14.51, p = 0.631$), indicating that the other fixed effects did not vary with fatigue or item repetition. We further removed the three-way interaction of SOA, type of relatedness and target-distractor relatedness from m2 (m3), and another model comparison (m2 vs. m3) showed that the three-way interaction was significant ($\chi^2_{(4)} = 13.08, p = 0.011$).

The effect of target-distractor relatedness was further analyzed separately for each condition. Table 2 shows the regression coefficients (b), standard errors (SE), p values and Cohen's d (Judd, Westfall, & Kenny, 2017) of the factor target-distractor relatedness. Syllable-relatedness significantly facilitated naming responses at SOAs of -100 and 0 ms. Body-relatedness showed a significant facilitation at 0-ms SOA, but null effects were found at other SOAs. Rhyme-relatedness showed null effects at all SOAs.

To compare the body effect with the rhyme effect, additional analyses were conducted on a smaller dataset excluding the syllable conditions. Model comparisons demonstrated that the

Table 1 Mean naming latencies and error rates (standard errors in the parentheses) in each condition

SOA	Type of relatedness	Naming latency (ms)		Error rate (%)	
		Related	Unrelated	Related	Unrelated
-100 ms	Syllable	597 (11)	618 (12)	1.4 (0.4)	1.3 (0.5)
	Body	613 (12)	613 (12)	1.5 (0.4)	1.0 (0.3)
	Rhyme	608 (11)	604 (11)	1.4 (0.5)	0.6 (0.3)
0 ms	Syllable	574 (12)	623 (12)	0.5 (0.3)	1.4 (0.4)
	Body	606 (13)	624 (12)	1.3 (0.4)	1.4 (0.4)
	Rhyme	605 (13)	613 (12)	0.9 (0.3)	1.0 (0.4)
+100 ms	Syllable	580 (12)	594 (13)	0.8 (0.4)	1.4 (0.4)
	Body	587 (12)	599 (14)	0.4 (0.2)	0.8 (0.4)
	Rhyme	594 (12)	597 (13)	1.0 (0.5)	1.5 (0.5)

Table 2 Regression coefficients, standard errors, *p* values and Cohen's *d* of the factor target-distractor relatedness in the LMEM analyses

SOA	Type of relatedness	<i>b</i>	<i>SE</i>	<i>p</i>	<i>d</i>
-100 ms	Syllable	0.024	0.006	<0.001*	0.171
	Body	-0.001	0.006	0.848	-0.008
	Rhyme	-0.006	0.006	0.363	-0.042
0 ms	Syllable	0.068	0.011	<0.001*	0.453
	Body	0.028	0.008	0.001*	0.197
	Rhyme	0.012	0.007	0.081	0.080
+100 ms	Syllable	0.015	0.008	0.078	0.098
	Body	0.013	0.007	0.084	0.083
	Rhyme	0.002	0.008	0.766	0.015

**p* value significant at the 0.05 level after the Holm–Bonferroni correction for multiple comparisons

three-way interaction of SOA, type of relatedness, and target-distractor relatedness was not significant ($\chi^2_{(2)} = 0.59$, $p = 0.744$) and that the two-way interaction of type of relatedness and target-distractor relatedness was significant ($\chi^2_{(1)} = 3.93$, $p = 0.047$), suggesting that the body effect was larger than the rhyme effect.

Discussion

To distinguish the proximate phonological units in Mandarin spoken word production, the time course of syllabic and subsyllabic processing was investigated with the PWI paradigm in the current study. Compared with unrelated distractors, syllable-related distractors significantly facilitated the naming responses at -100- and 0-ms SOAs, whereas a significant effect of subsyllabic relatedness only appeared at 0-ms SOA (i.e., body-related distractors showed a facilitation effect at 0-ms SOA, and rhyme-related distractors showed null effects at all SOAs).

As predicted by the model of Roelofs (2015), the effect of syllabic relatedness occurred earlier than that of subsyllabic relatedness, supporting atonal syllables as the proximate phonological units in Mandarin spoken word production (O'Seaghdha et al., 2010). Our finding constitutes a nice contrast with the finding of Meyer and Schriefers (1991) in a Dutch PWI study with auditory distractors. For monosyllabic targets, their begin-related distractors (comparable to our body-related distractors) showed facilitation effects at -150-, 0- and +150-ms SOAs, whereas end-related distractors (comparable to our rhyme-related distractors) showed facilitation effects at 0- and +150-ms SOAs only. Importantly, the body effect occurred earlier in Dutch (i.e., since -150-ms SOA) than in Mandarin word production (i.e., since 0-ms SOA), which can be accounted for by the proximate units principle. In our Mandarin PWI study, atonal syllables are firstly

retrieved as the proximate units, so the facilitation effect of body-relatedness originated from the subsequent segmental specification process and thus occurred later relative to the case of Dutch word production, where segments are the proximate units. Our null effect of body-relatedness at -100-ms SOA is inconsistent with the alternative proposal, which assumes that both syllable-related and body-related effects should be arising from the same processing stage. The nonsignificance of our syllabic and subsyllabic effects at +100-ms SOA was probably due to relatively weaker activation of phonological information with visually presented distractors (i.e., relative to auditory distractors in the study of Meyer and Schriefers).

Interestingly, the effect of rhyme-relatedness failed to reach significance at all SOAs although the related pattern is very similar to that of body-relatedness. Previous Cantonese PWI studies also have found that the rhyme effect is not very robust with visual distractors (Wong & Chen, 2008, 2009) but more reliable with auditory distractors (Wong & Chen, 2008, 2015). Thus, it is not surprising that non-significant rhyme effects were obtained in the current study with visual distractors. The advantage of body-relatedness over rhyme-relatedness suggests the significance of word-initial segments even in Mandarin word production, which is inconsistent with Roelofs' (2015) assumption that segments within a syllable are selected in parallel during Mandarin word production. In other words, serial processing of segmental information seems to be universal across Germanic languages and Chinese. Future studies are needed to verify the reliability of this finding (e.g., using the ERP technique with exceptionally high temporal resolution).

Conclusions

The current results provide strong evidence for the proximate units principle, from a new perspective (i.e., the time course of syllabic and subsyllabic processing). While the proximate phonological units are language-specific, serial processing of segmental information seems to be universal across Germanic languages and Chinese.

Acknowledgements This research was supported by a Direct Grant for Research from the Chinese University of Hong Kong (4052098).

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.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.

- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, 14, 177–208.
- Chen, T. M., & Chen, J. Y. (2013). The syllable as the proximate unit in Mandarin Chinese word production: An intrinsic or accidental property of the production system? *Psychonomic Bulletin & Review*, 20(1), 154–162.
- Chen, J. Y., Chen, T. M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language*, 46(4), 751–781.
- Chen, J. Y., O'Seaghdha, P. G., & Chen, T. M. (2016). The primacy of abstract syllables in Chinese word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(5), 825.
- Da, J. (2004). A corpus-based study of character and bigram frequencies in Chinese e-texts and its implications for Chinese language instruction. In P. Zhang, T. W. Xie, & J. Xu (Eds.), *Proceedings of the Fourth International Conference on New Technologies in Teaching and Learning Chinese* (pp. 501–511). Beijing: Tsinghua University Press.
- 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. (2009). Exploring phonological encoding through repeated segments. *Language and Cognitive Processes*, 24(5), 685–712.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321.
- Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of prediction from a connectionist model. *Journal of Memory and Language*, 27, 124–142.
- Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, 2, 255.
- Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1–2), 101–144.
- Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, 68, 601–625.
- Kureta, Y., Fushimi, T., & Tatsumi, I. F. (2006). The functional unit in phonological encoding: Evidence for moraic representation in native Japanese speakers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(5), 1102.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2016). lmerTest: Tests in linear mixed effects models. *R Package Version 2.0-33*. Available at: <https://CRAN.R-project.org/package=lmerTest>
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge: MIT Press.
- Levelt, W. J. M. (1999). Models of word production. *Trends in Cognitive Sciences*, 3(6), 223–232.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75.
- 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(4), 563–578.
- Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, 30, 69–89.
- Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 1146–1160.
- O'Seaghdha, P. G., Chen, J. Y., & Chen, T. M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition*, 115(2), 282–302.
- R development core team. (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org>
- Roelofs, A. (2006). The influence of spelling on phonological encoding in word reading, object naming, and word generation. *Psychonomic Bulletin & Review*, 13(1), 33–37.
- 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.
- Santiago, J. (2000). Implicit priming of picture naming: A theoretical and methodological note on the implicit priming task. *Psicológica*, 21(1), 39–59.
- Schiller, N. O. (2008). The masked onset priming effect in picture naming. *Cognition*, 106(2), 952–962.
- Timmer, K., Ganushchak, L. Y., Mitlina, Y., & Schiller, N. O. (2014). Trial by trial: selecting first or second language phonology of a visually masked word. *Language, Cognition and Neuroscience*, 29(9), 1059–1069.
- Timmer, K., Vahid-Gharavi, N., & Schiller, N. O. (2012). Reading aloud in Persian: ERP evidence for an early locus of the masked onset priming effect. *Brain and Language*, 122(1), 34–41.
- Verdonschot, R. G., Kiyama, S., Tamaoka, K., Kinoshita, S., La Heij, W., & Schiller, N. O. (2011). The functional unit of Japanese word naming: Evidence from masked priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(6), 1458.
- Verdonschot, R. G., Lai, J., Chen, F., Tamaoka, K., & Schiller, N. O. (2015). Constructing initial phonology in Mandarin Chinese: Syllabic or subsyllabic? A masked priming investigation. *Japanese Psychological Research*, 57(1), 61–68.
- 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.
- Wong, A. W. K., & Chen, H.-C. (2008). Processing segmental and prosodic information in Cantonese word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1172–1190.
- Wong, A. W. K., & Chen, H.-C. (2009). What are effective phonological units in Cantonese spoken word planning? *Psychonomic Bulletin & Review*, 16(5), 888–892.
- Wong, A. W. K., & Chen, H.-C. (2015). Processing segmental and prosodic information in spoken word planning: Further evidence from Cantonese Chinese. *Japanese Psychological Research*, 57(1), 69–80.
- Wong, A. W. K., Huang, J., & Chen, H.-C. (2012). Phonological units in spoken word production: Insights from Cantonese. *PLoS One*, 7, e48776.
- Wong, A. W. K., Wang, J., Ng, T. Y., & Chen, H.-C. (2016). Syllabic encoding during overt speech production in Cantonese: Evidence from temporal brain responses. *Brain Research*, 1648, 101–109.
- You, W., Zhang, Q., & Verdonschot, R. G. (2012). Masked syllable priming effects in word and picture naming in Chinese. *PLoS One*, 7(10), e46595.
- Zhang, Q. (2008). Phonological encoding in monosyllabic and bisyllabic Mandarin word production: Implicit priming paradigm study. *Acta Psychologica Sinica*, 3, 002.
- Zhang, Q., & Yang, Y. (2005). The phonological planning unit in Chinese monosyllabic word production. *Psychological Science (China)*, 28(2), 374–378.