BRAIN RESEARCH 1184 (2007) 234–244

available at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

Electrophysiological estimates of the time course of tonal and orthographic encoding in Chinese speech production

Qingfang Zhang[a,*], *Markus F. Damian*[b], *Yufang Yang*[a]

[a]*State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, China*
[b]*Department of Experimental Psychology, University of Bristol, UK*

ARTICLE INFO

Article history:
Accepted 25 September 2007
Available online 4 October 2007

Keywords:
Speech production
Orthographic encoding
Tonal encoding
Dual-choice Go/noGo task
N200
LRP

ABSTRACT

Recent electrophysiological studies have investigated the time course of semantic, syntactic, and phonological encoding in European language spoken production, such as English or Dutch. The present study investigated the time course of tonal and orthographic encoding during Chinese word production. Participants were shown pictures and carried out a dual-choice Go/noGo decision based on tonal information (whether a picture name was tone 1 or 2, or tone 3 or 4) or orthographic information (whether or not the picture name was written with a left–right structure character). Analyses of N200 effects and LRPs (lateralized readiness potentials) indicated that tonal information was retrieved prior to orthographic information. These results imply that orthographic codes are unlikely to contribute to phonological encoding in spoken word production. Furthermore, a late effect for the N200 in the Go/noGo=tone condition was observed, which may be related to internal self-monitoring of suprasegmental information.

1. Introduction

A central issue in language production concerns the time course of semantic, syntactic, and phonological information. To maintain fluency in speaking, the retrieval of these different types of information has to be orchestrated with millisecond precision. Behavioral data suggest that a word's conceptual/semantic and syntactic properties are retrieved before its phonological form is available (e.g., Dell and O'Seaghdha, 1991, 1992; Levelt et al., 1991; Peterson and Savoy, 1998; Schriefers et al., 1990). Electrophysiological studies on language production in European languages using the lateralized readiness potential (LRP) and the N200 have provided converging evidence supporting this information retrieval sequence (Rodriguez-Fornells et al., 2002; Schmitt et al., 2000, 2001; van Turennout et al., 1997, 1998). In the current study, we investigated spoken production in Chinese speakers and specifically aimed to track the time course of two aspects related to phonological encoding in speaking, namely access to (i) tonal information and (ii) orthographic representations.

Much recent research has been devoted to elucidating the characteristics of phonological encoding in spoken production. A central issue concerns how information about phonological segments and their respective order is combined with suprasegmental codes, such as stress. Based on a multitude of empirical sources of evidence, Levelt et al. (1999) proposed a

---

[*] *Corresponding author.* Institute of Psychology, Chinese Academy of Sciences, Datun Road 10A, Beijing, 100101, China. Fax: +86 10 64872070.
E-mail address: zhangqf@psych.ac.cn (Q. Zhang).

Abbreviations: LRP, lateralized readiness potential; WEAVER, word form encoding by activation and verification; EEG, electroencephalogram; EOG, electro-oculogram

0006-8993/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2007.09.067

fine-grained model of phonological encoding, termed WEAVER (word form encoding by activation and verification). The WEAVER model, in agreement with behavioral (Meyer, 1990, 1991; Wheeldon and Levelt, 1995; Wheeldon and Morgan, 2002) and electrophysiological data (van Turennout et al., 1997), assumes that segmental encoding proceeds in an incremental fashion from the beginning of a word to its end. Furthermore, according to WEAVER, segmental encoding (the retrieval of a word's phonological segments and their respective order) and metrical encoding (access to a word's metrical frame, i.e., at least the number of syllables and the location of lexical stress) occur in parallel. Relatively few studies have been devoted to the investigation of metrical encoding in language production. Only two studies have investigated metrical stress encoding in internally generated speech; both behavioral data (Schiller et al., 2006) and ERP data (Schiller, 2006) indicate that participants are able to carry out a lexical stress decision on object names significantly faster when the picture name is stressed on the initial, than on the final, syllable. These two studies suggest that metrical encoding, much like segmental encoding, is an incremental process.

1.1. *Tonal information in Chinese spoken production*

In stress languages such as English and Dutch, stress position of a word is fixed, and hence stress is typically not lexically distinctive. For example, "cognition" is stressed at the second syllable. No other word exists in English which has the same segments and order, but is stressed at the first or the last syllable. By contrast, in tonal languages such as Mandarin Chinese, it is tone which is lexically distinctive (Chen et al., 2002): a large number of monosyllabic words exist with the same segments but different tones. For example, hu3 (rise falling tone, "tiger") and hu2 (low rising one, "lake") represent two different words with distinct meaning in Chinese. Therefore tone is extremely important for distinguishing word meaning. Both tone and stress are considered suprasegmental types of representations. Chen et al.'s (2002) study suggests that tone in Chinese functions much like stress in English during speech production. Hence, if one wants to express oneself fluently in Chinese, retrieval of tone information is centrally important in speaking. Yet, despite its importance, only very few studies have investigated the role of tonal codes in Chinese spoken production. This makes an investigation into the temporal aspects of access to tonal information in spoken word production imperative. Indefrey and Levelt (2004) estimated a critical time window of between 275- and 400-ms post stimulus onset for access to segmental information; assuming that segmental and suprasegmental properties of words are accessed in parallel, we predict information about tone to be available roughly in the same time window.

1.2. *Orthographic effects in spoken production*

The spoken production of a word involves the retrieval of phonological representations. A few behavioral studies have recently explored the possibility that literate speakers may also mandatorily co-activate a different representational format, namely orthographic information. This claim is somewhat counterintuitive as spelling codes would prima facia appear irrelevant for phonological encoding. It is motivated by the observation that in language comprehension, rather than production, there is substantial and growing evidence for the co-activation of orthographic codes (e.g., Chéreau et al., 2007; Dijkstra et al., 1995; Donnenwerth-Nolan et al., 1981; Hallé et al., 2000; Jakimik et al., 1985; Muneaux and Ziegler, 2004; Racine and Grosjean, 2005; Seidenberg and Tanenhaus, 1979; Taft and Hambly, 1985; Ventura et al., 2004; Ziegler and Ferrand, 1998; Ziegler et al., 2004). These findings suggest a high degree of interconnectedness of linguistic codes in different formats in the mental lexicon, with access to phonological representations entailing parallel access to the orthographic format.

If so, then it is not implausible that a similar process should take place in spoken production as well. At present, empirical evidence that would speak to the issue of co-activation between phonological and orthographic codes in this domain is limited. Gaskell et al. (2003) studied the fact that the definite article "the" is typically pronounced as "thee" when occurring before a noun starting with a vowel, and as "thuh" otherwise. Participants were auditorily presented with nouns and were asked to shadow them together with the determiner. It was shown that the chosen form of the determiner depended not only on the pronunciation and stress of the target word, but crucially also on its spelling. For instance, despite the fact that "union" and "yellow" both begin with the same phoneme, "union" was more likely to be preceded by the article "thee", presumably because the first letter of "union", but not of "yellow", corresponds to a vowel. Wheeldon and Monsell (1992) investigated long-lasting priming in language production, and specifically explored whether production of a word in a study phase (e.g., in response to a definition) can facilitate production of a form-related word in a subsequent test phase (e.g., in a picture naming task). They showed that the production of homographic homophones (bat), but not heterographic homophones (sun-son), yielded facilitatory effects over substantial periods of time. The fact that long-term facilitation of responses primed by homophones depends on the presence or absence of spelling overlap underscores the potential importance of orthographic codes in spoken production. Damian and Bowers (2003) used a form preparation paradigm in which a small number of responses, typically elicited by prompt words, are produced repeatedly within an experimental block, and the presence or absence of form overlap between the responses is manipulated. They replicated previous studies in showing a reliable priming effect in the homogenous condition in which all response words shared initial sound and spelling (e.g., "camel"–"coffee"–" cushion"), compared to a heterogeneous condition in which this was not the case (e.g., "camel"–"gypsy"–"cushion"). Crucially, no priming effect was obtained in an inconsistent condition in which all response words shared initial sound, but differed in spelling (e.g., "camel"–"kayak"–"kidney"). Hence, when retrieving the phonological codes of the response words, although information about spelling is irrelevant to the speaking process, orthographic codes may nevertheless be activated.

On the other hand, some studies have suggested that effects of orthography in spoken production may perhaps not be as automatic as hypothesized. Subsequent studies with the paradigm used by Damian and Bowers (2003), but conducted with Dutch (Roelofs, 2006; see also Schiller, 2007) and French (Alario et al., 2007) speakers, failed to replicate the originally reported effect, warranting caution about the claim that orthographic activation is mandatory in spoken production, and highlighting the need for further studies in order to clarify the role of orthography in speech production (and perception).

The aim of the study below was to assess the time course of access to orthographic information in Chinese speakers. The orthographic system of Chinese consists of a number of levels: strokes, radicals, characters, and words. A character is composed of one or more radicals, which, in turn, are composed of one or more strokes. Modern Chinese characters can be broadly differentiated into two categories (Li, 1993): simple and complex. Most complex characters consist of a semantic radical on the left and a phonetic radical on the right. For example, 柏 (/bai3/, cypress) is composed of a semantic radical (木, /mu1/, wood) on the left, and a phonetic radical (白, /bai2/, white) on the right. Such a Chinese character can be said to have a left–right structure. A radical can appear in different positions within a complex character. For example, 白 is at the top of 皂 (/zao4/, soap), at the bottom of 皆 (/jie1/, all), to the left of 皑 (/ai2/, pure white), and to the right of 柏 (adopted from Ding et al., 2004). Increasing evidence from a range of paradigms suggests that reading a complex character involves the processing of its radicals and their positional relationship (Ding et al., 2004; Taft et al., 2000, 1999), hence a character's structural information is an important element of the mental representation of orthography.

An interesting property of the Chinese orthographic system is that it is very "deep", i.e., the correspondences between spelling and sound are weak. Recently the argument has been raised that automatic orthographic effects in spoken production may be stronger in languages with deep than with shallow orthographies. The argument is based on the finding that effects from English speakers originally reported by Damian and Bowers (2003) did not replicate easily in a language such as Dutch which has a shallower orthographic transparency. Roelofs (2006) pointed out that "cross-linguistic differences in the degree to which orthography and phonology interact in speech production (are) perhaps related to differences in orthographic depth between languages". If true, studies conducted on Chinese speakers should be particularly likely to demonstrate automatic orthographic effects.

In the study reported below, we assess the time course of access to orthographic information about an object name, and contrast it with the time course for tonal representations. Specifically, we will assess electrophysiological markers – outlined below – for each type of code regarding their time of availability. The relative timing of markers for the two types of codes may enable us to distinguish mandatory from strategic, or task-dependent, retrieval modes: a relatively similar time course for tonal and orthographic retrieval may imply that both codes are centrally involved in lexical retrieval, whereas a substantial delay of access to orthographic, relative to tonal, codes would suggest that spelling properties are only accessed in a processing step subsequent to lexical access proper.

1.3. *Electrophysiological markers of the time course of spoken production*

The relative time course of tonal versus orthographic encoding in Chinese word production was investigated by using the N200 and LRP components, which in previous studies have been shown to be informative in constraining theoretical models (e.g., Levelt et al., 1999; Schiller, 2006). A tone decision task was used to examine the stage of suprasegmental encoding during language production, whereas an orthographic decision task was used to investigate access to spelling properties of object names.

The N200 is a negative-going waveform. In a Go/noGo task, participants are asked to respond to one type of stimuli and to withhold response for another type. Compared to the waveform on the Go trials, an ERP component, namely N200, is typically observed on the noGo trials. It is visible at a fronto-central region occurring between 100 and 300 ms after stimulus onset (Jodo and Kayama, 1992; Sasaki et al., 1993; Simson et al., 1977). It has been suggested that the amplitude of the N200 is a function of neural activity required for response inhibition. Hence, the emergence of N200 suggests that the information which is used to determine whether or not a response is to be given must have been encoded. The peak latency of the N200 can therefore be used to determine the moment in time at which this information has become available. Note that the N200 occurs later in time when it is related to language processing (see Kutas and Schmitt, 2003).

As is the case for N200, the LRP (lateralized readiness potential) has been proved a sensitive index for estimating the timing of information processing (e.g., Rodriguez-Fornells et al., 2002; Schmitt et al., 2000; van Turennout et al., 1997). It is derived from the readiness potential, which is a slow, negative-going potential that starts to develop about 0.5 s before the execution of a voluntary hand movement and reaches its maximum just after the response is initiated. A series of studies (i.e., Kutas and Donchin, 1980) have shown that the LRP can be used as an index for exact response preparation prior to an overt Go response, or in the absence of any overt response. The LRP is computed according to the following equation:

$$\mathrm{LRP} = \text{ right hand}(\mathrm{C3}-\mathrm{C4}) - \text{ left hand}(\mathrm{C3}-\mathrm{C4})$$

The electrode sites C3 and C4 are located above the left and the right motor cortices, respectively. The LRP yields largest amplitudes at the motor cortices contralateral to the response hand.

The present study used a dual-choice Go/noGo task adopted from pioneering work by van Turennout et al. (1997, 1998). Participants were asked to perform a dual decision along two dimensions, tone and orthography. The tonal decision was to determine whether the name of the depicted object was of one tone type (tone 1 or tone 2) or the other (tone 3 or tone 4), while the orthographic decision was to decide whether or not the picture's written name consisted of a left–right structure character.

In the experiment, the instruction was, for example, "press the left button if the picture's name is a left–right structure character (i.e., 狗, /gou3/, *dog*), and the right button if the picture's name is not a left–right structure character (i.e., 缶, /

yu2/, *fish*). However, respond only if the picture's name has tone 1 or tone 2, and not if the picture's name has tone 3 or tone 4." Thus, depending on the tonal and orthographic characteristics, a response was given with either the left or right hand, or no response at all.

The experimental design consisted of two conditions. In one half of the experiment, the responding hand was contingent on tonal information (hand = tone) and the Go/noGo decision on orthographic information (Go/noGo = orthography). In the other condition, the response contingencies were reversed, i.e., the responding hand was contingent on orthographic information (hand = orthography) and the Go/noGo decision on tonal information (Go/noGo = tone). The logic of varying response contingencies, again adopted from van Turennout et al. (1997) and subsequent studies, is as follows: if information associated with the response hand dimension (e.g., tone) is available before information associated with the Go/noGo dimension (e.g., orthography), then on noGo trials, the negative deflection typical of the LRP should be observed, which reflects a – eventually abandoned – response preparation. Crucially, given the assumptions about relative timing of the two dimensions, an LRP should not be observed on noGo trials when the response dimensions are reversed (i.e., hand = orthography; Go/noGo = tone). An LRP should be observed on Go trials in both conditions.

Although the dual-choice Go/noGo task can be criticized on the basis that is does not constitute a "pure" instance of language production, it has been productively used to examine various aspects of language production in a number of previous studies (Rodriguez-Fornells et al., 2002; Schmitt et al., 2000, 2001; van Turennout et al., 1997). In our procedure participants were trained to name pictures in the practice session before the main session, and the results indicated that they used the intended picture names. It is worth noting that the assumption that overt and covert spoken production involves very similar cognitive and cortical processes receives somewhat mixed support in the literature. For instance, Palmer et al. (2001) demonstrated in an fMRI study that regions active during overt speaking were similar to those active during covert task performance, with the addition of certain regions typically associated with motor aspects of spoken production. On the other hand, Christoffels et al. (2007) compared overt and covert picture naming in Dutch and found – among other differences – different activation patterns in the insula during overt and covert naming.

2. Results

2.1. *Pre-experiment*

The aim of this pre-test was to measure mean reaction times on the target stimuli for simple tonal and orthographic decision tasks, respectively (i.e., without the dual-task paradigm). The results were taken to indicate how long it takes in general to perform a simple tone decision (i.e., if the tone of a picture name is tone 1 or tone 2, press the left button, if tone 3 or tone 4, press the right button) or orthographic decision task (if a picture name is a left–right character, press the left button, if not, press the right button). Thirteen native Chinese speakers from Beijing Forestry University participated in the pretest. Errors and reaction times deviating more than 3 standard deviations (SD) from the mean were excluded from the data analysis. For the orthographic decision, the mean reaction time was 1314 ms (SD = 431 ms), with a mean error rate of 5.23%. For the tone decision, the mean reaction time was 1318 ms (SD = 395 ms), with a mean error rate of 7.36%. A paired *t*-test showed that the two conditions did not differ significantly from each other on reaction times or error rates (both $ts < 1$). The very similar time for participants to accomplish these two simple decision tasks suggests comparable degrees of difficulty.

2.2. *Push-button reaction times*

One participant was excluded from all further analyses due to high error rates (>35%). Incorrect responses and reaction times longer than 2000 ms were excluded from data analysis. These criteria accounted for 23.3% of the data. The number of rejections was not significantly different for the two response contingency conditions. The mean reaction times for correct Go responses were averaged across left and right responses for the remaining 15 participants. The mean reaction time was 1369 ms (SD = 101 ms) and 1406 ms (SD = 142 ms) for the Go/noGo = tone (hand = orthography) and Go/noGo = orthography (hand = tone) condition, respectively. A paired sample *t*-test of the reaction times for the two conditions showed no significant difference ($t(14) = 1.60$, n.s.). Likewise, the error rates were not significantly different between the two response conditions: Go = tone: 13.3%, Go = orthography: 15.8% ($t(14) = 1.43$, n.s.); NoGo = tone: 0.6%, NoGo = orthography: 0.4% ($t(14) = 2.03$, $p < 0.06$).

2.3. *N200 analysis*

Three participants were excluded from the analysis due to no significant amplitude difference between Go and noGo average waves generated in both conditions. The electrophysiological signals were averaged separately for the Go and noGo trials. The N200 effect was obtained by subtracting waveforms on noGo trials from those on Go trials in the different Go/noGo contingency conditions (tone versus orthography). The maximum number of trials per condition per individual was 144. On average, the number of accepted trials for both conditions was 111. The minimum number of accepted trials was 86 in the go = tone (hand = orthography) and 94 in the go = orthography (hand = tone) condition. Fig. 1 shows grand average ERP waveforms on the Go and noGo trials in the Go/noGo = tone and orthography condition for the 12 participants at midline sites (Fz and FCz). Both response contingency conditions showed clear evidence of N200 effects (see arrows in Fig. 1). Waveforms on noGo trials were more negative than on Go trials in both conditions, and this tendency occurred visibly earlier in the Go/noGo = tone condition than in the Go/noGo = orthography condition.

Fig. 2 shows the grand average of difference waves (noGo minus go) for the Go/noGo = tone and orthography conditions at Fz and FCz sites. It is clear that the difference wave in the Go/noGo = tone condition is bimodal, with a first peak at around 300 ms and a second peak at around 600 ms, whereas

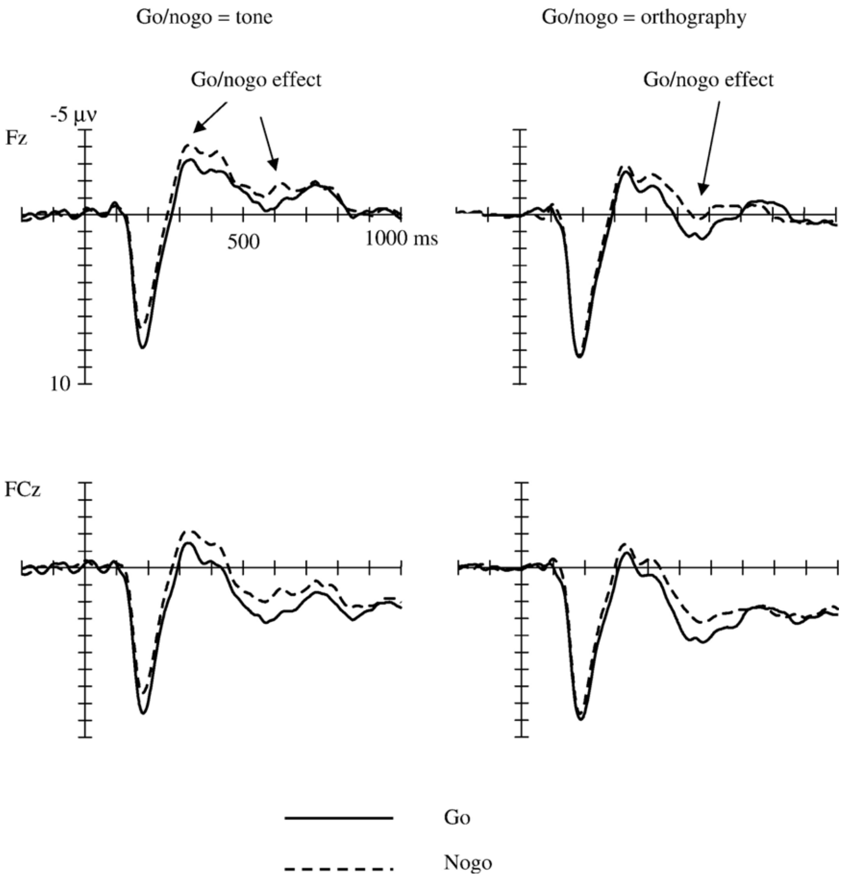


Fig. 1 – Grand average ERP waveforms on the Go and noGo trials in the Go/noGo=orthography or tone condition for 12 participants at Fz and FCz sites.

the difference wave in the Go/noGo=orthography condition exhibits only a single peak, around 500 ms. To estimate the onset latency of the difference wave and determine the time window of data analysis, serial *t*-tests at Fz and FCz sites in the time window 200–800 ms after picture onset were performed in both conditions. The time intervals after picture onset in which the difference wave significantly ($p < 0.05$, one-tailed) diverges from the zero baseline are shown in Table 1.

Based on the above *t*-tests analysis, peak latencies and peak amplitude of the difference waves were analyzed for each participant in the time window in the Go/noGo=tone condition (200–450 ms after picture onset), as well as in the time window in the Go/noGo=orthography condition (400–

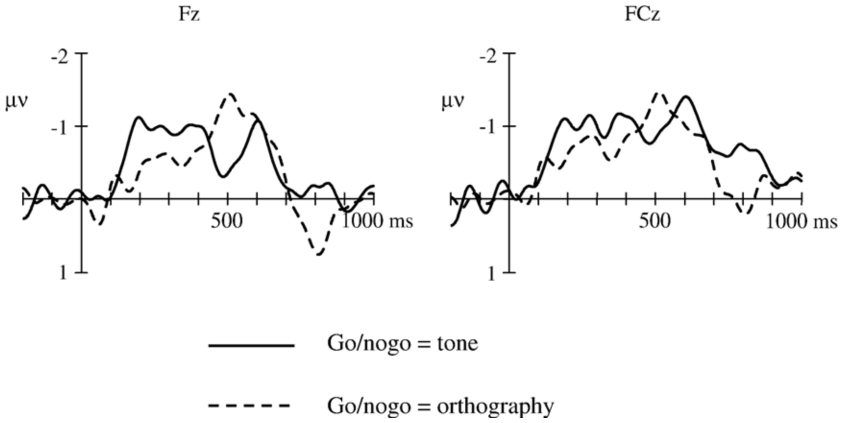


Fig. 2 – Grand average difference waves (noGo/Go) for the Go/noGo=orthographic or tone conditions at Fz and FCz sites.

Table 1 – Time intervals of significant divergence ($p<0.05$) of the difference wave from zero baseline in the time window 200–800 ms after picture onset

Contingency conditions	Different time intervals (ms) of significant t tests in the time window of 200–800 ms after picture onset at Fz and FCz sites	
	Fz	FCz
Tone	236–438 $-1.87<t(11)<-2.04$ 598–610 $-1.80<t(11)<-1.81$	234–440 $-1.84<t(11)<-1.82$ 584–616 $-1.85<t(11)<-1.84$
Orthography	450–600 $-1.85<t(11)<-1.83$	470–540 $-1.82<t(11)<-1.81$

700 ms). Omnibus ANOVAs were computed on the N200 peak latencies and peak amplitudes of the first early component with two within-participant variables: contingency condition (Go/noGo=tone versus orthography) and electrode site (Fz and FCz). Greenhouse–Geisser correction was used when appropriate.

2.3.1. *Peak latency of the first early component*

The main effect of contingency condition was significant ($F(1,11)=53.580$, $p<0.001$), reflecting a difference in peak latencies of the N200. When the Go/noGo decision was contingent on orthographic information, the mean peak latency of the N200 effect was 571 ms (SD=89 ms), whereas it was 301 ms (SD=88 ms) when contingent on tonal information. Neither the main effect of electrode site ($F<1$), nor the interaction between contingency condition and electrode site ($F(1,11)=1.284$, n.s.), was significant. Fig. 3 shows the scalp distribution of the N200 effects for the Go/noGo=tone condition (mean amplitude of the time window 280–330 ms after picture onset) and for the Go/noGo=orthography condition (mean amplitude of the time window 550–600 ms), respectively. It is observed that the first early components of the difference waves have a similar scalp distribution in both conditions, i.e., it is generated by the same neural populations. Although the peak amplitude of the first early negative component in the Go/noGo=orthography is larger than in the Go/noGo=tone condition, both are distributed largely in frontal and central regions.

2.3.2. *Peak amplitude of the first early component*

Neither the main effect of contingency condition ($F<1$), the main effect of electrode site ($F<1$), nor the interaction between condition and electrode site ($F<1$) were significant. The mean amplitude difference (across both electrode sites) of the two N200 effects was very small (0.15 $\mu\nu$).

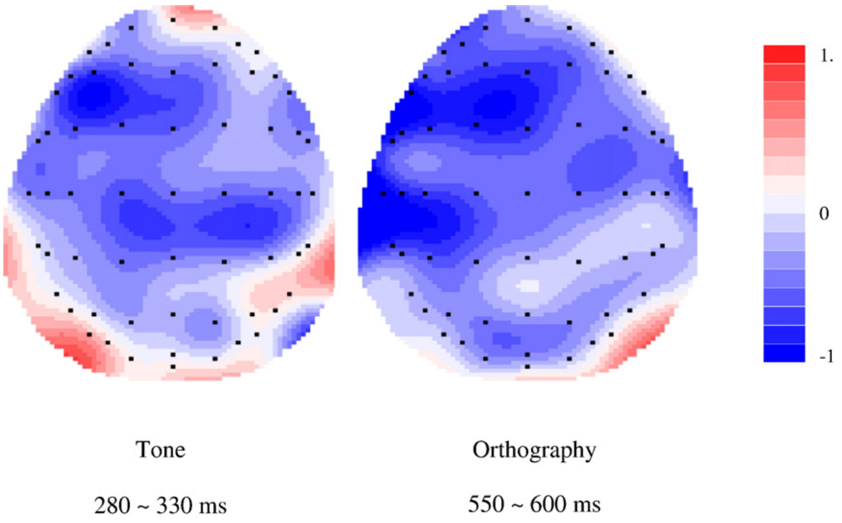
2.3.3. *Onset latency of the first early component*

Following Schmitt et al. (2000), the onset latency of the difference wave was defined as the point at which four consecutive t-tests yielded significant results (in the same direction). Based on the above t-tests analysis, the mean onset latencies of the first component at the two electrode sites were 235 ms and 460 ms in the Go/noGo=tone and orthography, respectively.

2.4. *LRP analysis*

Four stimulus-locked LRP average waveforms for each participant were calculated: (1) hand=tone, Go=orthography, (2) hand=tone, noGo=orthography, (3) hand=orthography, Go=tone, (4) hand=orthography, noGo=tone. The maximum number of trials per condition per individual was 72.

In order to assure that LRP results refer to the same sample as the N200 results, we restricted analysis to those 12 participants used in the N200 analysis. On average, the number of accepted trials for Go and noGo LRPs in the hand=tone condition was 51 and 63, respectively, and the



Tone

280 ~ 330 ms

Orthography

550 ~ 600 ms

Fig. 3 – Scalp distribution of the N200 effects for the Go/noGo=tone condition (mean amplitudes of the time window 280–330 ms) and the Go/noGo=orthography condition (mean amplitudes of the time window 550–600 ms).

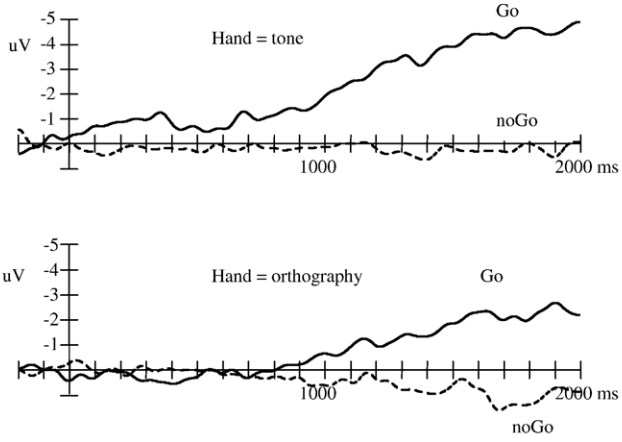


Fig. 4 – Grand average of Go and noGo LRP in the hand = tone (top panel) and the hand = orthography (bottom panel) condition.

hand = orthography condition was 53 and 66, respectively. There was no significant difference on the number of accepted trials between Go and noGo LRPs in both conditions. The minimum number of accepted trials for Go and noGo LRPs in the hand = tone condition was 43 and 55, respectively, and the hand = orthography was 43 and 50, respectively. Fig. 4 shows the grand average of Go and noGo LRPs for the response contingency conditions: hand = tone (top panel), and hand = orthography (bottom panel) at the motor cortex sites (C3 and C4). The typical LRP patterns for Go responses were obtained under both the hand = tone or hand = orthography conditions.

The LRPs were measured by mean amplitudes relative to the pre-stimulus baseline (−200 to 0 ms before picture onset). The onset latency of the LRP was measured by one-tailed serial *t*-tests between 300 and 1200 ms after picture onset against a zero mean. The *t*-tests were carried out stepwise with a step size of 2 ms. As in the analysis on the N200, the onset latency of the LRP was defined as the point at which four consecutive *t*-tests yielded significant results (in the same direction).

The mean onset latency for the Go LRP in the hand = tone condition was 766 ms (from that time all $t\ (11) < -1.850$ all $p < 0.05$, one-tailed), while Go LRP in the hand = orthography condition was 1314 ms (from that time all $t\ (11) < -1.825$, all $p < 0.05$, one-tailed). When compared via a paired sample *t*-test, the two Go LRP onset latencies were significantly different. No significant difference from baseline was obtained for the noGo LRP in both conditions.

3. Discussion

The present experiment investigated the time course of tonal and orthographic encoding in spoken Chinese word production with high temporal resolution electrophysiological methods. N200 and LRP data were used to examine the absolute and relative time course of access to tone and orthography during implicit picture naming. The N200 analysis is based on the assumption that increased negativity on the noGo trials in comparison with the Go trials reflects the moment in time at which the relevant information necessary to withhold a response must have been encoded, while the LRP reflects the moment in time at which the relevant information is available for response preparation. The N200 and LRP data showed that tonal information was encoded substantially prior to orthographic information in Chinese spoken word production.

Two different aspects of the N200 effect were examined: (1) peak latency and (2) onset latency. The peak latency of the N200 effect provides an upper limit on the time by which information about whether an actual response needs to be made or withheld must have become available. In terms of temporal course of information processing, the onset of the effect might be as relevant as its peak (see Schmitt et al., 2000). The mean peak latency of the N200 effect was 270 ms earlier when the Go/noGo decisions were contingent on tonal, compared to when they were contingent on orthographic information; the mean difference in the onset latencies was 225 ms. Both peak latency and onset latency data therefore suggest considerable differences between the Go/noGo = tone and orthography conditions, suggesting that tonal information was available prior to orthographic information.

In the Go/noGo = tone condition, ERP waveforms of Go and noGo responses diverged between 235 and 440 ms after picture onset. This interval falls nicely within the time window of phonological encoding (275–400 ms) estimated by Indefrey and Levelt (2004). We therefore suggest that the peak latency (310 ms) of this first component reflects the moment in time at which tonal information must have been encoded to withhold a response. Note that it is likely that there is a delay of some time between the retrieval and availability of the information and the generation of the N200 peak. In contrast, in the Go/noGo = orthography condition, the ERP waveforms of Go and noGo responses diverged only between 460 and 600 ms after picture onset. This interval is very likely beyond the time window of phonological encoding in picture naming. Assuming that the peak latency of this component in the Go/noGo = orthography condition reflects the moment in time at which orthographic information was available, spelling properties are activated with substantial delay relative to tonal representations and are likely accessed too late to contribute to lexical access in spoken word production. This conclusion casts doubt on earlier claims (e.g., Damian and Bowers, 2003; Gaskell et al., 2003) that for literate speakers, access to orthographic representations may constitute a mandatory component of phonological encoding — if true, a relatively similar time course for the two types of information would have been predicted. Indeed, we believe that our data should be interpreted as positive evidence against the claim that orthographic codes impact on phonological encoding. Speakers, when required to gain access to orthographic codes in order to carry out an orthographically based decision task, do so apparently only in a processing step subsequent to phonological encoding. This in turn suggests that orthographic access is either optional and dependent on task requirements or so late that it is unlikely to be relevant for lexical access in spoken production.

The serial *t*-tests reported in Table 1 indicate that in the Go/noGo = tone condition, there is a second, and much later, effect between 580 and 620 ms after picture onset. Similar findings have been previously reported by Schiller et al. (2003). They investigated the time course of segmental and metrical encoding in Dutch speech production and found a late effect between 610 and 640 ms after picture onset in the Go/noGo = metrical condition. Because an effect occurring at such a late point past target onset is likely beyond the time window of phonological encoding, it was suggested that internal self-monitoring of to-be-generated speech was the underlying cause. In our data, the late effect is similarly unlikely to reflect phonological encoding proper, and hence it is possible that before participants decide whether to carry out or to withhold a response in the Go/noGo = tone condition, they undergo a process of first generating the picture name, and then subsequently monitoring it for tonal information. On the other hand, in the Go/noGo = orthography condition, no late effect was observed. If the self-monitoring hypothesis used to explain the late effect in the tone condition is correct, it would follow that no such monitoring is present in the access to orthographic codes, which could be taken as additional evidence that access to orthography is not an essential part of phonological encoding in spoken production.

An alternative explanation for the late effect in the tone condition hinges on the fact that the categories for the tone decision each comprised two tones (see Section 1.3). As a consequence, the task likely required participants to first retrieve individual tones, and then consciously to classify them into the two response categories. For this reason, the first peak in the wave forms may be related to phonological encoding proper, whereas the second peak may reflect a decisional categorization process.[1] Both accounts of the late effect in the tone condition are at present speculative, and further empirical evidence is required to resolve the issue.

The onset latency of the LRPs was also examined. The results indicated that the onset latencies of Go LRPs were 766 and 1314 ms in the hand = tone and orthography condition, respectively, which also reflects a sequence of availability of tonal and orthographic information. Note that the onset of the Go LRP was later than the peak latency of the N200 in both conditions, which may relate to the fact that they are different ERP components and represent different respective cognitive processes. In the literature, it is hypothesized that the N200 reflects response inhibition, whereas the LRP reflects response preparation in a dual-choice Go/noGo task. Fuster (1997) argued that retrieved information is evaluated by the frontal cortex, reflected by N200 effects. The evaluated information is then transmitted to the motor cortex, reflected by the LRPs. According to Fuster's account, the peak latency of the N200 effect should appear earlier than the onset latency of LRP in the same contingency condition. This is indeed what we observed in the present study. However, the exact relationship between N200 and LRP remains tentative and in need of further investigation (Schmitt et al., 2000).

Somewhat disappointingly, we failed to obtain evidence for LRPs on noGo trials. To reiterate, adopting the logic pioneered by van Turennout et al. (1997) and assuming a relative difference in the availability of tonal and orthographic information, a noGo LRP would have been predicted in the hand = tone/response = orthography condition, but not vice versa. We can at present only speculate on why we did not find this pattern. Note that one previous study that adopted this approach (Rodriguez-Fornells et al., 2002) also failed to obtain interpretable noGo LRPs. There is one possible reason for the absence of LRPs on noGo trials: the respective N200 effects have onsets/peaks that precede the onsets of the go-LRPs by quite some time; this early inhibitory information – if used – might have prevented motor preparation on noGo trials. Despite the absence of the predicted effect in our data, we believe that the N200 results speak rather clearly to the central research issue, namely the time course of orthographic and tonal information.

The results from the pretest showed that there was no significant difference with respect to reaction times and error rates on the two simple decision tasks, which suggests comparable task difficulty. In addition, it has been suggested that the magnitude of the N200 is a function of the neural activity required for response inhibition (Jodo and Kayama, 1992; Sasaki and Gemba, 1993), and that it is sensitive to task difficulty not only in monkeys (Gemba and Sasaki, 1989) but also in priming task in humans (Kopp et al., 1996). The insignificant difference between the two peak amplitudes of the two N200 effects in the present study also implies comparable task difficulty of orthographic and tonal decisions. Thus, the results of the N200 and the LRP should not be attributed to task difficulty, but rather reflect the order of information availability during language production.

One possibility that deserves further investigation in future research is that the pattern of orthographic activation following the availability of phonological codes in word production may to some extent be specific to the target language Chinese. It could be argued that the ideographic nature of Chinese characters favors a holistic retrieval mode, whereas in languages with alphabetic scripts, it is not impossible that partial orthographic codes can be activated based on partial availability of phonological codes, simply because both codes map onto each other sublexically. Future experiments should therefore check whether the pattern shown in our data replicates in languages with alphabetic scripts.

Studies that adopt the dual-choice Go/noGo methodology introduced by van Turennout and colleagues use meta-linguistic tasks to assess the properties of spoken production: rather than naming objects, participants perform tasks which tax properties of the codes retrieved in response to pictorial stimuli. It is worth pointing out that previously published work using this approach (see Introduction) investigated access to information (e.g., semantic, syntactic, phonological) which is mandatory in speaking. By contrast, retrieval of the codes necessary to carry out the orthographic task in our study is not strictly speaking necessary for object naming per se; for this reason it could be argued that the observed differences in N200 peak latencies between tonal and orthographic access may be rather large. On the other hand, our study was designed to evaluate the possibility that literacy makes access to orthographic codes a central (yet of course not mandatory) component of lexical access in speaking. If so, then orthography and phonological form should be co-activated and

---

[1] We would like to thank an anonymous reviewer for pointing out this possibility.

modulate each other, resulting in similar time courses of respective access. This does not seem to be the case, which suggests that orthographic access is not a component of phonological encoding in spoken production.

In conclusion, N200 and LRP components were used to investigate the time course of orthographic and tonal encoding in Chinese word production with a dual-choice Go/noGo task. The results of the N200 effect and LRP indicate that tonal encoding occurred earlier than orthographic encoding. Such a pattern appears incompatible with the notion that orthographic access constitutes a mandatory component of phonological encoding in spoken production. Furthermore, there was a late effect on difference wave for Go and noGo trials in the Go/noGo = tone condition, which we suggest may be related to internal self-monitoring of phonological information.

4. Experimental procedures

4.1. *Participants*

Sixteen native Mandarin speakers participated in the experiment (7 females and 9 males, with a mean age of 19.4 years; range 18–21 years). All participants were right-handed, neurologically healthy, with normal or corrected-to-normal vision and normal hearing. They were paid for their participation.

4.2. *Materials*

Eighty target pictures with names corresponding to monosyllabic Chinese characters were selected from a database of standardized pictures for picture naming tasks in Mandarin production (Zhang and Yang, 2003). The stimuli included two orthographic categories: half of the picture names were written with left–right structure characters, and half of them were not. Each of the two groups of pictures contained 20 pictures with tone 1 or tone 2, whereas the other 20 pictures had tone 3 or tone 4. Eight pictures were used as practice stimuli, the remaining 72 pictures as experimental stimuli (see Appendix A).

4.3. *Design*

Each participant received eight different instruction sets. In four experimental sets, the responding hand was contingent on tonal information, and the Go/noGo decision was contingent on orthographic information. In the other four sets, the responding hand was contingent on orthographic information and the Go/noGo decision was contingent on tonal information. The left- and right-hand assignment, as well as the Go and noGo responses, was counterbalanced for each picture. Each picture was presented eight times to each participant, i.e., once in each set. The presentation order of the eight sets was systematically varied across participants.

4.4. *Procedure*

Participants were tested individually in front of a computer screen in a soundproof chamber. At first, participants were asked to familiarize themselves with the pictures and to memorize their corresponding name. Each picture was presented in the center of a computer screen for 3 s, and the picture name was shown below. Then each picture was presented without the printed name, and participants were asked to name the picture as correctly and fast as possible. If the participant's response was other than the expected one, the experimenter would correct the participant, and the testing for this item would be repeated. This procedure, typical of studies that use object naming tasks to investigate spoken production, guaranteed that each participant knew and used the intended names of the pictures.

During the subsequent experimental session, participants were asked to carry out a dual-choice Go/noGo task without overtly naming the picture. Each set began with 16 practice trials (2 pictures in each orthographic–tonal combination category; each picture was presented two times in the practice session), followed by 72 experimental trials (18 pictures in each orthographic–tonal category). Each participant completed 8 sets in the experimental session, with each set consisting of 72 trials.

Each trial was constructed as follows: a fixation cross appeared in the center of the computer screen for 500 ms. After a random interval of 600 to 1600 ms, the picture was presented for 2000 ms. Then a blank screen appeared for 500 ms, followed by the next trial.

4.5. *Apparatus and recordings*

The electroencephalogram (EEG) was recorded with 64 electrodes secured in an elastic cap (Electro cap International). The vertical electro-oculogram (VEOG) was monitored with an electrode placed above and below the left eye. The horizontal EOG (HEOG) was recorded by a bipolar montage using two electrodes placed on the right and left external cantus. The bilateral mastoids served as reference points and the GND electrode on the cap served as ground. Electrode impedances were kept below 5 kΩ for the EEG and eye movement recording.

The electrophysiological signals were amplified with a bandpass from 0.05 to 70 Hz and digitized at a rate of 500 Hz. Epochs of 2200 ms were obtained (−200 ms to 2000 ms) including a 200 ms pre-stimulus baseline. The EEG and EOG signals were filtered with a high-frequency cutoff point of 30 Hz. The artifact rejection criteria were from −100 μv to 100 μv. Push-button response latencies were measured from picture onset, with a time-out point set at 2500 ms, i.e., responses given after 2500 ms were registered as missing. Trials with time-outs and errors were excluded from the data analysis. Hand responses were made by pressing either the left button with the left hand, or the right button with the right hand, on a hand-held button box.

Acknowledgments

This research was supported by Grants from the National Natural Science Foundation of China (30400134), Young Scientist Foundation of Institute of Psychology (07CX102010). We thank two anonymous reviewers for critical comments and helpful suggestions on an earlier draft.

Appendix A. Picture names used in experiment

Left–right structure picture names with tone 1 or tone 2 names	Left–right structure picture names with tone 3 or tone 4 names	Non left–right structure picture names with tone 1 or tone 2 names	Non left–right structure picture names with tone 3 or tone 4 names
钟 (/zhong1/, bell)	蚁 (/yi3/, ant)	书 (/shu1/, book)	斧 (/fu3/, axe)
猫 (/mao1/, cat)	碗 (/wan3/, bowl)	弓 (/gong1/, bow)	鸟 (/niao3/, bird)
杯 (/bei1/, cup)	椅 (/yi3/, chair)	鹰 (/ying1/, eagle)	耳 (/er3/, ear)
鸭 (/ya1/, duck)	狗 (/gou3/, dog)	花 (/hua1/, flower)	手 (/shou3/, hand)
狮 (/shi1/, lion)	鼓 (/gu3/, drum)	刀 (/dao1/, knife)	马 (/ma3/, horse)
枪 (/qiang1/, machgun)	眼 (/yan3/, eye)	山 (/shan1/, mountain)	鼠 (/shu3/, mouse)
针 (/zhen1/, needle)	脚 (/jiao3/, feet)	桌 (/zhuo1/, table)	尺 (/chi3/, ruler)
猪 (/zhu1/, pig)	腿 (/tui3/, leg)	龟 (/gui1/, turtle)	虎 (/hu3/, tiger)
鸡 (/ji1/, chicken)	锁 (/suo3/, lock)	窗 (/chuang1/, window)	伞 (/san3/, umbrella)
球 (/qiu2/, ball)	炮 (/pao4/, cannon)	床 (/chuang2/, bed)	井 (/jing3/, well)
糖 (/tang2/, candy)	帽 (/mao4/, cap)	云 (/yun2/, cloud)	蟹 (/xie4/, crab)
旗 (/qi2/, flag)	豹 (/bao4 /, leopard)	牛 (niu2/, cow)	鹿 (/lu4/, deer)
锤 (/chui2/, hammer)	镜 (/jing4/, mirror)	门 (/men2/, door)	象 (/xiang4/, elephant)
猴 (/hou2/, monkey)	锯 (ju4/, saw)	鱼 (/yu2/, fish)	叉 (/cha4/, fork)
钳 (/qian2/, pliers)	袜 (/wa4 /, socks)	羊 (/yang2 /, goat)	兔 (/tu4/, rabbit)
耙 (/pa2/, rake)	剑 (/jian4/, sword)	鼻 (/bi2/, nose)	凳 (/deng4/, stool)
鞋 (/xie2/, shoe)	树 (/shu4/, tree)	梨 (/li2/, pear)	燕 (/yan4/, swallow)
蛇 (/ she2/, snake)	哨 (/shao4/, whistle)	勺 (/shao2/, spoon)	号 (/hao4/, trumpet)

REFERENCES

Alario, F.-X., Perre, L., Castel, C., Ziegler, J.C., 2007. The role of orthography in speech production revisited. Cognition 102, 464–475.

Chen, T.Y., Chen, T.M., Dell, G.S., 2002. Word-form encoding in Mandarin as assessed by the implicit priming task. J. Mem. Lang. 46, 751–781.

Chéreau, C., Gaskell, M.G., Dumay, N., 2007. Reading spoken words: orthographic effects in auditory priming. Cognition 102, 341–360.

Christoffels, I.K., Formisano, E., Schiller, N.O., 2007. Neural correlates of verbal feedback processing: an fMRI study employing overt speech. Hum. Brain Mapp. 28, 868–879.

Damian, M.F., Bowers, J.S., 2003. Effects of orthography on speech production in a form-preparation paradigm. J. Mem. Lang. 49, 119–132.

Dell, G.S., O'Seaghdha, P.G., 1991. Mediated and convergent lexical priming in language production: a comment on Levelt et al. (1991). Psychol. Rev. 98, 604–614.

Dell, G.S., O'Seaghdha, P.G., 1992. Stages of lexical access in language production. Cognition 42, 287–314.

Dijkstra, T., Roelofs, A., Fieuws, S., 1995. Orthographic effects on phoneme monitoring. Can. J. Exp. Psychol. 49, 264–271.

Ding, G., Peng, D., Taft, M., 2004. The nature of the mental representation of radicals in Chinese: a priming study. J. Exper. Psychol., Learn., Mem., Cogn. 30, 530–539.

Donnenwerth-Nolan, S., Tanenhaus, M.K., Seidenberg, M.S., 1981. Multiple code activation in word recognition: evidence from rhyme monitoring. J. Exper. Psychol., Learn., Mem., Cogn. 7, 170–180.

Fuster, J.M., 1997. The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe. Raven Press, New York.

Gaskell, M.G., Cox, H., Foley, K., Grieve, H., O'Brien, R., 2003. Constraints on definite article alternation in speech production: to "thee" or not to "thee"? Mem. Cogn. 31, 715–727.

Gemba, H., Sasaki, K., 1989. Potential related to no-go reaction of go/no-go hand movement task with color discrimination in human. Neurosci. Lett. 101, 262–268.

Hallé, P.A., Chéreau, C., Segui, J., 2000. Where is the /b/ in "absurde" [apsyrd]? It is in French listeners' minds. J. Mem. Lang. 43, 618–639.

Indefrey, P., Levelt, W.J.M., 2004. The spatial and temporal signatures of word production components. Cognition 92, 101–144.

Jakimik, J., Cole, R.A., Rudnicky, A.I., 1985. Sound and spelling in spoken word recognition. J. Mem. Lang. 24, 165–178.

Jodo, E., Kayama, Y., 1992. Relation of a negative ERP component to response inhibition in a Go/noGo task. Electroencephalogr. Clin. Neurophysiol. 82, 477–482.

Kopp, B., Mattler, R., Goetry, R., Rist, F., 1996. N2, P3 and the lateralized readiness potential in a noGo task involving selective response priming. Electroencephalogr. Clin. Neurophysiol. 99, 19–27.

Kutas, M., Donchin, E., 1980. Preparation to respond as manifested by movement-related brain potentials. Brain Res. 202, 95–115.

Kutas, R., Schmitt, B.M., 2003. Language in microvolts. In: Banich, M.T., Mack, M.A. (Eds.), Mind, Brain, and Language: Multidisciplinary Perspectives. Lawrence Erlbaum Associates Incorporated, New York, NY, pp. 171–209.

Levelt, W.J, Schriefers, M., Vorberg, H., Meyer, D., Pechmann, A.S., Havinga, T., 1991. The time course of lexical access in speech production: a study of picture naming. Psychol. Rev. 98, 122–142.

Levelt, W.J.M., Roelofs, A., Meyer, A.S., 1999. A theory of lexical access in speech production. Behav. Brain Sci. 22, 1–75.

Li, D., 1993. A Study of Chinese Characters. Peking University Press, Beijing.

Meyer, A.S., 1990. The time course of phonological encoding in language production: the encoding of successive syllables of a word. J. Mem. Lang. 29, 524–545.

Meyer, A.S., 1991. The time course of phonological encoding in language production: phonological encoding inside a syllable. J. Mem. Lang. 69–89.

Muneaux, M., Ziegler, J.C., 2004. Locus of orthographic effects in spoken word recognition: novel insights from the neighbor generation task. Lang. Cogn. Processes 19, 641–660.

Palmer, E.D., Rosen, H.J., Ojemann, J.G., Buckner, R.L., Kelley, W.M., Petersen, S.E., 2001. An event-related fMRI study of overt and covert word stem completion. NeuroImage 14, 182–193.

Peterson, R.R., Savoy, P., 1998. Lexical selection and phonological encoding during language production: evidence for cascaded processing. J. Exper. Psychol., Learn., Mem., Cogn. 24, 539–557.

Racine, I., Grosjean, F., 2005. The cost of deleting a schwa at the time of recognition of French words. Can. J. Exp. Psychol. 59, 240–254.

Rodriguez-Fornells, A., Schmitt, B.M., Kutas, M., Münte, T.F., 2002. Electrophysiological estimates of the time course of semantic and phonological encoding during listening and naming. Neuropsychologia 40, 778–787.

Roelofs, A., 2006. The influence of spelling on phonological encoding in word reading, object naming, and word generation. Psychon. Bull. Rev. 13, 33–37.

Sasaki, K., Gemba, H., 1993. Prefrontal cortex in the organization and control of voluntary movement. In: Ono, T., Squire, L.R., Raiche, M.E., Perrett, D.I., Fukuda, M. (Eds.), Brain Mechanisms of Perception and Memory: From Neuron to Behavior. Oxford University Press, New York, pp. 473–496.

Sasaki, K., Gemba, H., Nambu, A., Matsuzaki, R., 1993. No-go activity in the frontal association cortex of human subjects. Neurosci. Res. 18, 249–252.

Schiller, N.O., 2006. Lexical stress encoding in single word production estimated by event-related brain potentials. Brain Res. 1112, 201–212.

Schiller, N.O., 2007. Phonology and orthography in reading aloud. Psychon. Bull. Rev. 14, 460–465.

Schiller, N.O., Bles, M., Jansma, B.M., 2003. Tracking the time course of phonological encoding in speech production: an event-related brain potential study. Cogn. Brain Res. 17, 819–831.

Schiller, N.O., Jansma, B.M., Peters, J., Levelt, W.J.M., 2006. Monitoring metrical stress in polysyllabic words. Lang. Cogn. Processes 21, 112–140.

Schmitt, B.M., Münte, T.F., Kutas, M., 2000. Electrophysiological estimates of the time course of semantic and phonological encoding during implicit naming. Psychophysiology 37, 473–484.

Schmitt, B.M., Schiltz, K., Zaake, W., Kutas, M., Münte, T.F., 2001. An electrophysiological analysis of the time course of conceptual and syntactic encoding during tacit picture naming. J. Cogn. Neurosci. 13, 510–522.

Schriefers, H., Meyer, A.S., Levelt, W.J.M., 1990. Exploring the time course of lexical access in language production: picture–word interference studies. J. Mem. Lang. 29, 86–102.

Seidenberg, M.S., Tanenhaus, M.K., 1979. Orthographic effects on rhyme monitoring. J. Exp. Psychol. Hum. Learn. Mem. 5, 546–554.

Simson, R., Vaughan, H.G., Ritter, W., 1977. The scalp topography of potentials in auditory and visual Go/noGo tasks Electroencephalogr. Clin. Neurophysiol. 43, 864–875.

Taft, M., Hambly, G., 1985. The influence of orthography on phonological representations in the lexicon. J. Mem. Lang. 24, 320–335.

Taft, M., Zhu, X., Peng, D., 1999. Positional specificity of radicals in Chinese character recognition. J. Mem. Lang. 40, 498–519.

Taft, M., Zhu, X., Ding, G., 2000. The relationship between character and radical representation in Chinese. Acta Psychol. Sin. 32(Suppl.), 3–12.

van Turennout, M., Hagoort, P., Brown, C.M., 1997. Electrophysiological evidence on the time course of semantic and phonological processes in speech production. J. Exper. Psychol., Learn., Mem., Cogn. 23, 787–806.

van Turennout, M., Hagoort, P., Brown, C.M., 1998. Brain activity during speaking: from syntax to phonology in 40 milliseconds. Science 280, 572–574.

Ventura, P., Morais, J., Pattamadilok, C., Kolinsky, R., 2004. The locus of the orthographic consistency effect in auditory word recognition. Lang. Cogn. Processes 19, 57–95.

Wheeldon, L., Monsell, S., 1992. The locus of repetition priming of spoken word production. Q. J. Exp. Psychol., Human Exp. Psychol. 44(A), 723–761.

Wheeldon, L., Levelt, W.J.M., 1995. Monitoring the time course of phonological encoding. J. Mem. Lang. 34, 311–334.

Wheeldon, L., Morgan, J.L., 2002. Phoneme monitoring in internal and external speech. Lang. Cogn. Processes 17, 503–535.

Zhang, Q., Yang, Y., 2003. The determiners of picture naming latency (in Chinese). Acta Psychol. Sin. 35, 447–454.

Ziegler, J.C., Ferrand, L., 1998. Orthography shapes the perception of speech: the consistency effect in auditory word recognition. Psychon. Bull. Rev. 5, 683–689.

Ziegler, J.C, Ferrand, L., Montant, M., 2004. Visual phonology: the effects of orthographic consistency on different auditory word recognition tasks. Mem. Cogn. 32, 732–741.