Brain & Language 144 (2015) 16–25

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Seriality of semantic and phonological processes during overt speech in Mandarin as revealed by event-related brain potentials

Xuebing Zhu[b,d], Markus F. Damian[c], Qingfang Zhang[a,b,*]

[a] Department of Psychology, Renmin University of China, Beijing, China
[b] Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China
[c] School of Experimental Psychology, University of Bristol, UK
[d] Institute of Linguistic Studies, Shanghai International Studies University, Shanghai, China

ARTICLE INFO

Article history:
Received 23 June 2014
Accepted 19 March 2015
Available online 13 April 2015

Keywords:
Speech production
Semantic interference
Phonological facilitation
Electroencephalography

ABSTRACT

How is information transmitted across semantic and phonological levels in spoken word production? Recent evidence from speakers of Western languages such as English and Dutch suggests non-discrete transmission, but it is not clear whether this view can be generalized to other languages such as Mandarin, given potential differences in phonological encoding across languages. The present study used Mandarin speakers and combined a behavioral picture–word interference task with event-related potentials. The design factorially crossed semantic and phonological relatedness. Results showed semantic and phonological effects both in behavioral and electrophysiological measurements, with statistical additivity in latencies, and discrete time signatures (250–450 ms and 450–600 ms after picture onset for the semantic and phonological condition, respectively). Overall, results suggest that in Mandarin spoken production, information is transmitted from semantic to phonological levels in a sequential fashion. Hence, temporal signatures associated with spoken word production might differ depending on target language.

1. Introduction

Speaking, as a highly skilled behavior in daily life, is marked by astonishingly high-speed retrieval from the mental lexicon and low error rates (Levelt, 1992). One of the key requirements of spoken production is to select an appropriate target at a given time and to focus the execution of goal-directed articulation. Over the past few decades, the speech production system has been envisaged as a system of interrelated layers of mental representations, such as semantic, syntactic and phonological codes (Caramazza, 1997; Dell, 1986; Levelt, Roelofs, & Meyer, 1999). A central theoretical issue concerns how information flows within this cognitive system and its underlying neural implementation in speech production (Abdel Rahman & Sommer, 2003; Caramazza, 1997; Starreveld & La Heij, 1995; Starreveld & La Heij, 1996a, 1996b). Serial-discrete models (Levelt et al., 1999) argue that only a single selected lexical-semantic/syntactic node ("lemma") spreads its activation to the phonological level, and semantic processing must be completed before phonological processing. Non-serial models dispute some of these assumptions: cascaded models (Humphreys, Riddoch, & Quinlan, 1988; Morsella & Miozzo, 2002) propose that multiple lexical-semantic candidates which are co-activated during retrieval of the target word transmit activation to the phonological level. Interactive models (Dell, 1986) additionally assume that transmission of activation between semantic and phonological encoding is bidirectional. In both cascaded and interactive but not in serial models, phonological processing can begin on the basis of early partial information provided by semantic processes.

Recent empirical findings provide support against a strictly serial view, and for some degree of cascadedness. For instance, in a task in which two line drawings of objects are superimposed and one is to be named based on its color, a facilitation effect is observed when target and distractor objects overlap in their phonemes (Meyer & Damian, 2007; Morsella & Miozzo, 2002; but see Jescheniak et al., 2009). Similarly, when the color of a line drawing is named while ignoring the object, priming is found when target color and object names overlap in their form (Kuipers & Heij, 2009; Navarrete & Costa, 2005; note that henceforth, we use the term "form" to refer to surface properties – sound or spelling – of lexical items). These findings suggest that multiple lexical candidates are phonologically activated, which contradicts a central tenet of the seriality view. At the same time, the existing evidence suggests that cascading is not "universal" such that all activated units at higher level necessary transmit activation to lower levels. For instance, Kuipers and Heij (2009; see also Dumay & Damian,

---

[*] Corresponding author at: Department of Psychology, Renmin University of China, 59 Zhongguancun Street, Haidian District, Beijing 100872, China.
E-mail address: qingfang.zhang@ruc.edu.cn (Q. Zhang).

http://dx.doi.org/10.1016/j.bandl.2015.03.007
0093-934X/

2011) have suggested a principle of "limited cascadedness" according to which properties associated with the identity of the target dimension (such as an object's name) will cascade to the form level, whereas modifying dimensions (such as its color or size) will not. Cascadedness might additionally be modulated by factors such as attention and task demands (Mädebach, Jescheniak, Oppermann, & Schriefers, 2011). Nevertheless, overall the currently available evidence on phonological activation of multiple lexical codes contradicts a strictly serial view of lexical access in spoken production, and suggests that at least under some circumstances, multiple lexical candidates can transmit activation to the phonological level.

A different way to tackle the issue of activation transmission between semantic and phonological stages in word production is to employ one of the most widely used paradigms in speech production, namely the picture–word interference (hereafter PWI) task. In this task, participants are instructed to name a target picture while ignoring a distractor word which is either visually superimposed on the target, or presented in spoken format. A semantic relationship between a context word such as "dog" and a target picture such as "cat" slows naming relative to an unrelated word (e.g., "table"), whereas a phonological relationship (e.g., "key") speeds up latencies (Glaser & Düngelhoff, 1984; Schriefers, Meyer, & Levelt, 1990; Starreveld & La Heij, 1995). These two phenomena have been termed "semantic interference" and "phonological facilitation", and numerous studies with PWI have shown that those effects provide important constraints on models of speech production. For instance, the prominent WEAVER++ model (Word-form Encoding by Activation and VERification; Roelofs, 1992, 1997) postulates that semantic interference arises at a processing stage of lexical-semantic retrieval, where co-activated entries ("lemmas") engage in competition with one another. By contrast, phonological facilitation arises mainly at the segmental level (with the possibility of weaker priming also arising at the morpheme level): distractor words activate corresponding segments, and therefore partially pre-activate the segments which form the target response, resulting in faster encoding for related than unrelated distractors.

The PWI task can be used to explore how semantic and phonological processing stages relate to each other, via employing not only the semantically and phonologically related distractors, but additionally, by including "mixed" distractors which are semantically as well as form-related (e.g., picture: "cat"; distractor: "calf"). Factorially crossing semantic and phonological relatedness allows to determine whether the two variables have statistically additive or interactive effects. Based on "additive factors logic" (Sternberg, 1969) the idea is that if the two experimental variables exert statistically additive effects, then they affect different and separate processing stages, with strictly serial information transmission between the stages. By contrast, if the two variables show non-additive effects, then either they act on a single processing stage, or they affect two processing stages but these two stages are themselves closely related in terms of processing, for instance, via cascaded transmission, or interactivity (i.e., feedback). The currently available results clearly demonstrate a statistical interaction between semantic and phonological relatedness in PWI tasks (Damian & Martin, 1999; Starreveld & La Heij, 1995, 1996b; Taylor & Burke, 2002). More specifically, the pattern that is typically found is that the semantic interference effect is attenuated when a distractor is also form-related to the target name; hence, "rabbit-rat" acts predominantly as a form-related pair whereas the semantic effect which should arise from shared category membership is much diminished. This general pattern has been interpreted as supporting non-discrete models of word production (Damian & Martin, 1999).

1.1. Event-Related Potentials (ERPs) and spoken word production

The bulk of evidence concerning spoken word production has traditionally come from chronometric studies. However, response latencies merely index the "end point" in a cascade of mental processes which precede initiation of a response. Hence questions associated with the time course of various types of mental activities (i.e., how processing stages unfold over time) are difficult to address with chronometry. A complementary approach is to employ electroencephalography (EEG). By tracking electrical activity along the scalp, brain responses to specific sensory, cognitive, or motor events can be assessed millisecond-by-millisecond as they unfold.

The EEG approach is well-established in various areas of language research. However, until relatively recently it was assumed that EEG could not be measured for spoken responses because artefacts from muscular activity associated with articulation distort the signal (Wohlert, 1993). Hence, many empirical studies used manual responses as a substitute for spoken ones (Van Turennout, Hagoort, & Brown, 1997; Zhang & Zhu, 2011). Yet, it has recently become clear that the problems associated with overt articulation are tractable, and a number of studies have combined spoken production tasks with EEG (e.g., Blackford, Holcomb, Grainger, & Kuperberg, 2012; for reviews, see Ganushchak, Christoffels, & Schiller, 2011; Indefrey, 2011; Strijkers & Costa, 2011) and MEG (magnetoencephalography, Levelt, Praamstra, Meyer, Helenius, & Salmelin, 1998; Maess, Friederici, Damian, Meyer, & Levelt, 2002; Salmelin, Schnitzler, Schmitz, & Freund, 2000). In these EEG studies, classical ERP components have been replicated during overt naming. For instance, the N400 complex, first reported by Kutas and Hillyard (1980) in semantic violations, is widely interpreted as a central index of lexical and semantic processing (for review, see Lau, Phillips, & Poeppel, 2008) and phonological processing (i.e., Chen, Lee, Kuo, Hung, & Cheng, 2010; Valdes-Sosa et al., 1993) in language comprehension. Importantly, this negative-ongoing waveform apparently also reflects phonological processing in spoken production (Blackford et al., 2012; Dell'Acqua et al., 2010), and hence indicates priming resulting from the convergence of phonological processing from pictures and distractors in the PWI task. Moreover, Dell'Acqua et al. (2010) used ERPs combined with a subtraction technique to explore the time course of activation of semantic and phonological representations in the PWI task. Difference ERP waveforms were generated in the semantic condition and in the phonological condition by subtracting ERP waveforms in the unrelated condition. In the time window of 250–450 ms, they found significant differences on mean amplitude for both semantic and phonological relatedness. Furthermore, the peak latencies of semantically related distractors (320 ms) coincided temporally with those of phonologically related distractors (321 ms). These estimates are difficult to reconcile with a strictly serial information transmission model (see previous section) which would predict a more sequential pattern.

We should note that in the still limited literature on EEG studies exploring spoken production, it is at present typical to focus on ERP differences between experimental and baseline conditions, rather than (or sometimes in addition to) identifying components such as N400, etc., which are associated with specific particular mental processes. Undoubtedly, this is the case because EEG research on production is relatively less well developed than corresponding research on comprehension.

1.2. Cross-linguistic differences in phonological encoding?

Much of our understanding of how speakers plan and produce words is based on evidence from Indo-European languages such as English, German, Spanish and Dutch. Relatively little attention has been paid to the possibility that the architecture of word production, and specifically, the mechanisms of phonological encoding, might differ between languages. For instance, the WEAVER++ model (Roelofs, 1997) postulates that the form network for typical Western languages such as English and Dutch encompasses three levels: [1] the morpheme corresponding to the target is activated, [2] segments corresponding to the target name are chosen (labelled for order); simultaneously, a metrical frame is activated which conveys information about stress patterns, [3] segments and metrical frame are merged into "syllable motor programs" (see Fig. 1). However, this architecture might not be universal across languages, and it has recently been suggested (O'Seaghdha, Chen, & Chen, 2010) that languages might differ in the "proximate unit" of phonological encoding (i.e., the primary selectable unit below the word level, carrying particular salience as a speech planning unit). For instance, for speakers of Chinese Mandarin, syllables have particular prominence as Mandarin has a limited inventory of syllables with relatively simple structure (e.g., V: a; CV: ba; CVC: ban, VC: an), and clear syllable boundaries. Furthermore, unlike in English or Dutch where re-syllabification is allowed under certain circumstance (e.g., "get it" pronounced as "ge.tit"), syllables in Chinese are rarely re-syllabified. Furthermore, the Chinese orthographic system is non-alphabetic and maps characters onto spoken syllables, but in contrast to alphabetic languages does not explicitly represent speech sounds.

Intriguingly, recent empirical findings using the "implicit priming" task with Mandarin speakers support the possibility of differential proximate units between Indo-European languages and Mandarin. In this task, numerous reported studies have documented priming effects based on manipulating word-initial phonemes (e.g., Meyer, 1991). However, no such effects of segmental overlap were found with Mandarin speakers (e.g., Chen, Chen, & Dell, 2002; O'Seaghdha et al., 2010; You, Zhang, & Verdonschot, 2012) for whom only syllabic overlap generated priming. Furthermore, syllabic overlap without shared tone resulted in weaker priming, and shared tone by itself generated no priming. Based on these results, Roelofs (2015) recently postulated for Mandarin Chinese phonological encoding the following four levels (see Fig. 1): [1] a morpheme corresponding to the target is activated, [2] atonal syllable nodes are activated; simultaneously, a tonal frame is activated, [3] segments are activated, [4] segments and tonal frames are merged into syllable motor programs. If this framework is accurate, phonological encoding for Mandarin speakers would involve an additional processing layer compared to Western languages.

1.3. *The current study – exploring spoken word production in mandarin speakers via ERPs*

Given the possibility that languages might differ to some degree in how phonological encoding is carried out, the current study targeted word production in Mandarin speakers via a PWI task. Simultaneously, ERPs were measured. As in previous studies based on "additive factors logic" which were conducted on speakers of English and Dutch (Damian & Martin, 1999; Starreveld & La Heij, 1995, 1996b; Taylor & Burke, 2002), semantic and phonological overlap between distractors and target names was factorially crossed. As summarized above, the earlier PWI studies had consistently indicated non-additivity in latencies with regard to semantic and phonological effects. The aim of the current study was to investigate whether this was also the case in spoken Mandarin.

If spoken production in non-alphabetic languages such as Mandarin follows the same principles as those established for alphabetic languages, we would predict (i) a statistical interaction between semantic and phonological relatedness in response times, and (ii) a largely "parallel" time signature of both types of overlap in ERPs. By contrast, if phonological encoding in Mandarin is accomplished via a different "proximate unit" and involves a different (and possibly more complex) sequence of processing stages (see Roelofs, 2015) then it may well be that results differ from those found with Indo-European languages. From a serial framework of word production, for Mandarin speakers we may predict (i) an additive pattern of semantic and phonological effects in response latencies, and (ii) a strongly sequential pattern of semantic and phonological effect in ERPs, with an "early" semantic stage followed by a "later" phonological stage.

In revisiting the relation between semantic and phonological variables in PWI tasks, we felt it was important to ensure that form-based effects could clearly be attributed to the processing stage of phonological encoding. Hence, in our experiment, phonologically related distractors were chosen such that they shared a syllable (but differed in tone) with the target; by contrast, orthographic overlap was entirely avoided (in languages with alphabetic orthographic systems, the two dimensions are almost unavoidably confounded). The issue of phonological vs orthographic effects in PWI will be highlighted in much greater detail in the Discussion.

2. Materials and methods

2.1. *Participants*

Thirty-seven native Mandarin speakers (17 males; age range 18–24 year, mean = 21.6 year) from Beijing Forest University

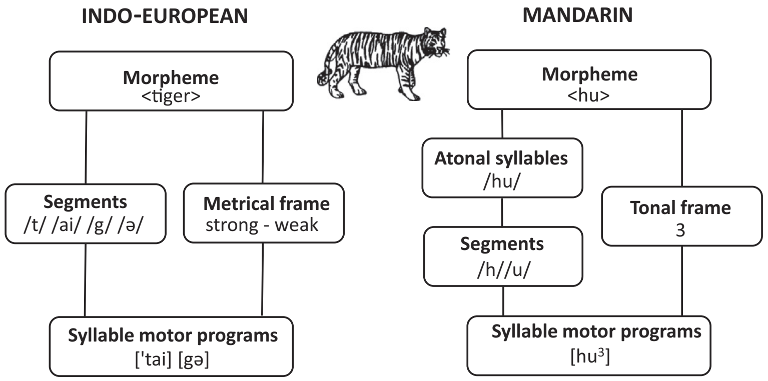

Fig. 1. Processing stages involved in phonological encoding of the concept "tiger" in English and Mandarin.

participated in the study. All participants were neurologically healthy, right-handed, with normal or corrected-to-normal vision.

2.2. Stimuli and design

Twenty-seven black and white line pictures (23 targets and four practice items) were selected from a database created by Zhang and Yang (2003). Each target picture was paired with four distractor words. A semantic word (condition S) was chosen that belonged to the same category as the target picture but had no phonological overlap (i.e., 毛衣, /mao2yi1/, "sweater" as target, 衬衫, /chen4shan1/, "shirt" as a distractor). A phonological word (P) was chosen that shared a syllable which always differed in tone with the first character of the picture name (i.e., 茂盛, /mao4sheng5/, "luxuriant"). A semantic plus phonological word (SP) was chosen that belonged to same category as the target and shared a syllable with a different tone with the first character of the picture name (i.e., 帽子, /mao4zi5/, "hat"). Finally, an unrelated word (U) was selected that stood in no obvious relationship to the target picture (i.e., 座位 /zuo4wei4/, "seat"). Among the 23 targets, six pictures had monosyllabic names, and seventeen pictures had disyllabic names. Distractors in each condition were statistically matched for number of strokes and written frequency based on normative information from the database of Chinese Lexicon (2003).

In a design in which semantic and phonological relatedness are crossed, two form-related conditions exist (P and SP), as well as two semantically-related conditions (S and SP). Validity of the interpretation of the outcome critically depends on the assumption that the two corresponding conditions are closely matched in terms of overlap with the target. The degree of form overlap between target and distractor was identical in the P and the SP condition: in both conditions, target and distractor shared a single syllable whose tone always differed. Concerning the degree of semantic overlap in the S and the SP condition, a debate exists in the literature on whether closely related distractors cause more (Vigliocco, Vinson, Lewis, & Garrett, 2004), less (Mahon, Costa, Peterson, Vargas, & Caramazza, 2007), or the same (Hutson & Damian, 2014) amount of interference than distractors which are only weakly related. Because this issue remains unresolved, it is important to ensure that the two related conditions have similar degrees of semantic overlap. We assessed the relative degree of semantic relatedness via a rating task carried out by sixteen native Mandarin speakers (5 males, age from 22 to 32 years old) who did not take part in the main experiments. Target picture names were paired with their corresponding distractor words, and presented side-by-side in random order. The word pairs were rated on a 5-point scale, with 5 indicating that word pairs were highly semantically related and 1 indicating that word pairs were semantically unrelated. For the S condition, the mean value was 4.36 with a range of 3.65 to 5 across subjects (SD = 0.49) and a range of 3.50 to 4.36 across items (SD = 0.30). For the SP condition, the mean value was 4.01 with a range of 3.22 to 4.87 across subjects (SD = 0.55) and a range of 2.88 to 4.56 across items (SD = 0.50). For the P condition, the mean was 1.27 with a range of 1 to 1.87 across subjects (SD = 0.22) and a range of 1 to 2.50 across items (SD = 0.34). For the U condition, the mean was 1.18 with a range of 1 to 1.65 across subjects (SD = 0.21) and a range of 1 to 1.69 (SD = 0.19). The rating scores in both S and SP conditions were high and close to each other, indicating that the two related conditions were well-matched in terms of semantic overlap.

Further potential variables which should be taken into account are imageability and/or concreteness. Previous studies have shown that in tasks such as lexical decision, highly imageable or concrete words are processed more quickly than less imageable or abstract words (Tsai et al., 2009; West & Holcomb, 2000; Zhang, Guo, Ding, & Wang, 2006). To investigate the possibility of differences among the experimental conditions in our Materials, a rating task on imageability and concreteness was carried out by fourteen native Mandarin speakers (4 males, age from 24 to 30 years old) who did not take part in the main experiment. Distractor words were rated on a 5-point scale, with 5 indicating that words were highly imageable (or concrete) and 1 indicating that words were difficult to imagine (or abstract). For the U condition, the mean value was 4.03 ± .36 for imageability and 4.09 ± .38 for concreteness. For the S condition, the mean value was 4.81 ± .30 for imageability and 4.91 ± .13 for concreteness. For the SP condition, the mean value was 4.71 ± .29 for imageability and 4.84 ± .16 for concreteness. For the P condition, the mean value was 3.51 ± .39 for imageability and 3.54 ± .41 for concreteness. Hence, rating scores in all conditions were high (i.e., distractor words were on average easy to imagine and concrete) and relatively similar to each other.

The experimental design factorially crossed semantic relatedness (related vs. unrelated) and phonological relatedness (related vs. unrelated) as within-participants and within-items variables (see Damian & Martin, 1999; Starreveld & La Heij, 1995, 1996b; Taylor & Burke, 2002 for a similar design).Within an experimental block, a participant saw the 23 target pictures four times (once under each condition), for a total of 92 trials. This block was repeated three times, thus the entire experiment consisted of 276 experimental trials. The order of items within a block was pseudo-randomized for each participant with the constraint that a particular target did not re-occur for at least five trials, and the first phoneme of a target name was never the same on consecutive trials. It should be noted that a design in which each target is presented and named multiple times is quite common in research on spoken word production, and generally considered to be unproblematic. Nevertheless, in the latencies analysis reported below, we included "Repetition" as an additional factor to check for potential effects of multiple target presentation.

2.3. Procedure and apparatus

Participants were tested individually in front of a computer screen in a sound-proof room. Participants first were asked to familiarize themselves with the experimental stimuli by viewing each target for 3000 ms with the correct name printed below. Then, participants were asked to name the pictures.

Each trial involved the following sequence: A fixation point (+) presented in the middle of the screen for 500 ms, followed by a blank screen for 500 ms. Then, the target picture plus distractor word was presented. Target pictures and distractors disappeared when participants initiated a voice response. Participants were asked to name the target as quickly and accurately as possible while ignoring the distractor. An inter-trial interval of 2000 ms concluded each trial. The experiment took about 40 min in total. The experiment was performed using E-Prime Professional Software.

2.4. EEG acquisition and analysis

EEG signals were recorded with 64 electrodes secured in an elastic cap (Electro Cap International). The left mastoid electrode served as reference. All electrode impedances were kept below 5 kΩ. Electrophysiological signals were amplified with a band-pass filter of 0.05 and 100 Hz and digitized continuously at a rate of 500 Hz. The EEG-data were re-referenced off-line to the average of both mastoids (Wang, Bastiaansen, Yang, & Hagoort, 2011; Zhang & Zhu, 2011), for a similar procedure) and low-pass filtered (high cutoff = 30 Hz, 24 dB/cot). The data were segmented from 200 ms before to 1000 ms after the onset of the pictures, with baseline correction from −200 to 0 ms preceding pictures onset and any artifact exceeding ±100 µV. Prior to off-line averaging, all single-trial waveforms were screened for eye movements, electrode drifting, amplifier blocking and artifacts. Seven participants were excluded from the EEG analysis because of large electrode drift and excessive artifacts. Behavioral and EEG data analyses were carried out for the remaining 30 participants. Only trials with correct response were considered for ERP analyses. To avoid contamination of the ERPs due to muscular and mouth movement activity, trials with a naming response faster than 600 ms (6.6%) were removed (Costa, Strijkers, Martin, & Thierry, 2009). Also, to avoid contamination from very slow responses, trials with a naming response slower than 2000 ms (0.24%) were removed. Mean amplitude measurements were performed in three consecutive time windows: 0–250 ms, 250–450 ms and 450–600 ms, which were chosen based on the visible ERP peaks (i.e., Costa et al., 2009), as well as on previous studies (i.e., Dell'Acqua et al., 2010). Nine regions of interest (ROIs) were conducted, the voltage of each lateral ROI was the mean amplitude of three electrodes, i.e., left-anterior (pooled F3, F5, FC3), mid-anterior (Fz), right-anterior (pooled F4, F6, FC4), left-central (pooled C3, C5, CP3), mid-central (Cz), right-central (pooled C4, C6, CP4), left-posterior (pooled P3, P5, PO3), mid-posterior (Pz), and right-posterior (pooled P4, P6, PO4) regions.

A repeated measures analysis of variance (ANOVA) was performed on the ERP amplitude means with the factors semantic relatedness, phonological relatedness and ROIs (left-anterior, left-central, left-posterior, middle-anterior, middle-central, middle-posterior, right-anterior, right-central, and right-posterior), conducted separately for each time window. Greenhouse–Geisser correction was applied where appropriate. Furthermore, to examine the effects of semantic, phonological and mixed distractor conditions, planned comparisons were conducted to compare the unrelated to each of the other conditions. Onset latencies were determined by conducting paired t-tests at every sampling point (2 ms) starting from picture onset (0 ms) until 600 ms. The widely used technique by Guthrie and Buchwald (1991) suggested that for our data, differences could be considered reliable when at least a sequence of 22 consecutive milliseconds exceeded the 0.05 significance level in our case. Therefore, the first point at which ERPs started to diverge was taken as the onset latency.

3. Results

3.1. Behavioral results

Data from incorrect responses (0.89%), other responses such as mouth clicks (1.01%), naming latencies longer than 2000 ms or shorter than 200 ms (0.24%), and those deviating by more than three standard deviations from a participant's mean (1.40%) were removed from all analyses. Fig. 2A presents the mean pictures naming latencies and standard errors by Distractor Type, suggesting semantic and phonological effects which are additive.

The main objective of the experiment was to identify a potential interaction between semantic and phonological relatedness. Because semantic relatedness and phonological relatedness were factorially crossed in this experiment, "additive factors logic" (Sternberg, 1969) could be applied.

ANOVAs were performed on the response latency means with semantic relatedness and phonological relatedness as within-participants and within-items variables. A significant effect of semantic relatedness was found, $F1(1,29) = 42.12$, $MSE = 612$, $p < .001$; $F2(1,22) = 42.10$, $MSE = 493$, $p < .001$, and so was phonological relatedness, $F1(1,29) = 47.54$, $MSE = 434$, $p < .001$; $F2(1,22) = 31.38$, $MSE = 517$, $p < .05$. Crucially, semantic and phonological relatedness did not statistically interact, $F1(1,29) = 0.66$, $MSE = 474$, $p = .43$; $F2(1,15) = 0.46$, $MSE = 580$, $p = .51$. We conducted a further ANOVA in which we additionally included the variable "Repetition". This analysis showed a significant effect of Repetition, $F1(2,58) = 4.76$, $MSE = 10726$, $p < .05$; $F2(2,44) = 32.24$, $MSE = 853$, $p < .001$. As in the initial analysis, we found significant effects of semantic relatedness, $F1(1,29) = 42.45$, $MSE = 1849$, $p < .001$; $F2(1,22) = 41.93$, $MSE = 1499$, $p < .001$, and of phonological relatedness, $F1(1,29) = 48.23$, $MSE = 1262$, $p < .001$; $F2(1,22) = 31.45$, $MSE = 1539$, $p < .001$, and no interaction. Most importantly, repetition did not interact with either type of relatedness, nor was a three-way interaction found (all $ps > .39$), suggesting that the statistical pattern was stable across repetitions.

The additive relationship between semantic and phonological relatedness in these analysis constitutes, technically speaking, a null finding (i.e., a non-significant interaction, although in the context of highly significant main effects). Within a conventional null-hypothesis testing statistical approach, it is problematic to draw strong inferences from null findings. Bayesian analysis (e.g., Rouder, Morey, Speckman, & Province, 2012) provides an estimate of the degree of confidence one can have in a null finding, something which is difficult or impossible within a traditional statistical framework. Our results revealed that the model with only the two main effects was superior to the full model including the interaction, with a Bayes factor of 23.3. Values greater than 3 are considered "some evidence", greater than 10 "strong evidence", and greater than 30 "very strong evidence" (Jeffreys, 1961), hence the Bayesian analysis indicated "strong to very strong evidence" for the null interaction. This further underscores the notion that semantic and phonological variables exhibit an additive relationship.

Before we make strong conclusion based on the additivity, analyses on the potential influence of imageability and concreteness were conducted. We performed a linear mixed effects model (Baayen, Davidson, & Bates, 2008) including fixed effects of semantic relatedness (semantically related vs. unrelated), phonological relatedness (phonologically related vs. unrelated) and concreteness (or imageability), and by-subject and by-item random intercepts. Results showed that there was no improvement in the fit when the full model with a three-way interaction of semantic relatedness, phonological relatedness and imageability (or concreteness) compared with a model without the three-way interaction, $\chi^2(1, N = 7965) = .44$, $p = .51$ (for imageability) and χ^2 (1, $N = 7965) = .95$, $p = .33$ (for concreteness). Those results indicated that our result of additivity between semantic and phonological relatedness was not influenced by concreteness and imageability.

Error rates (overall less than 1%) were considered too low to allow for a meaningful statistical analysis.

3.2. Electrophysiological results

The main objective of the experiment was to identify a potential interaction between semantic and phonological relatedness, and to identify the time signature of both effects. To this aim, three consecutive time windows (0–250 ms, 250–450 ms and 450–600 ms) were chosen based on the visible ERP peaks as well as on previous studies (Costa et al., 2009; Dell'Acqua et al., 2010). Amplitude means were analyzed via ANOVAs conducted separately for each time window, with the variables semantic and phonological relatedness factorially crossed (as in the analysis of the behavioral results), and additionally including the variable ROI (left-anterior, left-central, left-posterior, middle-anterior, middle-central, middle-posterior, right-anterior, right-central, and right-posterior). Results are summarized in Table 1. At the very earliest time window (0–250 ms), no effect was observed in any of the electrodes in this early time window. Crucially, in the slightly later time window (250–450 ms), a main effect of semantic relatedness, but no phonological effect, was found, whereas in the latest time window

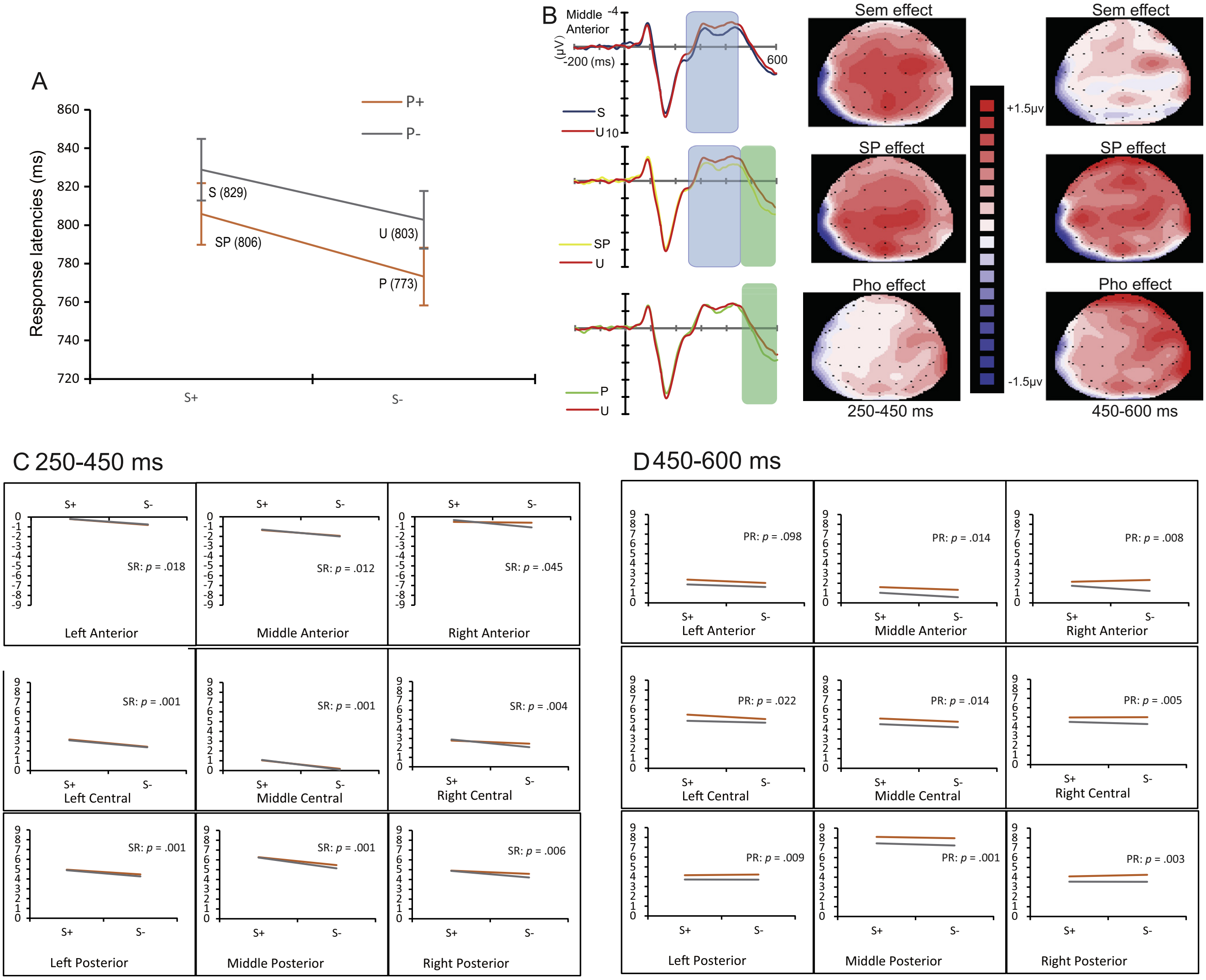

Fig. 2. (A) Mean picture naming latencies (with standard errors) dependent on semantic and phonological overlap with distractors. (B) Grand average ERP waveforms for distractor conditions at middle anterior and scalp distribution (related minus unrelated), with semantic effects indicated by blue shading in 250–450 ms and phonological effects indicated by green shading in 450–600 ms. Mean amplitude (μv) of different distractor conditions for nine regions of interest (ROIs) are shown in (C) 250–450 ms and (D) 450–600 ms ($p < .10$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$).

Table 1
Analysis of Variance for mean amplitude with ROI, Semantic Relatedness and Phonological relatedness in the time windows of 250–450 ms and 450–600 ms.

Source	*F*			
	df1	df2	250–450	450–600
ROI	8	232	22.38***	13.72**
SR	1	29	15.76***	0.55
PR	1	29	0.27	10.36**
ROI × SR	8	232	2.45*	1.41
ROI × PR	8	232	0.46	0.67
SR × PR	1	29	0.29	0.06
ROI × SR × PR	8	232	1.45	1.21

* $p < .05$.
** $p < .01$.
*** $p < .001$.

(450–600 ms), a main effect of phonological relatedness but no semantic effect was found. Semantic and phonological relatedness did not statistically interact under either time window.

To further explore this pattern, separate ANOVAs which crossed semantic and phonological relatedness were conducted on amplitude means under the two relevant time windows (250–450 ms and 450–600 ms) and for each ROI. Amplitude means are shown in Fig. 2C and D, along with information about statistical results. As can be seen, under the earlier time window (250–450 ms), a significant semantic effect was found at all ROIs, but no phonological effect emerged. By contrast, under the later time window (450–600 ms), a significant phonological effect was found under eight of the nine ROIs (the effect was marginally significant under the remaining ROI), but no semantic effect was found. Under neither time window nor ROI, a statistical interaction between semantic and phonological relatedness was found. As was the case for the behavioral results, the overall results imply that semantic and phonological variables exhibit an additive relationship, and the EEG results revealed that the two corresponding processing stages followed a sequential pattern.

The above analyses were based on pre-defined time windows. In a further analysis, we aimed to identify the temporal onset of semantic and phonological effects in our study. To this aim, we omitted the semantically as well as phonologically related condition from the analysis, and compared either the semantic or the phonological condition to their corresponding unrelated baselines, which identified onsets associated with either type of relationship.

Results showed a latency of 224 ms for the semantic condition, and of 506 ms for the phonological condition, i.e., a substantial differential in latencies associated with the two types of relationship. In a final analysis, we compared the semantically as well as phonologically related condition to the unrelated baseline, and two sequential waveforms emerged in different time windows: 250–450 ms and 450–600 ms (see Fig. 2D). Respective onset latencies were 236 ms and 494 ms, which are quite similar to the separate onsets of the semantic and phonological condition, respectively. Overall, the large difference in onset associated with the two types of relationship corroborates the sequential pattern of semantic followed by phonological effects suggested in the previous analysis.

4. Discussion

With a PWI task combined with EEG, we explored the time course and interplay between semantic and phonological variables in spoken word production. Behavioral results showed the classic semantic interference and phonological facilitation effects. More importantly, we obtained clear statistical additivity between semantic and phonological relatedness, which contrasts with previous studies in alphabetic languages.

In elucidating the behavioral additive pattern further, associated ERPs provide invaluable additional information based on the fine-grained temporal resolution. Across all conditions, classic exogenous P1/N1/P2 ERP components were found within a time window of 0–250 ms, but no effect was observed in any of the electrodes in this early time window. In the following, we investigated mean amplitudes and onset latencies associated with semantic and phonological processing under two consecutive time windows: 250–450 ms and 450–600 ms. Corresponding with our behavioral results, the electrophysiological data showed a dissociation of semantic and phonological processes. Wide scalp activity was associated with the semantic condition in the 250–450 ms time window only, but with the phonological condition in the 450–600 ms time window only. Within both time windows, the semantic and phonological ("mixed") manipulation exhibited a strictly "additive" pattern, i.e., at the "early" time window in which exclusively semantic effects (onset of 224 ms) were observed, effects in the "mixed" condition (onset of 236 ms) were very similar to those in the semantic condition but the simultaneous phonological overlap present in this condition was irrelevant. At the "later" time windows in which mainly phonological effects (onset of 506 ms) were found, "mixed" distractors acted as if they were only phonologically related (onset of 494 ms) and semantic overlap was irrelevant. Hence, results from ERPs showed a rather strict sequence of semantic and phonological effects, and no interaction between semantic and form overlap.

Our finding of N400 elicited by semantic overlap in the 250–450 ms window is overall in line with previous ERP studies in alphabetic languages. The general estimate from these studies is that lexical selection in spoken word production begins between 200 and 250 ms after stimulus onset (Indefrey, 2011; Indefrey & Levelt, 2004; Levelt et al., 1998; Maess et al., 2002). For instance, Costa et al. (2009) calculated point-by-point correlations to track the time course of the "cumulative semantic interference" effect (as a proxy for lexical selection) in the ERPs and estimated that lexical retrieval starts approximately 208 ms after picture onset and unfolds for about 180 ms. Piai, Roelofs, and van der Meij (2012) postulated 250 ms as the point at which the operation of word selection is initiated. In a PWI task reported by Dell'Acqua et al. (2010), the semantic manipulation elicited N400 in a time window of 250–450 ms. Evidence with MEG comes from the finding that brain responses between 150 and 225 ms after picture onset in left temporal regions reflect the difficulty with which lexical items are retrieved from the lexicon (Maess et al., 2002) Overall, the convergence of our results from Mandarin speakers with previous reports from alphabetic scripts indicates that the mechanism underlying semantic processing in spoken word production is general across languages.

More surprising is the finding that in our study, a phonological effect emerged in a "late" time window of 450–600 ms. The fact that effects of phonological overlap emerged at such a late time windows diverges from the temporal estimate for phonological encoding (275–445 ms) in a comprehensive meta-analysis of results from alphabetic languages (Indefrey & Levelt, 2004). In alphabetic studies, evidence on phonological effects, stemming from converging phonological activation between picture targets and distractors, emerged as a N400 modulation (Dell'Acqua et al., 2010; Jescheniak, Hahne, & Schriefers, 2003; Jescheniak, Schriefers, Garrett, & Friederici, 2002). In those studies, phonological relatedness was manipulated in segmental level, whereas we adopted syllabic overlap for the phonological condition in Chinese. Results showed that the negative difference wave between the ERP of the unrelated condition minus the ERP of the phonologically related condition was prominent in the time window of 450–600 ms and peaked around 500 ms (see Fig. 2D). Most importantly, using PWI, Dell'Acqua et al. (2010) found similar temporal signatures of semantic and phonological processing, with both types eliciting N400 (onset latency of 320 ms for semantic effect and 321 ms for phonological effect) in the 250–450 ms time window. By contrast, in our study, the semantic effect manifested itself in an early time window of 250–450 ms and was followed by a prominent phonological effect in a later time window of 450–600 ms, hence ERP results suggested a temporal dissociation between semantic and phonological stages in Chinese, with no sign of interaction between them. Overall, in behavioral and electrophysiological data, semantic and phonological variables exhibited an additive and sequential relationship. This inference contrasts with all previous studies, in which an interaction was obtained (see Section 1).

What do these results tell us about phonological encoding? One possibility is that the discrepancy between the current and previous results reflects cross-linguistic differences. As outlined in the Introduction, languages might differ in their primary phonological planning unit ("proximate unit"; O'Seaghdha et al., 2010). Indo-European languages such as English, German, Dutch, etc. presumably use phonological segments as the primary planning unit of spoken language, whereas spoken Mandarin is strongly oriented toward syllables as proximate units (which possibly, but not necessarily, results from a non-alphabetic orthographic system). Roelofs (2015) reported computational simulations of phonological encoding which contrasted English with Chinese (and Japanese) word production. Critically, whereas in Indo-European languages, segments are directly activated from morphemes, in Mandarin they are indirectly activated via an intermediate layer of "atonal syllables". Here we unfold two scenarios which would account for the dissociation between semantic and phonological effects in our key findings.

According to a first scenario, phonological effects in PWI primarily reflect facilitation at the segmental level (e.g., Roelofs, 1997). Distractor words activate corresponding segments and therefore partially pre-activate the segments which form the target response, resulting in faster encoding for related than unrelated distractors. Under this assumption, phonological effects for speakers of Mandarin emerge in ERPs delayed, relative to speakers of Indo-European languages, because phonological encoding involves an extra step in the former compared to the later (see Fig. 1). The additivity in latencies which we observed (see Fig. 2A) could be accounted for by an architecture which is "globally modular but locally interactive" (Dell & O'Seaghdha, 1991): adjacent processing stages interact (or information transmission is at least cascaded) yet such non-seriality is restricted such that overall, the system exhibits modular (serial) characteristics. Hence, for Indo-European languages, the statistical interaction in naming latencies could be accounted for by postulating interactivity between adjacent morpheme and segmental processing layers. In Mandarin, the additional processing layer of atonal syllables which intervenes between morpheme and segment levels renders these interactive (or cascading) properties irrelevant, and morphemes and segments act in an additive manner (as they did in our study).

According to a second scenario, the phonological effect in our study arose at the level of atonal syllables. Phonological overlap in our study implied shared syllables between distractor and target names, but these never carried the same tone. If Mandarin phonological encoding incorporates an atonal syllable level (as advocated, e.g., by Roelofs, 2015; see Fig. 1), it could be that related distractors pre-activated one of the target's atonal syllable representations, resulting in facilitation. It is not implausible to assume that, due to the relatively low number of atonal syllables and their high salience to Mandarin speakers, these are "informationally encapsulated". To account for the present results, one would have to assume that information transmission from the morpheme to the (syllable-based) phonological level is strictly discrete in processing terms.

It is acknowledged that it is at present difficult to adjudicate between the two scenarios, and further evidence is required to test the various assumptions. Unfortunately, it is not easy to identify experimental procedures which would allow clear insight into the various postulated processing levels. For instance, in the PWI task, it would be interesting to manipulate form overlap in various "grain sizes", corresponding to the various hypothetical levels of phonological encoding (atonal syllables, segments, tonal frames, tonally specified syllables). But these types of overlap are of course not independent (e.g., a tonal syllable implies overlap at the atonal syllable level as well as at the segmental level; an atonal syllable implies segmental overlap, etc.). Due to the scarcity of currently available methods and paradigms used to investigate spoken production, novel approaches are required to tackle this issue. But as we believe our own results demonstrate, the investigation of ERPs, in addition to response latencies, provides important information: the behavioral additivity between semantic and phonological variables in our experiment (see Fig. 2A) is interesting in and by itself, but information about the respective time course of the two types of relatedness (Fig. 2C and D) allows much stronger inferences concerning the differences between languages.

A final account of the discrepancy between the current and the previous results centers less on cross-linguistic differences in phonological encoding, but instead on the character of form overlap in the current study vs previous experiments. In Western languages with alphabetic scripts, phonology and orthography are strongly confounded. Hence phonological overlap in a PWI task typically implies orthographic relatedness and vice versa. Do form-based facilitation effects in PWI arise as a consequence of orthographic, or phonological, overlap (or both)? This question is difficult to answer with alphabetic languages (but see Lupker, 1982; Posnansky & Rayner, 1978; Underwood & Briggs, 1984). In a growing number of recent studies, researchers have therefore taken advantage of the fact that in languages with non-alphabetic scripts, orthographic and phonological relatedness can be dissociated, and have reported PWI studies with Chinese distractors (Bi, Xu, & Caramazza, 2009; Zhang, Chen, Weekes, & Yang, 2009; Zhang & Weekes, 2009; Zhao, La Heij, & Schiller, 2012). The pattern of results is somewhat complex, but it is clearly the case that both phonological and orthographic overlap between distractor and target name generate independent facilitation effects.

Recently, the possibility has been raised (Zhang et al., 2009) that phonological and orthographic effects in PWI tasks might arise at different processing levels, with phonological relatedness evoking priming at the phonological output level (as, e.g., proposed in the WEAVER model; Roelofs, 1997) whereas orthographic effects generate facilitation at the lexical-semantic ("lemma") level. Orthographically based priming effects could arise, for instance, because a printed distractor word evokes a cohort of orthographically similar neighbors. If the target is among them, as is the case for orthographically related distractors, then lexical-semantic (rather than form-based) retrieval of the target could be primed ("input priming"). If input priming accounts for facilitation effects from orthographically related distractors, then it would be unsurprising that earlier studies had shown an interaction between semantic and form-based effects in PWI task (Damian & Martin, 1999; Starreveld & La Heij, 1995, 1996b; Taylor & Burke, 2002): both effects would presumably affect lexical-semantic retrieval, making a statistical interaction a likely outcome. By contrast, in the present study due to the nature of the Chinese orthographic system we were able to exclude orthographic overlap altogether and to focus on "pure" phonological similarity. If the latter exclusively reflects phonological encoding, then perhaps information transmission between semantic and phonological level is indeed serial, and previous studies had found statistical non-additivity due to the orthographic confound.

It should be noted, however, that Zhao et al. (2012) more recently mounted a robust argument against the possibility of different loci for phonological and orthographic effects in PWI tasks, based on strong similarity of the two types of effects (plus their time course, as assessed by stimulus-onset manipulations) in their own experiments). Additionally, the "input priming" account would leave unexplained why a statistical interaction was obtained with English speakers even when distractor words were presented in spoken format (Damian & Martin, 1999), a case in which orthographic relatedness would appear to be less relevant. Finally, a strictly serial view of activation transmission is probably incompatible with the recent evidence on "cascadedness" summarized in the Introduction. Nevertheless, we argue that if the aim is to explore the relation between semantic and phonological processing stages in spoken word production via PWI task, it is advisable to avoid a possible orthographic confound, and to design studies which manipulate "pure" phonological overlap. This is difficult to accomplish with target languages using an alphabetic system, but easier with non-alphabetic languages, as in the current study.

From a methodological point of view, our results highlight not only the feasibility of ERP registration during tasks requiring overt naming, but also the advantage of obtaining insight into neural correlates of stages of cognitive processing when compared to behavioral measurement (see Qu, Damian, & Kazanina, 2012). Consideration of only the additive pattern in the behavioral results might have spurred the criticism that perhaps our study was not sensitive enough to detect the statistical interaction previously reported. The fact that semantic and phonological overlap dissociated regarding their time course in the ERPs, and that mixed distractors elicited a semantically based early effect followed by a phonologically based later effect, provides substantial and powerful evidence for a sequential and additive processing mode in Mandarin word production. In all, the combination of overt naming task with ERPs provides researchers with increased degrees of freedom to investigate the mechanisms underlying speech production.

5. Conclusion

The results from this study suggest some degree of cross-linguistic divergence of phonological encoding during spoken word production. Combined with previous findings, our view is that for speakers of Indo-European languages, information transmission between lexical-semantic and phonological levels might be cascaded/interactive, whereas for Mandarin speakers the two stages are accessed in a more discrete manner. Further investigations are needed to pinpoint the timing, coordination, and integration of different information types during spoken production

Acknowledgments

This research was supported by Grants from the National Natural Science Foundation of China (31170977, 31471074), the Fundamental Research funds for the Central Universities, and the Research Funds of Renmin University of China (No. 14NXLF12) to Qingfang Zhang.

References

Abdel Rahman, R., & Sommer, W. (2003). Does phonological encoding in speech production always follow the retrieval of semantic knowledge? Electrophysiological evidence for parallel processing. *Cognitive Brain Research, 16*, 372–382.

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*, 390–412.

Bi, Y. C., Xu, Y. D., & Caramazza, A. (2009). Orthographic and phonological effects in the picture–word interference paradigm: Evidence from a logographic language. *Applied Psycholinguistics, 30*, 637–658.

Blackford, T., Holcomb, P. J., Grainger, J., & Kuperberg, G. R. (2012). A funny thing happened on the way to articulation: N400 attenuation despite behavioral interference in picture naming. *Cognition, 123*, 84–99.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14*, 177–208.

Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language, 46*, 751–781.

Chen, Y., Lee, J., Kuo, W., Hung, D. L., & Cheng, S. (2010). An ERP study of Chinese speakers' rhyme judgments to Chinese and English words. *Neuroreport, 21*, 636–640.

Chinese Linguistic Data Consortium. (2003). 现代汉语通用词表[Chinese lexicon] (CLDC-LAC-2003-001). Beijing, China: Tsinghua University, State Key Laboratory of Intelligent Technology and Systems, and Chinese Academy of Sciences, Institute of Automation.

Costa, A., Strijkers, K., Martin, C., & Thierry, G. (2009). The time course of word retrieval revealed by event-related brain potentials during overt speech. *Proceedings of the National Academy of Sciences, 106*, 21442–21446.

Damian, M. F., & Martin, R. C. (1999). Semantic and phonological codes interact in single word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 345–361.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*, 283–321.

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). *Psychological Review, 98*, 604–614.

Dell'Acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., & Grainger, J. (2010). ERP evidence for ultra-fast semantic processing in the picture–word interference paradigm. *Frontiers in Psychology, 1*, 177.

Dumay, N., & Damian, M. F. (2011). A word-order constraint in single-word production?: Failure to replicate Janssen, Alario, and Caramazza (2008). *Psychological Science, 22*, 559–561.

Ganushchak, L. Y., Christoffels, I. K., & Schiller, N. O. (2011). The use of electroencephalography in language production research: A review. *Frontiers in Psychology, 2*, 208.

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

Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology, 28*, 240–244.

Humphreys, G. W., Riddoch, M. J., & Quinlan, P. T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology, 5*, 67–104.

Hutson, J., & Damian, M. F. (2014). Semantic gradients in picture–word interference tasks: Is the size of interference effects affected by the degree of semantic overlap? *Frontiers in Psychology, 5*, 872.

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. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*, 101–144.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Oxford University Press.

Jescheniak, J. D., Hahne, A., & Schriefers, H. (2003). Information flow in the mental lexicon during speech planning: Evidence from event-related brain potentials. *Cognitive Brain Research, 15*, 261–276.

Jescheniak, J. D., Oppermann, F., Hantsch, A., Wagner, V., Mädebach, A., & Schriefers, H. (2009). Do perceived context pictures automatically activate their phonological code? *Experimental Psychology, 56*, 56–65.

Jescheniak, J. D., Schriefers, H., Garrett, M. F., & Friederici, A. D. (2002). Exploring the activation of semantic and phonological codes during speech planning with event-related brain potentials. *Journal of Cognitive Neuroscience, 14*, 951–964.

Kuipers, J.-R., & Heij, W. L. (2009). The limitations of cascading in the speech production system. *Language and Cognitive Processes, 24*, 120–135.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203–205.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de) Constructing the N400. *Nature Reviews Neuroscience, 9*, 920–933.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition, 42*, 1–22.

Levelt, W. J., Praamstra, P., Meyer, A. S., Helenius, P., & Salmelin, R. (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience, 10*, 553–567.

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

Lupker, S. J. (1982). The role of phonetic and orthographic similarity in picture–word interference. *Canadian Journal of Psychology, 36*, 349–367.

Mädebach, A., Jescheniak, J. D., Oppermann, F., & Schriefers, H. (2011). Ease of processing constrains the activation flow in the conceptual-lexical system during speech planning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 649–660.

Maess, B., Friederici, A. D., Damian, M., Meyer, A. S., & Levelt, W. M. (2002). Semantic category interference in overt picture naming: Sharpening current density localization by PCA. *Journal Of Cognitive Neuroscience, 14*, 455–462.

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture–word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 503–535.

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., & Damian, M. F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory and Cognition, 35*, 494–503.

Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 555–563.

Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: Further evidence for a cascade model of lexical access. *Journal of Memory and Language, 53*, 359–377.

O'Seaghdha, P. G., Chen, J.-Y., & Chen, T.-M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition, 115*, 282–302.

Piai, V., Roelofs, A., & van der Meij, R. (2012). Event-related potentials and oscillatory brain responses associated with semantic and Stroop-like interference effects in overt naming. *Brain Research, 1450*, 87–101.

Posnansky, C. J., & Rayner, K. (1978). Visual vs. phonemic contributions to importance of initial letter in word identification. *Bulletin of the Psychonomic Society, 11*, 188–190.

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

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42*, 107–142.

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

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

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*, 356–374.

Salmelin, R., Schnitzler, A., Schmitz, F., & Freund, H. (2000). Single word reading in developmental stutterers and fluent speakers. *Brain, 123*, 1184–1202.

Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture–word interference studies. *Journal of Memory and Language, 29*, 86–102.

Starreveld, P. A., & La Heij, W. (1995). Semantic interference, orthographic facilitation, and their interaction in naming tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 686–698.

Starreveld, P. A., & La Heij, W. (1996a). The locus of orthographic-phonological facilitation: Reply to Roelofs, Meyer, and Levelt (1996). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 252–255.

Starreveld, P. A., & La Heij, W. (1996b). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 896–918.

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

Strijkers, K., & Costa, A. (2011). Riding the lexical speedway: A critical review on the time course of lexical selection in speech production. *Frontiers in psychology, 2*, 356.

Taylor, J. K., & Burke, D. M. (2002). Asymmetric aging effects on semantic and phonological processes: Naming in the picture–word interference task. *Psychology and Aging, 17*, 662–676.

Tsai, P. S., Yu, B. H. Y., Lee, C. Y., Tzeng, O. J. L., Hung, D. L., & Wu, D. H. (2009). An event-related potential study of the concreteness effect between Chinese nouns and verbs. *Brain Research, 1253*, 149–160.

Underwood, G., & Briggs, P. (1984). The development of word recognition processes. *British Journal of Psychology, 75*, 243–255.

Valdes-Sosa, M., Gonzales, A., Xiang, L., Xiao-Lei, Z., Yi, H., & Bobes, M. A. (1993). Brain potentials in a phonological matching task using Chinese characters. *Neuropsychologia, 31*, 853–864.

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*, 787–806.

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology, 48*, 422–488.

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

West, W. C., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience, 12*, 1024–1037.

Wohlert, A. B. (1993). Event-related brain potentials preceding speech and nonspeech oral movements of varying complexity. *Journal of Speech, Language and Hearing Research, 36*, 897–905.

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

Zhang, Q., Chen, H.-C., Weekes, B. S., & Yang, Y. F. (2009). Independent effects of orthographic and phonological facilitation on spoken word production in Mandarin. *Language Speech, 52*, 113–126.

Zhang, Q., Guo, C. Y., Ding, J. H., & Wang, Z. Y. (2006). Concreteness effects in the processing of Chinese words. *Brain and Language, 96*, 59–68.

Zhang, Q., & Weekes, B. S. (2009). Orthographic facilitation effects on spoken word production: Evidence from Chinese. *Language and Cognitive Processes, 24*, 1082–1096.

Zhang, Q., & Yang, Y. (2003). The determiners of picture naming latency. *Acta Psychologica Sinica, 35*, 447–454 (in Chinese).

Zhang, Q., & Zhu, X. (2011). The temporal and spatial features of segmental and suprasegmental encoding during implicit picture naming: An event-related potential study. *Neuropsychologia, 49*, 3813–3825.

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*, 272–280.