Neuroscience 163 (2009) 252–265

THE TIME COURSE OF SEGMENT AND TONE ENCODING IN CHINESE SPOKEN PRODUCTION: AN EVENT-RELATED POTENTIAL STUDY

Q. ZHANG[a]* AND M. F. DAMIAN[b]

[a]*State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Datun Road 10A, Beijing, 100101, China*

[b]*Department of Experimental Psychology, University of Bristol*

Abstract—The present study investigated the time course of segment and tone encoding in Chinese spoken production with an event-related brain potentials (ERPs) experiment. Native Chinese speakers viewed a series of pictures and made Go/noGo decisions along dimensions of segmental onset or tone information of picture names. Behavioral data and onset latency of the N200 effect indicated that segmental information became available prior to tonal information. Moreover, the results of scalp distributions and onset latency patterns of the N200 effect on segmental and tonal decisions suggest that segmental and metrical encoding is relatively disassociated in Chinese spoken production. Our findings provide additional evidence from Chinese as a kind of non-alphabetic language concerning theories of phonological encoding based on alphabetic languages.

Key words: spoken production, phonological encoding, Go/noGo task, N200, Chinese.

Models of speech production assume that spoken word production involves several planning stages, such as conceptual preparation, lexical-semantic ("lemma") selection, phonological encoding and articulation. To maintain fluency, a speaker must coordinate retrieval of these different types of information with millisecond precision. 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 pattern and syllable structure.

Perhaps the most detailed theoretical account of this process, based on a multitude of empiric sources of evidence, is found in the WEAVER (Word Form Encoding by Activation and VERification) framework introduced by Roelofs (1997; Levelt et al., 1999). 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 form encoding proceeds in an incremental fashion from the beginning of a word to its end. A morpheme initially activates all its corresponding phonological segments simultaneously, together with information about their order. In parallel to this "segmental spell out" procedure, suprasegmental codes containing an abstract grouping of syllables into phonological words are retrieved. Both segments and metrical structure are subsequently merged in a syllabification process which proceeds in strictly sequential fashion. According to the WEAVER, then, segmental and metrical encoding occurs in parallel yet independent processing streams and results in the generation of syllabified phonological words. These serve as the input to a phonetic encoding process in which the metrical information is used to set parameters for loudness, pitch and duration.

Assessing the validity of the assumptions embedded in this framework is made difficult by the fact that empiric investigations of the way in which suprasegmental information is retrieved, and how it is combined with segmental representations, are at present relatively scarce. Roelofs and Meyer (1998) used a spoken production task ("implicit priming") in which characteristic facilitation effects are observed due to word-initial phonological overlap of responses produced within an experimental block. They showed that this priming effect is restricted to cases in which responses additionally share their number of syllables and stress pattern, but is not contingent on responses sharing abstract consonant-vowel syllable structure. These results were interpreted as showing that suprasegmental coding mainly consists of the number of syllables and their respective stress pattern, but that contrary to previous claims (e.g. Sevald et al., 1995), syllable-internal structure does not play a role in phonological encoding. Two recent studies have investigated metrical stress encoding in internally generated speech; both behavioral data (Schiller et al., 2006) and event-related potential (ERP) results (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 specifically suggest that metrical encoding, much like segmental encoding, is an incremental process.

The time course of information processing is an important aspect of language processing since it can help to constrain theoretical models of psycholinguistics (Levelt et al., 1999; Schiller, 2006). Based on a meta-analysis of word production experiments, Indefrey and Levelt (2004) suggested that for single word utterances such as those in the naming of pictures with labels of moderate to high frequency of occurrence, phonological encoding takes

*Corresponding author. Tel: +86-10-64836909; fax: +86-10-64872070.
E-mail address: zhangqf@psych.ac.cn (Q. Zhang).
Abbreviations: EEG, electroencephalogram; EOG, electro-oculogram; ERP, event-related potentials; WEAVER, word-form encoding by activation and verification.

0306-4522/09 $ - see front matter © 2009 IBRO. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuroscience.2009.06.015

place roughly between 250 and 450 ms after stimulus onset. Over the last few years, several ERP studies have been carried out to estimate the time course of information processing during language production (e.g. Abdel Rahman and Sommer, 2003; Abdel Rahman et al., 2003; Rodriguez-Fornells et al., 2002; Schiller, 2006; Schiller et al., 2003; Schmitt et al., 2000, 2001a,b; van Turennout et al., 1997, 1998; Zhang et al., 2007). In the present study, we aim at tracking the temporal and spatial characteristics of segmental and suprasegmental encoding within the phonological encoding module in Chinese spoken production, and we examine with electrophysiological measures whether or not the two processes run in parallel.

Segmental and suprasegmental information in Chinese spoken production

According to current linguistic theory (e.g. Goldsmith, 1990), the form representation of a word includes two representational tiers: a segmental tier which represents phonemes in terms of vowels and consonants, and a prosodic (or suprasegmental) tier which represents syllable structure, pitch, stress, or tone. Segmental encoding involves the retrieval of segments (phonemes) along with information about their order. For instance, naming of a picture of *banana* involves retrieval of the individual phonemes /b/, /ə/, /n/, /æ/, /n/, /a/ (adopted from Schiller et al., 2006). Phonological encoding additionally involves the retrieval of suprasegmental information, such as syllabic structure. In general, a syllable can be divided into an optional onset, an obligatory peak or nucleus, and an optional coda. There are some important differences in terms of syllable structure between English and Chinese, such that both onset and coda can be more complex in English than in Chinese (Davenport and Hannahs, 2005; Duanmu, 1999; Zhang, 1996).

Apart from syllable-internal structure, stress and tone are also considered suprasegmental types of representations. In stress languages such as English or 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, and no other word exists in English which has the same segments and order as “cognition,” but is stressed at the first or the last syllable. By contrast, in tonal languages such as Chinese, tone 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 an extremely important property in order to distinguish word meaning in Chinese, and similarly if one wants to express oneself fluently, retrieval of tone information is of central relevance. Yet, despite its importance, only very few studies have investigated the role of tonal codes in Chinese spoken production. Chen et al. (2002) reported evidence from a spoken production task (“implicit priming”) suggesting that for Chinese speakers, tone functions much like stress in English: characteristic priming effects in this task which for English speakers are constrained by stress are for Chinese speakers constrained by tone. The relative scarcity of such studies on the role of tonal information in spoken word production makes an investigation into the temporal and spatial aspects of tone generation particularly interesting.

Studies of segmental and suprasegmental encoding

The question whether segmental and suprasegmental representations involve different neural and cognitive mechanisms is presently under debate. A good number of behavioral (Ferrand and Segui, 1998; Meijer, 1996; Sevald et al., 1995) and brain lesion studies (Cappa et al., 1997; Laganaro et al., 2002) have brought forward evidence for separate storage and processing systems for segmental and suprasegmental information. Congruent with this evidence, current accounts of language production such as those by Dell (1988); Levelt (1992) and Levelt et al. (1999) suggest that segmental and suprasegmental codes are stored and retrieved independently from each other.

Importantly, this assumption is based on studies conducted in alphabetic languages, and hence it may not necessarily extend to other languages, and specifically not when tone constitutes an important suprasegmental property. Studies that investigated suprasegmental characteristics of tonal languages, documenting slips of the tongue in Thai (Gandour, 1977) and of aphasics (Packard, 1986), suggest that tones are as susceptible to errors as consonants and vowels. Shen (1993) and Wan (1996) observed a similar phenomenon as Gandour (1977) in Mandarin Chinese. Therefore, tone is typically characterized by linguists as phonemic, i.e. despite a tone functioning as a suprasegmental, it has a unit-like representation much like a segmental sound. Based on this view, it is quite plausible that in Chinese, segmental (e.g. consonantal) and suprasegmental (e.g. tone) information may involve very similar neural correlates. However, the very few existing empiric investigations seem to argue against this possibility. Luo et al. (2006) found that in a speech perception task, lexical tone variation evoked a stronger preattentive response in the right hemisphere than in the left hemisphere, whereas consonant variation yielded the opposite pattern. Liu et al. (2006) demonstrated in a simple tone or vowel production task that tone production was less-left lateralized than vowel production, although both processes showed left-hemisphere dominance. These studies suggest important processing differences between segmental and suprasegmental processing, even in tonal languages. The present study adds to these findings by exploring the brain activation patterns associated with segmental and suprasegmental encoding in Chinese production. To this aim, the N200 components associated with each type are compared against each other concerning their temporal and spatial characteristics.

Electrophysiological markers of the time course of spoken production: N200

The N200 is a negative-going waveform. In a Go/noGo task, participants are asked to respond to one type of stimulus, and to withhold their response to another type. Compared to the waveform on the Go trials, a particular

ERP component, namely N200, is typically observed on the noGo trials. This component is visible at a frontocentral region typically 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 (Jodo and Kayama, 1992; Sasaki and Gemba, 1993) 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 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 tends to occur later in time when it is related to language processing, compared to non-linguistic tasks (see Kutas and Schmitt, 2003).

Experimental paradigm

The experiment was carried out in Mandarin Chinese. Native Mandarin speakers saw pictures with names corresponding to monosyllabic nouns in the center of the computer screen. In separate experimental blocks, participants performed decisions based either on segmental, or on tonal, properties. They were asked to covertly name a picture and to press a button either if its name began with a particular target onset, or had a particular target tone. In the segmental decision task, for instance, if the target onset was /sh/ and the picture was 蛇 (/she2/, snake), participants were required to press a button with the dominant hand. The tonal decision task required determination of whether the name of the depicted object was of a particular tone (tone 1, 2, 3 or 4). For instance, if the target tone was tone 2 and the picture was 蛇, Participants were required to press a button with the dominant hand. Thus, by asking participants to monitor their own internal speech production, two difference waves could be calculated, one representing tone encoding and the other representing segment encoding.

In the field of speech perception, monitoring for phonological characteristics is a relatively widely used task (for an overview, see Connine and Titone, 1996). Concerning spoken production, a parallel procedure involves participants monitoring for a particular target in a covertly generated response. For instance, Wheeldon and Levelt (1995) presented Dutch participants with good knowledge of English with English words, and asked them to silently generate the Dutch translation, and to monitor the internally generated speech for a particular phoneme. By varying the position of the target phoneme, it was shown that reaction times steadily increased for later targets, suggesting that the internal code became incrementally available. Wheeldon and Levelt furthermore suggested that internal monitoring takes place at the processing level of the syllabified phonological word. More recent studies have used the phoneme monitoring technique when responses were elicited with semantically related prompt–response pairs which participants memorized prior to testing (e.g. “fish–dolphin”; Wheeldon and Morgan, 2002), or with responses elicited by pictures (e.g. Ganushchak and Schiller, 2006, 2008, 2009; Özdemir et al., 2007; Rodriguez-Fornells et al., 2002; Schiller, 2006; Schmitt et al., 2000, 2001a,b; van Turennout et al., 1997, 1998).

Concerning tonal properties, a number of studies investigating speech perception have employed tone monitoring tasks (e.g. Ye and Connine, 1999). With regard to spoken production, to our knowledge the only existing study which asked participants to monitor silently generated names of pictorial stimuli was reported by Zhang et al. (2007).

EXPERIMENTAL PROCEDURES

Participants

Twenty native Mandarin speakers participated in the experiment (10 females and 10 males, with a mean age of 21.7 years; range 20–25 years). Nineteen participants were right-handed and one was left-handed. The Edinburgh Handedness Inventory (Oldfield, 1971) was used to determine handedness. All participants were neurologically healthy, with normal or corrected-to-normal vision and normal hearing. They were paid for their participation.

Materials

To maximize comparability between segmental and tonal decision tasks, one would ideally design the experiment such that exactly the same target pictures are used under both conditions, with participants monitoring their names either for segmental or tonal characteristics. However, the pool of available candidates to choose from is severely limited when using pictorial stimuli, rendering this possibility impossible to achieve. Briefly, for the tone decision task one could select four sets of target names corresponding to the four tones (tone 1, tone 2, tone 3, and tone 4, respectively), and intermix items within each set with candidates from the other tone sets in order to form corresponding “match” and “no match” responses. To render the segment decision task comparable, it would be advantageous to choose four sets of target names, each containing a particular target segment, and again intermix each group with the remaining candidates to form “match” and “no match” responses. In fact, ideally each set of candidates with a particular tone would have equal numbers of occurrences of each target segment contained in them. Unfortunately, the number of available pictures with the relevant phonological characteristics is not sufficient, and hence the procedure was adapted to include more than four target segments, and stimuli were drawn from a set which was largely, but not entirely, overlapping for the two conditions. One hundred and nine target pictures with names corresponding to monosyllabic Chinese characters were selected from a database of standardized pictures in Chinese (Zhang and Yang, 2003). One hundred of them were used in the segmental decision task and 84 of them were used in the tone decision task, with 75 pictures used in both types of tasks. The pictures used here consisted of everyday objects with which participants were highly familiar (see Appendix 1 for details).

In the tone decision task, the four different tones of Chinese were used as targets. Four corresponding experimental blocks, plus four practice blocks, were generated. In each block, participants were asked to monitor for one particular target tone. Eighty-four pictures in the tone decision task corresponded to the four types of tones. The number of pictures in each block was 42, and half of pictures were of a particular tone, and half were not. In the tone decision task, each picture was presented twice, once as target and once as a nontarget. The 21 target pictures in a particular block were divided into three equal numbers and were presented as nontarget pictures in the other three blocks. Seven additional pictures were used as practice stimuli before each tone decision block.

In the segment decision task, 10 different onsets were used as targets, namely /y/, /zh/, /x/, /sh/, /q/, /ch/, /b/, /g/, /h/, and /j/. Ten corresponding experimental blocks, plus one practice block, were generated. In each block, participants were asked to monitor for one particular target onset. The number of targets in the experimental blocks was eight with onset of /y/ and /zh/, 10 with onset of /×/, /q/, /ch/, /b/, /g/, /h/, and /j/, and 14 with onset /sh/. In each block, half of the pictures had a particular onset, and the other half had not. In the segment decision task, each picture was presented once, half of the pictures as the target and half as the nontarget. Six additional pictures were used as practice stimuli.

Despite the fact that the stimuli used for the two monitoring tasks were largely overlapping (75 out of 100) the possibility must be excluded that potential deviations in the markers of segmental and tonal monitoring tasks arise from residual differences between the two sets. Hence it was ensured that they were matched on a number of characteristics. The corresponding picture names on the two decision tasks had similar written word frequencies (per million) (segment: mean=374; tone: mean=373, $t<1$, $P=0.88$). Word frequency was based on normative information reported by the Beijing Language Institute (1986). Differences between the two stimulus sets may also exist in the ease of object recognition and/or conceptual access. To exclude this possible confound, a word-picture matching control task was carried out (e.g. Jescheniak and Levelt, 1994; Özdemir et al., 2007). A fixation dot was presented for 500 ms, and after a 500 ms blank screen a character was presented for 800 ms. Then a picture was presented after a random interval between 1000 and 1800 ms. Participants decided with a manual response as quickly as possible whether or not the character matched the name of the picture. The idea is that this task requires access to the picture's visual and conceptual characteristics but not its name; consequently, latencies should exclusively reflect these non-lexical properties. Stimuli in the segment and tone decision sets showed very similar word–picture matching latencies (segment: mean=690 ms, SD=60 ms; tone: mean=686 ms, SD=59 ms $t<1$, $P=0.80$), implying that they were matched on visual and conceptual ease of access.

Further differences between the two stimulus sets may exist in articulatory difficulty and voice-key sensitivity. To ensure that this was not the case, we carried out a delayed naming task, a procedure which is assumed to allow speakers to complete all cognitive processes involved in identification and phonological encoding of the target, prior to the onset of motor execution (e.g. Kemeny et al., 2006; Rastle et al., 2005). Pictures were presented for 600 ms followed by a blank screen for a random interval between 1000 and 1800 ms. Then a visual cue was presented, and participants named pictures as accurately and quickly as possible. Latencies for the segment and tone decision stimulus sets were virtually identical (segment: mean=432 ms, SD=32; tone: mean=431 ms, SD=31, $t<1$, $P=0.93$), hence no differences in articulatory difficulty and voice-key sensitivity exist between the two sets.

Design

The experiment consisted of three stages. First, participants were familiarized with the set of experimental pictures by viewing each target picture for 3000 ms with the picture name printed below each picture. Then, 109 pictures were successively presented on the computer screen and participants were asked to provide the corresponding name. If their responses were not as expected, the experimenter corrected them until the participants could name the pictures with the correct words. This procedure, typical in studies that use object naming tasks to investigate spoken production, guaranteed that each participant knew and used the intended names of the pictures. Third, the experiment proper was administered. Each participant completed the two types of experimental blocks—the segment and tone decision task—twice. The order of the two blocks was counterbalanced among participants. Half of the participants completed the segment decision first and the tone decision second, then the tone decision and segment decision again (ABBA). The order of the tasks was reversed for the other half of the participants (BAAB). Participants received seven practice trials prior to each tonal decision block and six practice trials prior to each segment decision block. The order of the experimental blocks within each block (segment or tone decision) was randomized. There was a short break between blocks, and the next block started after participants indicated that they were ready to continue. The entire experiment lasted about 2 h, including the running of the pretest (see below), placement of the electrode cap, and the resting periods between blocks.

Procedure

Participants were tested individually in front of a computer screen in a soundproof chamber. During the experimental session, participants were instructed to carry out a Go/noGo task and to press a response button as soon as possible without overtly naming the picture. The space bar of the computer keyboard was used as the response key. Before each block was run, instructions were presented visually in order to let participants know the particular contingency (segment or tone) for the upcoming block.

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 500–1200 ms, a picture was presented for 2500 ms. Then a blank screen appeared for 1000 ms, followed by the next trial. The random interval between fixation and picture was used to avoid participants' systematic expectation in the form of a contingent negative variation (Walter et al., 1964). Participants were asked to put their dominant hand on the space bar and not to speak, or blink their eyes while a picture was presented on the screen.

Apparatus and recordings

The electroencephalogram (EEG) was recorded with 64 electrodes secured in an elastic cap (Electro Cap International, Neuroscan Inc., El Paso, TX, USA). The vertical electro-oculogram (EOG) (VEOG) was monitored with electrodes 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 right mastoid served as reference during the recording, but before the EEG was analyzed, the signal was re-referenced to software linked mastoids. Electrode impedances were kept below 5 $k\Omega$.

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 (−200–2000 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 $\mu\nu$ to 100 $\mu\nu$. 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.

Pre-experiment

A pretest assessed participants' ability to classify segment or tone of a Chinese character. This pretest consisted of three different tasks. Only those participants whose average accuracy on the pretest was more than 85% were allowed to participate in the main experiment. In the first task, four Chinese characters were presented on the computer screen per trial. Three of four characters had the same segmental onset whereas one had a different segmental onset. Participants were asked to identify the deviant word. This task consisted of two trials for familiarizing participants with the procedure, and 20 trials for testing their segmental onset discrimination ability. In the second task, four Chinese characters

Table 1. Mean reaction times (in millisecond; standard deviations in parentheses) for correct Go responses, and mean error rates in the Go/noGo responses contingent on segment and tone decisions

Response contingency condition	Time 1		Time 2	
	RT (SD)	Error rate (%)	RT (SD)	Error rate (%)
Segment	1017 (136)	3.050	868 (128)	1.400
Tone	1044 (162)	2.215	958 (168)	1.440

were presented on the computer screen per trial. Three of these words had the same tone whereas one had a different tone. This task consisted of four trials for familiarizing participants with the procedure, and 12 trials for testing their tone discrimination ability. In the third task, participants saw a Chinese character on the screen and simultaneously heard a sound which spells the pinyin of a character. In half of the cases, the auditorily presented characters had the same tones with the visually presented character, in the other half, they had different tones. Participants were asked to respond as to whether or not the auditorily presented character's tone matched the visually presented character's tone.

Out of the 20 participants who completed the three pretests, all made less than 15% errors, hence all participated in the following main experiment.

RESULTS

Push-button reaction times

Incorrect responses and reaction times longer than 2000 ms were excluded from the data analysis. These criteria accounted for 6.30% of the data in the Go/noGo=segment condition and 5.86% of the data in the Go/noGo=tone condition. The proportion of rejections was not significantly different for the two response contingency conditions. The mean reaction times for correct Go response and the mean error rate are shown in Table 1. Because each task was administered twice in the ABBA design, two corresponding latency means and error rates were computed for each task and half. Latencies decreased from the first to the second half, suggesting a practice effect. However, latencies on the segment decision task benefited more from repetition than latencies on the tone decision task. A paired sample *t*-test of the reaction times for the two types of tasks showed no significant difference at the time of the first test ($t(19)=1.21$, $P=0.24$), but a significant difference at the time of the second test ($t(19)=4.17$, $P<0.01$). Error rates were not significantly different between the two response conditions, either at the time of the first ($t(19)=1.19$, $P=0.25$) or the second ($t(19)=0.08$, $P=0.94$) test. In summary, accuracy measures did not differ significantly across the two tasks, suggesting that they were roughly matched on overall difficulty. Response latencies were slower in the metrical than in the segmental task at the second test, which reflects the fact that decisions based on segments take less time than decisions based on tones (see Discussion for an interpretation of this pattern).

N200 analysis

In a first step, we checked whether or not the two main conditions showed an overall difference. Following the Schiller et al. (2003) method, grand average waveforms for segment and tone conditions were computed across the Go and noGo trials in both conditions, respectively. Fig. 1 shows the two grand average ERP waveforms across Go and noGo trials for 20 participants at nine electrode sites (F3, Fz, F4, FC3, FCz, FC4, C3, Cz, and C4). Serial *t*-tests showed no significant differences between two curves at any time. This implies that there was no task effect in the data, which is what we expected because almost the same items were used and the task difficulty turned out to be the same as shown in the behavioral data. The waveforms for the two conditions were almost identical in the time window from 0 to 300 ms after picture onset. It has been suggested that the waveforms in this time window may be related to

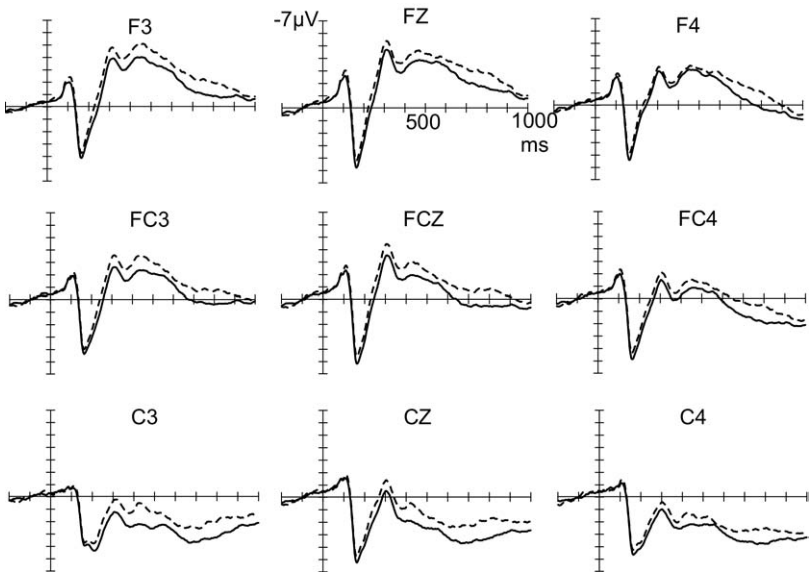

Fig. 1. Grand average waveforms for Go/noGo=segment (dotted lines) and tone (solid lines) conditions across Go and noGo trials over nine electrode sites (F3, Fz, F4, FC3, FCz, FC4, C3, Cz, and C4).

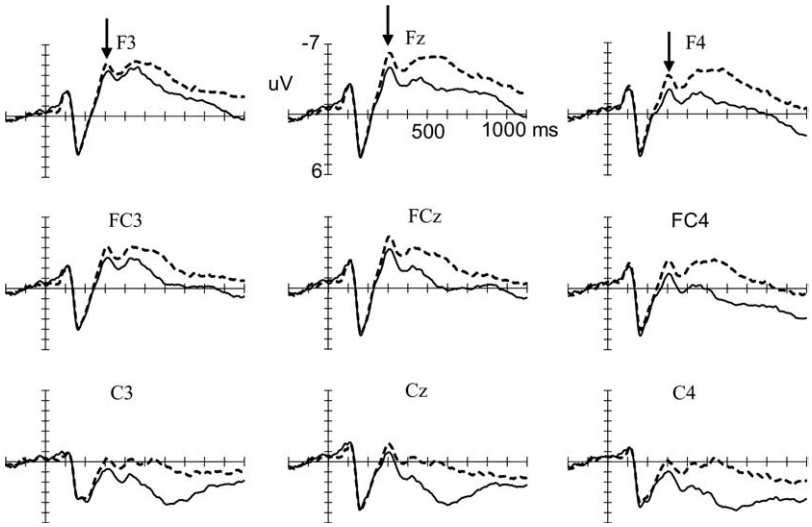

Fig. 2. Grand average waveforms for Go (solid lines) and noGo (dotted lines) trials in Go/noGo=segment condition over nine electrode sites (F3, Fz, F4, FC3, FCz, FC4, C3, Cz, and C4).

attentional mechanisms (Eimer, 1993; Patel and Azzam, 2005) and lexical selection (Indefrey and Levelt, 2004), so the comparison between two decision tasks suggests that the sets used in both tasks produced neither different attentional, nor different lexical selection, conditions. Secondly, because the pictures used for the two decision tasks were slightly different, we compared the pattern of 75 common pictures between two decision tasks with the pattern between 100 pictures on segment decision and 84 pictures on tone decision, and the comparison indicated that the two patterns were identical. Hence, in the following analysis we used the waveforms of all available stimuli, i.e. 100 pictures in the segment condition and 84 pictures in the tone condition.

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 (segment vs. tone). Fig. 2 and Fig. 3 show grand average ERP waveforms on the Go and noGo trials in the Go/noGo=segment and tone condition for 20 participants at nine fronto-central electrode sites (F3, Fz, F4, FC3, FCz, FC4, C3, Cz, and C4). Both response contingency conditions showed clear evidence of N200 effects, with the waveform on the noGo trials more negative than the Go trials (see arrows in Fig. 2 and Fig. 3).

Fig. 4 shows the grand average of difference waves (noGo minus go) for the Go/noGo=segment and tone conditions at nine electrode sites (the same locations as Fig. 2 and Fig. 3). Omnibus ANOVAs were computed on the N200 peak latencies and peak amplitudes of the difference wave with three within-participant variables: contingency condition (Go/nogo=Segment vs. tone) and region (left, middle and right) and electrode sites (frontal, frontal–central, and central). Greenhouse–Geisser correction was used when appropriate. For each participant,

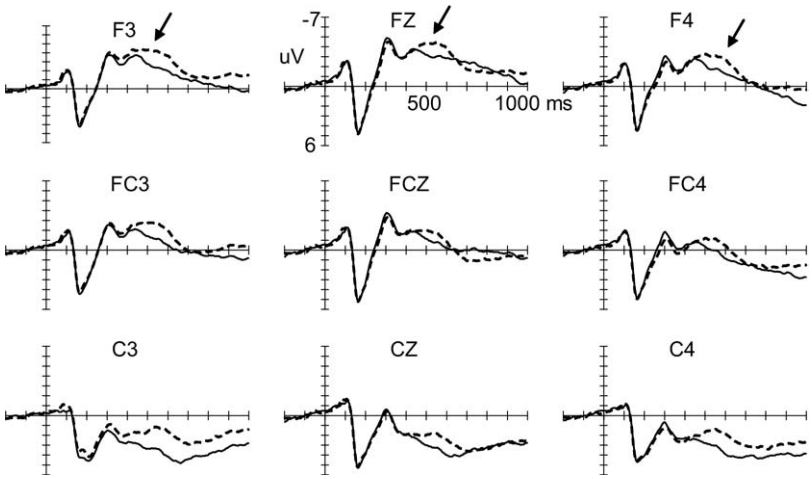

Fig. 3. Grand average waveforms for Go (solid lines) and noGo (dotted lines) trials in Go/noGo=tone condition over nine electrode sites (F3, Fz, F4, FC3, FCz, FC4, C3, Cz, and C4).

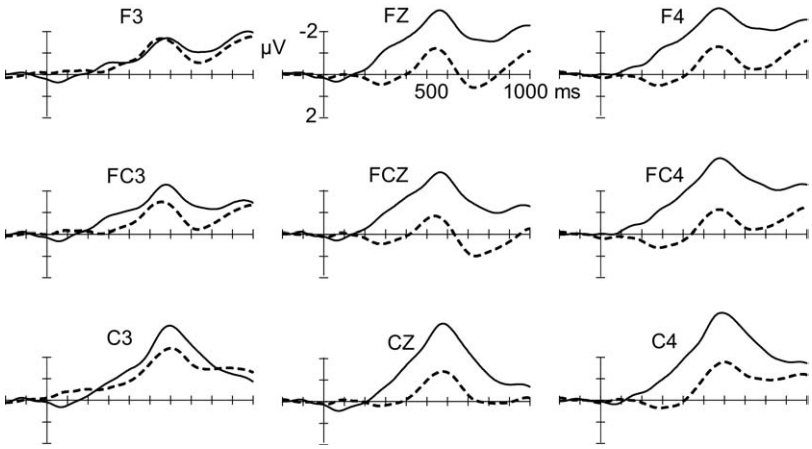

Fig. 4. Grand average difference waveforms (nogo minus go) for Go/noGo contingency on segment and tone conditions. Segment (solid lines) and tone (dotted lines) condition over nine electrode sites (F3, Fz, F4, FC3, FCz, FC4, C3, Cz, and C4).

peak latency and peak amplitude of the N200 effect between 200 and 800 ms were measured at each of the nine electrode sites for correct trials.

Peak latency

The main effect of contingency condition was not significant ($F(1,19)<1$), reflecting no difference in peak latencies. The mean peak latencies of the N200 across nine electrode sites were 592 ms (SD=131 ms) and 599 ms (SD=125 ms) when the Go/noGo decision was contingent on segmental and tonal information, respectively. Other effects were not significant.

Peak amplitude

The main effect of contingency condition was significant ($F(1,19)=11.05$, $P<0.01$), reflecting a difference in peak amplitude, with higher peaks in the segmental than in the tonal condition. The main effect of electrode sites was also significant ($F(2,38)=31.52$, $P<0.01$). Importantly, the interaction between contingency condition and region was significant ($F(2,38)=23.31$, $P<0.001$) (see Fig. 5A), which reflects the fact that scalp distributions of the two contingency conditions differed. The interaction between region and electrode site was also significant ($F(4,38)=4.05$, $P<0.05$). Other effects were not significant.

Onset latency

Onset latencies of the difference wave were estimated with the procedure outlined in Li et al. (2008a) which is also used in a number of other language processing studies (e.g. Li et al., 2008b; Rodriguez-Fornells et al., 2002; van den Brink et al., 2001; van den Brink and Hagoort, 2004). Serial *t*-tests were conducted over nine electrode sites with a step size of 10 ms in the time window 200–600 ms after picture onset, for both Go/noGo contingency conditions (i.e. 200–210 ms, 210–220 ms). The onset latency of the N200 was defined as the point at which five consecutive *t*-tests yielded one-tailed significant results (in the same direction). Table 2 shows the onset latency of the N200 in both conditions over nine electrode sites. Concerning the N200 at the frontal–central region (Fz, FCz, and Fz), the mean onset latencies were 283–293 ms and 483–493 ms when the Go/noGo decision was contingent on segmental and tonal information, respectively. Fig. 5A shows the scalp distributions of the N200 effects in the Go/noGo=segment (566–616 ms) and tone (576–626 ms) conditions. McCarthy and Wood (1985) suggested that in case of differences in effect size (maximum–minimum voltage range) between conditions, the data should be normalized (into the same min–max range) prior to any comparison. Following McCarthy and Wood's normalization method (see also Haig and Gordon, 1997), the normalized value *N* is computed by the following equation:

$$N=(x-n)/(m-n)$$

Where x represents each site value, m represents the maximum average site value, and *n* represents the minimum average site value. According to this equation, *N* is variable between 0 and 1. Fig. 5B shows the *n*-values of the three triples of electrodes on the left, middle and right sides of the scalp in the Go/noGo=segment and tone conditions.

DISCUSSION

The question of whether segmental and suprasegmental types of representation involve different neural and cognitive mechanisms in spoken production is at present not entirely settled, but for alphabetic languages there is mounting evidence that they are indeed separate. As a consequence, the leading cognitive models of spoken production assume that the two types of information are stored independently from each other, and are retrieved in parallel. The present study adds to this debate by (i) providing information about the relative time course of segmental and suprasegmental retrieval in Chinese, and (ii) identifying the relative degree of lateralization of the two types of codes.

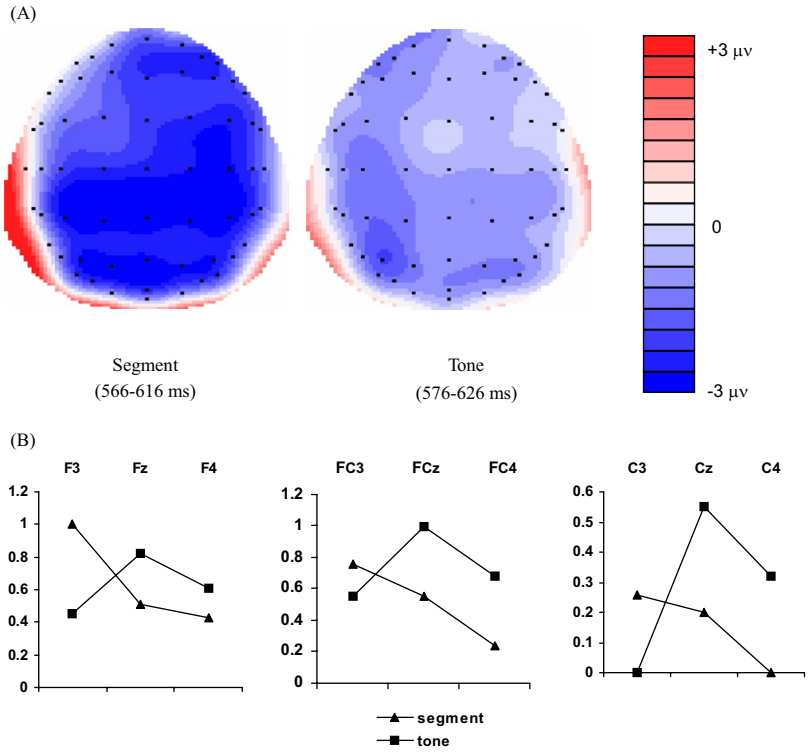

Fig. 5. (A) Scalp distribution of the N200 effects for the Go/noGo=segment condition (mean amplitudes of the time window 566–616 ms) and the Go/noGo=tone condition (mean amplitudes of the time window 576–626 ms). (B) The *n*-values from three triples of electrodes on the left, middle and right sides of the scalp after normalization (range: 0–1) in the Go/noGo=segment and tone conditions. For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.

To begin with, behavioral data were compared to rate participants' ability to classify segments vs. tones of a Chinese character. Accuracy measures did not differ significantly across the two tasks, suggesting that they were roughly matched on overall difficulty. Response latencies were slower in the metrical than in the segmental task at the second test, which reflects the fact that decisions based on segments take less time than decisions based on tones. The logic we followed was based on two previous studies. Schmitt et al. (2000) investigated the time course of semantic and phonological encoding with a dual-choice Go/noGo task. Mean reaction times were 617 ms for sim-

Table 2. Onset latencies of N200 effect in Go/noGo responses, contingent on segment and tone decisions, over nine electrode sites (F3, Fz, F4, FC3, FCz, FC4, C3, Cz, and C4)

Go/noGo contingency condition	Region					
	Left		Middle		Right	
	Site		Site		Site	
Segment	F3	420–430	Fz	260–270	F4	240–250
		$t(19)=-2.120$*		$t(19)=-2.221$*		$t(19)=-2.267$*
	FC3	280–290	FCz	280–290	FC4	230–240
		$t(19)=-2.156$*		$t(19)=-2.171$*		$t(19)=-2.171$*
	C3	270–280	Cz	310–320	C4	260–270
		$t(19)=-2.154$*		$t(19)=-2.235$*		$t(19)=-2.300$*
	Overall	323–333		283–293		243–253
Tone	F3	430–440	Fz	490–500	F4	490–500
		$t(19)=-2.216$*		$t(19)=-2.223$*		$t(19)=-2.177$*
	FC3	440–450	FCz	500–510	FC4	510–520
		$t(19)=-2.315$*		$t(19)=-2.006$*		$t(19)=-2.290$*
	C3	450–460	Cz	460–470	C4	460–470
		$t(19)=-2.381$*		$t(19)=-2.131$*		$t(19)=-2.179$*
	Overall	440–450		483–493		487–497

* $P<0.05$.

ple semantic decisions, and 841 ms for simple phonological decision; the 224-ms difference was significant. Heim et al. (2003) investigated phonological processing during language production with the fMRI method. Participants were asked to perform either a semantic decision task (SEM: natural or manmade?) or one of two phonological decision tasks on the initial phoneme of a picture name (Phon1: /b/ or not?; Phon2: vowel or not?). Mean reaction times were 646 ms for the semantic task, 812 ms for the Phon1 task, and 1015 ms for the Phon2 task, with differences significant between all tasks. Both Schmitt et al. (2000) and Heim et al. (2003) suggested that the pattern of longer reaction times for phonological relative to semantic processing during language production indicated that the simple choice based on semantics could be carried out faster than the choice based on phonology. At the same time, they argued that, based on the absence of significant differences in error rates, the tasks were matched on overall difficulty. Following this logic we likewise suggest that the segmental and tonal decision tasks used in the present study are roughly comparable in terms of difficulty. At the same time, it is clearly the case that simple segmental judgments can be carried out faster than tonal judgments. One should note that the mean error rates on segments decisions were slightly higher than on tone decisions, indicating a potential speed–accuracy tradeoff.

With regard to the ERP data, we interpret the results as indicating a relatively clear dissociation, both temporal and spatial, between segmental and suprasegmental codes. The N200 analysis is based on the assumption that increased negativity on noGo trials in comparison with Go trials reflects the moment in time at which the relevant information necessary to withhold a response must have been encoded. With regard to the temporal characteristics of the electrophysiological markers, the *peak latency* of the N200 effect indicated that segmental information (592 ms) is retrieved almost simultaneously with metrical information (599 ms) in spoken word production. The virtually identical numerical values indicating peak latency in the two conditions would appear to imply that segmental (onset) and suprasegmental (tonal) information became available simultaneously and in parallel. According to Wheeldon and Levelt (1995), the processing level at which participants carry out internal monitoring for segmental information is the phonological word, i.e. a syllabified phonological representation of an utterance. At this level, segmental and suprasegmental information has been merged into a common phonological code, and so it is not unexpected that peak latencies corresponding to the two types of judgments should show similar temporal properties.

On the other hand, the *onset latency* of the N200 effect showed a different pattern. When the Go/noGo decision was contingent on segmental information, the onset latencies of the N200 across three frontal–central sites (Fz, FCz, and Cz) were around 200 ms (segments: 283–293 ms vs. tone: 483–493 ms) earlier than when the decision was contingent on tonal information. This pattern suggests that segmental information was available ahead of tonal information. The push-button reaction times also showed that segmental information was encoded prior to tonal information, and the data from onset latencies are therefore to some extent consistent with the behavioral results. The relatively few previous studies in which not only peak, but also onset, latencies were computed generally showed relatively clear agreement between the two characteristics. For example, Schmitt et al. (2000) found that the mean difference in the onset latencies of two N200 effects corresponding to semantic and phonological information retrieval was 119 ms, while in the peak latencies it was 89 ms, resulting in a discrepancy of only 30 ms. In the present study, by contrast, we find virtually identical peak latencies, coupled with substantially asynchronous onsets. The interpretation of this pattern is not straightforward because the relative properties of onset vs. peak latencies of N200 have yet to be fully identified.

One possibility is that the temporal markers reflect two parallel but separate processing streams which start at different times but finish simultaneously. In the WEAVER++ model (Levelt et al., 1999), segments are activated in parallel, and simultaneously, a metrical frame is retrieved. Subsequently, the retrieved segments are inserted into the metrical frame from left to right in an incremental fashion, resulting in a syllabified phonological word. Concerning the tone monitoring task, previous evidence suggests that it is based on the level of the phonological word. In an implicit priming paradigm conducted with Mandarin speakers, Chen et al. (2002) did not find an effect of overlapping tone by itself, and they therefore proposed that tone cannot usefully be prepared in advance. This finding is consistent with a tone's characteristics as described in the introduction. Tone is lexical distinctive, and identical segments that differ in tone would be associated with different motor movements. In fact, Chen et al. found that any attempt to prepare the tone in the absence of information about segmental content may slow production. These finding therefore imply that tone is monitored in our study based on the availability of a phonological word. Concerning segmental monitoring, perhaps participants in our study may have been able to base their decisions either on the level of the phonological word, or on the preceding processing level in which a set of phonemes is made available which has not yet been merged with suprasegmental codes. Speaking against this possibility is that Wheeldon and Levelt (1995) have convincingly argued that phoneme monitoring in a production task takes place at the level of the syllabified phonological word, and that earlier processing levels are not consciously available to speakers. Therefore, both segmental and tonal monitoring would appear to be based on the same representation (the phonological word) and it is not clear why the temporal markers for each task indicate differential availability.

It should be noted that peak amplitudes differed significantly between the two types of tasks, with the segmental condition generally showing a larger N200 amplitude than the tonal condition. 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 it is sensitive to task difficulty not

only in monkeys (Gemba and Sasaki, 1989) but also in a priming task in humans (Kopp et al., 1996). It is possible that this amplitude difference is related to processing speed, as suggested by Sasaki and Gemba (1993). They found that the N200 effect in the frontal region was larger for quick than for slow processes. Schmitt et al. (2000) also found this pattern: faster semantic processes were associated with higher amplitudes than relatively slower phonological processes. In addition, the pattern of error rates suggests that the difficulty of tone decisions was comparable to the one of segmental decisions, because there was no significant difference on error rates between two tasks (see Schmitt et al., 2000; Heim et al., 2003). Our finding on behavioral and amplitude data generally agrees with this pattern: the faster segmental encoding was associated with higher amplitudes than the relatively slower tonal encoding. Therefore, the significantly larger peak amplitude on segmental decisions than tone decisions is unlikely to be attributable to the possibility that segmental decisions were more difficult than tone decisions.

Abdel Rahman and Sommer (2003) proposed that the N200 (onset or peak latency) may indicate the termination of processing of specific information in a dual-choice Go/noGo task, but is not related to the relative timing of the beginning of these processes. If these components speak to termination, the onset and the peak of the N200 may indicate an inconsistent pattern: the onset indicates that segment retrieval terminates earlier than tone retrieval, whereas the peak indicates that both retrieval processes terminate simultaneously. So far, it is not clear whether the observed N200 pattern speaks to availability or rather to differences in termination of a process. Based on the above discussion—the N200 effect indicating either information availability or termination—it seems reasonable to view the peak latency of the N200 effect as providing an upper limit on the time course of segmental and tonal encoding during implicit naming.

The N200 reported here does not closely resemble a "classic" N200, either in the time course or in scalp distribution. It is, however, generally in line with existing findings in the field of language production. Concerning the time course, Kutas and Schmitt (2003) clearly pointed out that the N200 tends to occur later in time when it is related to language processing, compared to non-linguistic tasks. The peak latencies were around 600 ms in both contingency conditions, and these latencies are comparable with other existing studies on a general level. For example, Schiller (2006) found that peak latencies of N200 were 475 and 533 ms when the decisions were dependent on initial and final stress, respectively. Schmitt et al. (2000) found that the mean peak latency of the N200 was 473 ms for initial phoneme decision. Rodriguez-Fornells et al. (2002) found a 563 mean peak latency for a phonological decision task. Concerning the scalp distribution, various patterns have been reported in previous electrophysiological language production studies with Go/noGo task. For instance, Schiller (2006) investigated the time course of lexical stress encoding in language production. The negative maximum was found at the frontal region when the decision was dependent on initial stress, whereas the negative maximum was found at the posterior region when the decision was dependent on final stress (see their Fig. 2 and Fig. 3 for details). Schiller et al. (2003) investigated the time course of phonological encoding during speech production planning. They found that the N200 effect showed a reversal in polarity (negative for metrical decision and positive for syllabic decision at frontal–central): in the syllabic condition, the negative maximum occurred at the left-frontal region, whereas the negative maximum appeared at the right-posterior region in the metrical condition (see their Fig. 5 for details). The N200 effects at the frontal–central region were interpreted as reflecting response inhibition and were analyzed within a language production framework. Following this research tradition, we analyzed the N200 effects at the frontal–central region to explore the time course of segmental and metrical encoding in Chinese spoken production. Obviously, the differences between existing studies concerning time course and location of these effects are in need of further explicit investigation.

One possible reason for not observing "classic" N200 might be the proportion of Go and noGo trials. It has been suggested that the amplitude of the waveform on the noGo trials increases with a higher proportion of Go relative to noGo trials (Donkers and van Boxtel, 2004). Donkers and van Boxtel also reported similar amplitudes for Go and noGo trials for 50% go probability. To our knowledge, all existing language production studies have used 50% Go and 50% noGo trials, and this may decrease the N200 amplitude at the frontal–central electrode sites. In addition, Nieuwenhuis et al. (2003) and Donkers and van Boxtel (2004) suggested that the N200 in Go/noGo tasks reflects response conflict monitoring. It is possible that the scalp distributions change due to this reason. Another possibility might be that the time course of the linguistic task was delayed in comparison with a non-linguistic one. The negative maximum of the N200 may be delayed and appeared at a later time window, and then change the N200's scalp distribution. This was what we found here: the N200s are most pronounced in the later positivity of the signal, but not in the negativity in the signal as is the case for the classic N200. It should be noted that for the N200 effect reported in Schmitt et al. (2001), the maximum difference between Go and noGo trials similarly appeared on the positive waveforms when Go/noGo decisions were based on conceptual representations. Independent of whether N200 reflects response inhibition or conflict monitoring, we believe that we can infer the time course of segmental and metrical encoding from its onset and peak latencies.

The patterns of onset latencies over nine sites were also distinct in the two contingency conditions (see Table 2). When the Go/noGo decision was contingent on segmental information, the onset latency was shortest at the right-frontal region (FC4: 230–240 ms) and longest at the left-frontal region (F3: 420–430 ms). It therefore seems that the N200 appeared at the right-frontal region first, then at the frontal–central region, and finally at the left-frontal region. When the Go/noGo decision was contingent on

tonal information, the onset latency was shortest at the left-frontal region (F3: 430–440 ms) and longest at right-frontal region (FC4: 510–520 ms). The N200 was observed firstly at the left-frontal region, then at the frontal–central region, and finally at the right-frontal region. The occurrence patterns of the N200 effects in both contingency conditions were opposite. These results additionally indicate that segment and tone encoding are two relatively independent processes.

Chinese lexical tone is carried by the vowel of a word's syllable (i.e. its medial part) whereas our segmental decision task was based on the initial consonant of a word. As outlined in the introduction, several studies have suggested that both segmental and metrical encoding proceeds incrementally from the beginning of a word towards its end (Schiller, 2006; Schiller et al., 2006; Wheeldon and Levelt, 1995; Wheeldon and Morgan, 2002; van Turennout et al., 1997). This implies that participants may have been able to carry out the segment task based on a portion of the word which was available slightly prior to the one on which the tone decision task was based. Fortunately, this discrepancy is likely to be very minor: Wheeldon and Levelt (1995) found that in phoneme monitoring, the reaction time difference between the onset and the offset of the first CVC syllable was only 55 ms. van Turennout et al. (1997) found a slightly larger 80 ms difference between monitoring for a word's onset and its offset with a dual-choice Go/noGo task, but with experimental targets which were 50% longer than those of Wheeldon and Levelt. In combination these results suggest that a discrepancy in the availability of initial consonant and central vowel in our study may have caused a slight asynchrony in temporal markers. But the measured difference in N200 onset latencies between segment and tone encoding in our study was around 200 ms, a discrepancy which clearly cannot be accounted for by the relative availability of onset consonant and central vowel. In future studies it may be advisable to base both the tonal and the segmental monitoring task on the central vowel, and hence to equate the two decision tasks in terms of the availability of respective codes. Again, it will not be straightforward to identify appropriate pictorial stimuli to accomplish this objective, due to the limitations inherent in using pictures as targets. In summary, despite the reservations associated with the interpretation of our findings, we obtain a rather clear dissociation between the temporal markers associated with segment vs. tone decision tasks.

Turning to the spatial characteristics of the N200 waves corresponding to the two types of tasks, we find clear evidence of differential lateralization of segmental vs. tonal processing, as evidenced by the different scalp distributions shown in Fig. 5A.

As can be seen in Fig. 5, at the gross level, segment and tone decision induced similar brain activation patterns. These results are generally in accordance with previous neuroimaging studies (Gandour et al., 2000; Hsieh et al., 2001). However, the peak amplitude of the N200 in the Go/noGo=segment condition was larger than the Go/noGo=tone condition, and the interaction between contingency condition and sites was significant. The N200 effect was significantly larger on the right side of the scalp than on the left in response to segmental decision but larger in amplitude on the left of the scalp than on the right in response to tonal decision. In other words, the brain activation associated with segmental decisions was more right-lateralized, and activation associated with tone (metrical information) decision was more left-lateralized. Comparisons between segment and tone conditions on the *n*-values also indicated that there exists a significant interaction between segment and tone decision tasks on amplitudes (see Fig. 5B). Different regions and different degrees of these regions involved in the tone and segmental decisions imply that segmental encoding and metrical encoding might run relatively independently. Hence the results seem to support the notion of a dissociation between tone and segment processing during spoken word production.

Curiously, however, the present lateralization results deviate from previous studies investigating segmental and suprasegmental processing in Chinese. Luo et al. (2006) found that in a speech perception task, the consonant contrast produced a more left-lateralized pattern whereas the lexical tone produced a more right-lateralized pattern. Luo et al. pointed out that hemispheric dominance depends mainly on acoustic cues before speech input is mapped onto a semantic representation in the processing stream. Liu et al. (2006) compared the production pattern of Chinese lexical tones and vowels with an adaptation paradigm. Their fMRI results found that tone production was less left-lateralized than vowel production, although both showed left-hemisphere dominance. The discrepancy between these previous, and our own, results, can likely be attributed to distinct processes which were involved in their and our studies. Luo et al. (2006) asked Chinese participants to ignore auditory stimuli and to watch a silent cartoon movie. This speech perception task involved only early auditory processing, and no semantic processing was required. Liu et al. (2006) asked Chinese speakers to repeatedly name Chinese characters and pinyin (romanized phonetic system for Chinese language) which varied in terms of tones and vowels; hence the study employed a pinyin and character reading task. By contrast, the task in the present study involved implicit picture naming, which is assumed to involve all processes of overt picture naming except articulation, and hence could be considered an approximation of spoken production. The exact locus at which tonal information in speaking is represented remains to be identified, and our results suggest a bias toward the left hemisphere. On the other hand, our finding that a segmental monitoring task shows larger activation in the right compared to the left hemisphere is admittedly puzzling and counterintuitive, and needs to be confirmed by independent studies. At minimum, the present results of scalp distributions indicate some degree of disassociation between segmental and metrical encoding during spoken word production.

CONCLUSION

In sum, we compared the time course of segmental and metrical encoding directly in the present experiment, and found that segment information became available prior to tone information. Moreover, the opposite onset latency patterns over nine electrode sites and the distinct scalp distributions of the N200 at both conditions indicate a dissociation of segment and tone encoding in Chinese spoken word production. Our findings provide additional evidence from a non-alphabetic language concerning theoretical models of phonological encoding which are typically based on alphabetic languages.

Acknowledgments—This research was supported by grants from the National Natural Science Foundation of China (30400134, 30870761) and Young Scientist Foundation of Institute of Psychology (07CX102010) to Qingfang Zhang, and an International Incoming Fellowship (IIF-2007/R1) from the Royal Society to Qingfang Zhang and Markus Damian.

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. Cogn Brain Res 16:372–382.

Abdel Rahman R, van Turennout M, Levelt JWM (2003) Phonological encoding is not contingent on semantic feature retrieval: an electrophysiological study on object naming. J Exp Psychol Learn Mem Cogn 29:850–860.

Beijing Language Institute (1986) Modern Chinese frequency dictionary [in Chinese]. Beijing: Beijing Language Institute Publisher.

Cappa SF, Nespor M, Ielasi W, Miozzo A (1997) The representation of stress: evidence from an aphasic patient. Cognition 65:1–13.

Chen TY, Chen TM, Dell GS (2002) Word-form encoding in mandarin as assessed by the implicit priming task. J Mem Lang 46:751–781.

Connine CM, Titone D (1996) Phoneme monitoring. Lang Cogn Proc 11:635–645.

Davenport M, Hannahs SJ (2005) Introducing phonetics and phonology. London: Hodder Arnold.

Dell GS (1988) The retrieval of phonological forms in production: tests of predictions from a connectionist model. J Mem Lang 27:124–142.

Donkers FCL, van Boxel GJM (2004) The N2 in go/no-go task reflects conflict monitoring not response inhibition. Brain Cogn 56:165–176.

Duanmu S (1999) Syllable structure in Chinese. In: The syllable: views and facts. Studies in generative grammar 45 (van der Hulst H, Ritter N, eds), pp 477–499. Berlin: Mouton de Gruyter.

Eimer M (1993) Effects of attention and stimulus probability on ERPs in a go/nogo task. Biol Psychol 35:123–138.

Ferrand L, Segui J (1998) The syllable's role in speech production: are syllables chunks, schemas, or both? Psychon Bull Rev 5:253–258.

Gandour J (1977) Counterfeit tones in the speech of southern Thai bidialectals. Lingua 41:125–143.

Gandour J, Wong D, Hsieh L, Weinzapfel B, van Lancker D, Hutchins G (2000) A crosslinguistic PET study of tone perception. J Cogn Neurosci 12:207–222.

Ganushchak L, Schiller NO (2006) Effects of time pressure on verbal self-monitoring. Brain Res 1125:104–115.

Ganushchak LY, Schiller NO (2008) Brain error-monitoring activity is affected by semantic relatedness: an event-related brain potentials study. J Cogn Neurosci 20:927–940.

Ganushchak LY, Schiller NO (2009) Speaking one's second language under time pressure: an ERP study on verbal self-monitoring in German-Dutch bilinguals. Psychophysiology 46:410–419.

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.

Goldsmith J (1990) Autosegmental and metrical phonology. Cambridge, MA: Basil Blackwell.

Haig AR, Gordon E, Hook S (1997) To scale or not to scale: McCarthy and Wood revisited. Electroencephalogr Clin Neurophysiol 103: 323–325.

Heim S, Opitz B, Müller K, Friederici AD (2003) Phonological processing during language production: fMRI evidence for a shared production-comprehension network. Cogn Brain Res 16(2):285–296.

Hsieh L, Gandour J, Wong D, Hutchins G (2001) Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. Brain Lang 76:227–252.

Indefrey P, Levelt WJM (2004) The spatial and temporal signatures of word production components. Cognition 92:101–144.

Jescheniak JD, Levelt WJM (1994) Word frequency effects in speech production: retrieval of syntactic information and of phonological form. J Exp Psychol Learn Mem Cogn 20:824–843.

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.

Kemeny S, Xu J, Park GH, Hosey LA, Wettig CM, Braun AR (2006) Temporal dissociation of early lexical access and articulation using a delayed naming task—an fMRI study. Cereb Cortex 16:587–595.

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 R, Schmitt BM (2003) Language in microvolts. In: Mind, brain, and language: multidisciplinary perspectives (Banich MT, Mack MA, eds), pp 171–209. New York, NY: Lawrence Erlbaum Associates Inc.

Laganaro M, Vacheresse F, Frauenfelder UH (2002) Selective impairment of lexical stress assignment in an Italian-speaking aphasic patient. Brain Lang 81:601–609.

Levelt WJM (1992) Accessing words in speech production: stages, processes and representations. Cognition 42:1–22.

Levelt WJM, Roelofs A, Meyer AS (1999) A theory of lexical access in speech production. Behav Brain Sci 22:1–75.

Li X, Hagoort P, Yang Y (2008a) Event-related potential evidence on the influence of accentuation in spoken discourse. J Cogn Neurosci 20:906–915.

Li X, Yang Y, Hagoort P (2008b) Pitch accent and lexical tone processing in Chinese discourse comprehension: an ERP study. Brain Res 1222:192–200.

Liu L, Peng D, Ding G, Jin Z, Zhang L, Li K, Chen C (2006) Dissociation in the neural basis underlying Chinese tone and vowel production. Neuroimage 29:515–523.

Luo H, Ni J-T, Li Z-H, Li X-O, Zhang D-R, Zeng F-G, Chen L (2006) Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. Proc Natl Acad Sci U S A 1003:19558–19563.

McCarthy G, Wood CC (1985) Scalp distribution of event-related potentials: an ambiguity associated with analysis of variance models. Electroencephalogr Clin Neurophysiol 62:203–208.

Meijer PJA (1996) Suprasegmental structures in phonological encoding: the CV structure. J Mem Lang 35:840–853.

Meyer AS (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 AS (1991) The time course of phonological encoding in language production: phonological encoding inside a syllable. J Mem Lang 30:69–89.

Nieuwenhuis S, Yeung N, Van den Wildenberg W, Ridderinkhof KR (2003) Electrophysiological correlates of anterior cingulate function

in a go/no-go task: effects of response conflict and trial type frequency. Cogn Affect Behav Neurosci 3:17–26.
Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97–113.
Özdemir R, Roelofs A, Levelt WJM (2007) Perceptual uniqueness point effects in monitoring internal speech. Cognition 105:457–465.
Packard JL (1986) Tone production deficits in nonfluent aphasic Chinese speech. Brain Lang 29:212–223.
Patel SH, Azzam PN (2005) Characterization of N200 and P300: selected studies of the event-related potential. Int J Med Sci 2: 147–154.
Rastle K, Croot KP, Harrington JM, Coltheart M (2005) Characterizing the motor execution stage of speech production: consonantal effects on delayed naming latency and onset duration. J Exp Psychol Hum Percept Perform 31:1083–1095.
Rodriguez-Fornells A, Schmitt BM, Kutas M, Münte TF (2002) Electrophysiological estimates of the time course of semantic and phonological encoding during listening and naming. Neuropsychologia 40:778–787.
Roelofs A (1997) The WEAVER model of word-form encoding in speech production. Cognition 64:249–284.
Roelofs A, Meyer AS (1998) Metrical structure in planning the production of spoken words. J Exp Psychol Learn Mem Cogn 24:922–939.
Sasaki K, Gemba H (1993) Prefrontal cortex in the organization and control of voluntary movement. In: Brain mechanisms of perception and memory: from neuron to behavior (Ono T, Squire LR, Raiche ME, Perrett DI, Fukuda M, eds), pp 473–496. New York: Oxford University Press.
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 NO (2006) Lexical stress encoding in single word production estimated by event-related brain potentials. Brain Res 1112:201–212.
Schiller NO, Bles M, Jansma BM (2003) Tracking the time course of phonological encoding in speech production: an event-related brain potential study. Cogn Brain Res 17:819–831.
Schiller NO, Jansma BM, Peters J, Levelt WJM (2006) Monitoring metrical stress in polysyllabic words. Lang Cogn Proc 21:112–140.
Schmitt BM, Münte TF, Kutas M (2000) Electrophysiological estimates of the time course of semantic and phonological encoding during implicit naming. Psychophysiology 37:473–484.
Schmitt BM, Rodriguez-Fornells A, Kutas M, Münte TF (2001a) Electrophysiological estimates of semantic and syntactic information access during tacit picture naming and listening to words. Neurosci Res 41:293–298.
Schmitt BM, Schiltz K, Zaake W, Kutas M, Münte TF (2001b) An electrophysiological analysis of the time course of conceptual and syntactic encoding during tacit picture naming. J Cogn Neurosci 13:510–522.
Sevald CA, Dell GS, Cole JS (1995) Syllable structure in speech production: are syllables chunks or schemas? J Mem Lang 34: 807–820.
Shen J-X (1993) Slips of the tongue and the syllable structure of mandarin Chinese. In: Essays on the Chinese language by contemporary Chinese scholars (Yau S-C, ed), pp 139–161. Paris, France: Centre de Recherdches Linguistiques sur l'Asie Orientale–Ecole des Hautes Etudes en Sciences Sociales.
Simson R, Vaughan HG, Ritter W (1977) The scalp topography of potentials in auditory and visual go/noGo tasks. Electroencephalogr Clin Neurophysiol 43:864–875.
van den Brink D, Brown CM, Hagoort P (2001) Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. J Cogn Neurosci 13:967–985.
van den Brink D, Hagoort P (2004) The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. J Cogn Neurosci 16:1068–1084.
van Turennout M, Hagoort P, Brown CM (1997) Electrophysiological evidence on the time course of semantic and phonological processes in speech production. J Exp Psychol Learn Mem Cogn 23:787–806.
van Turennout M, Hagoort P, Brown CM (1998) Brain activity during speaking: from syntax to phonology in 40 milliseconds. Science 280:572–574.
Walter WG, Cooper R, Aldridge VJ, McCallum WC, Winter AL (1964) Contingent negative variation: an electrical sign of sensorimotor association and expectancy in the human brain. Nature 203:380–384.
Wan I-P (1996) Tone errors in Mandarin Chinese. Master's thesis, State University of New York at Buffalo.
Wheeldon L, Levelt WJM (1995) Monitoring the time course of phonological encoding. J Mem Lang 34:311–334.
Wheeldon L, Morgan JL (2002) Phoneme monitoring in internal and external speech. Lang Cogn Proc 17:503–535.
Ye Y, Connine CM (1999) Processing spoken Chinese: the role of tone information. Lang Cogn Proc 14:609–630.
Zhang J (1996) On the syllable structures of Chinese relating to speech recognition. In: International Conference on Spoken Language Processing–1996, pp 2450–2453.
Zhang Q, Damian MF, Yang Y (2007) Electrophysiological estimates of the time course of tonal and orthographic encoding in Chinese speech production. Brain Res 1184:234–244.
Zhang Q, Yang Y (2003) The determiners of picture naming latency [in Chinese]. Acta Psychol Sin 35:447–454.

Appendix 1. Stimuli used in the Go/noGo task

Picture names in Chinese	Picture names in English	Picture names' pinyin (phonetics)	Picture names' tone
蚁	Ant	Yi	3
斧	Axe	Fu	3
球	Ball	Qiu	2
桶	Barrel	Tong	3
篮	Basket	Lan	2
熊	Bear	Xiong	2
床	Bed	Chuang	2
蜂	Bee	Feng	1
钟	Bell	Zhong	1
鸟	Bird	Niao	3
船	Boat	Chuan	2
书	Book	Shu	1
靴	Boot	Xue	1
弓	Bow	Gong	1
碗	Bowl	Wan	3
盒	Box	He	2
桥	Bridge	Qiao	2
刷	Brush	Shua	1
蝶	Butterfly	Die	2
炮	Cannon	Pao	4
帽	Cap	Mao	4
糖	Caramel	Tang	2
猫	Cat	Mao	1
虫	Caterpillar	Chong	2
链	Chain	Lian	4
椅	Chair	Yi	3
烟	Cigarette	Yan	1

Picture names in Chinese	Picture names in English	Picture names' pinyin (phonetics)	Picture names' tone
云	Cloud	Yun	2
牛	Cow	Niu	2
蟹	Crab	Xie	4
杯	Cup	Bei	1
鹿	Deer	Lu	4
狗	Dog	Gou	3
门	Door	Men	2
柜	Dresser	Gui	4
钻	Drill	Zuan	4
鼓	Drum	Gu	3
鸭	Duck	Ya	1
鹰	Eagle	Ying	1
耳	Ear	Er	3
象	Elephant	Xiang	4
眼	Eye	Yan	3
鱼	Fish	Yu	2
旗	Flag	Qi	2
花	Flower	Hua	1
笛	Flute	Di	2
蝇	Fly	Ying	2
脚	Foot	Jiao	3
叉	Fork	Cha	4
蛙	Frog	Wa	1
锅	Frying pan	Guo	1
羊	Goat	Yang	2
锤	Hammer	Chui	2
手	Hand	Shou	3
马	Horse	Ma	3
房	House	Fang	2
刀	Knife	Dao	1
梯	Ladder	Ti	1
腿	Leg	Tui	3
豹	Leopard	Bao	4
狮	Lion	Shi	1
虾	Lobster	Xia	1
锁	Lock	Suo	3
肺	Lung	Fei	4
枪	Machine gun	Qiang	1
镜	Mirror	Jing	4
猴	Monkey	Hou	2
山	Mountain	Shan	1
鼠	Mouse	Shu	3
钉	Nail	Ding	1
针	Needle	Zhen	1
鼻	Nose	Bi	2
橘	Orange	Ju	2
裤	Pants	Ku	4
桃	Peach	Tao	2
梨	Pear	Li	2
猪	Pig	Zhu	1
钳	Pliers	Qian	2
包	Pocketbook	Bao	1
兔	Rabbit	Tu	4
耙	Rake	Pa	2
鸡	Rooster	Ji	1
尺	Ruler	Chi	3
锯	Saw	Ju	4
剪	Scissors	Jian	3
鞋	Shoe	Xie	2
裙	Skirt	Qun	2
蛇	Snake	She	2
袜	Socks	Wa	4
勺	Spoon	Shao	2
凳	Stool	Deng	4
箱	Suitcase	Xiang	1
燕	Swallow	Yan	4
鹅	Swan	E	2
剑	Sword	Jian	4
桌	Table	Zhuo	1
虎	Tiger	Hu	3
树	Tree	Shu	4
号	Trumpet	Hao	4
龟	Turtle	Gui	1
伞	Umbrella	San	3
琴	Violin	Qin	2
表	Watch	Biao	3
壶	Watering can	Hu	2
井	Well	Jing	3
鞭	Whip	Bian	1
哨	Whistle	Shao	4
窗	Window	Chuang	1
狼	Wolf	Lang	2

(Accepted 6 June 2009)
(Available online 10 June 2009)