Available online at www.sciencedirect.com
SCIENCE DIRECT®
Brain and Language 96 (2006) 37–48
Brain and Language
www.elsevier.com/locate/b&l

Spoken word recognition in context: Evidence from Chinese ERP analyses

Yanni Liu [a,b], Hua Shu [b,*], Jinghan Wei [c]

[a] Department of Psychology, University of Michigan, Ann Arbor, MI, USA
[b] State Key Laboratory for Cognitive Neuroscience and Learning, School of Psychology, Beijing Normal University, Beijing, China
[c] Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, China

Accepted 19 August 2005
Available online 27 September 2005

Abstract

Two event-related potential (ERP) experiments were conducted to investigate spoken word recognition in Chinese and the effect of contextual constraints on this process. In Experiment 1, three kinds of incongruous words were formed by altering the first, second or both syllables of the congruous disyllabic terminal words in high constraint spoken sentences. Results showed an increase of N400 amplitude in all three incongruous word conditions and a delayed N400 effect in the cohort incongruous condition as compared with the rhyme incongruous and plain incongruous condition. In addition, unlike results in English, we found that the N400 effect in the rhyme incongruous condition disappeared earlier than in the plain incongruous condition. In Experiment 2, three kinds of nonwords derived from sentence congruous words were constructed by altering few or many phonetic features of the onset or the whole of the first syllable, and the resulting nonwords appeared as disyllabic terminal forms in either high or low constraint sentences. All three nonword conditions elicited the N400 component. In addition, in high constraint sentences but not in low, the amplitude and duration of the N400 varied as a function of the degree of phonetic mismatch between the terminal nonword and the expected congruous word.

Keywords: Spoken word recognition; Sentence context; ERP analysis; Chinese language

1. Introduction

Everyday, listeners perceive words in streams of speech. To understand the meaning of a single sentence, people not only access the meaning of each word, but also evaluate the grammatical and semantic relationships among words in a sentence. Current models generally agree that spoken word recognition in isolation involves the activation of a set of lexical candidates and the selection of the target words from the activated set although there is less consensus as to whether word-initial similarity is a necessary condition for activation (compare Connine, Blasko, & Titone, 1993; McClelland & Elman, 1986; with Marslen-Wilson, Moss, & van Halen, 1996; Marslen-Wilson & Zwitserlood, 1989). Word recognition in a sentence context additionally requires that the selected word be integrated into a higher-order meaning representation of the sentence context (Tyler & Wessels, 1983; Zwitserlood, 1989). Two central issues in the research on spoken word recognition are the degree of phonetic detail when the speech signal is mapped onto a representation in the mental lexicon, and whether there are conditions under which that information interacts with sentence context.

In the past 30 years, many behavioral studies have investigated the mapping process of a spoken word in isolation. The common experimental method is to introduce mismatches or mispronunciations into the sensory input relative to what is expected and to determine the amount of lexical activation that remains (Frauenfelder, Scholen, & Content, 2001). Spoken word recognition models vary in how well they tolerate initial phonological distortions to the sensory input. The early version of the Cohort model

* Corresponding author. Fax: +8610 5880 0567.
E-mail address: shuh@bnu.edu.cn (H. Shu).

0093-934X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandl.2005.08.007

required a complete match of word-initial information for the speech signal to activate a lexical representation (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987). This information was essential to activate a lexical representation because this portion of the word defined the set of lexical candidates. In essence, the Cohort model showed extreme intolerance and total lexical deactivation when there was mismatch in the initial part of the speech input. By contrast, another major model of spoken word recognition, the TRACE model (McClelland & Elman, 1986), showed relative tolerance and graded activation across positions. That is, the TRACE model could accommodate minor initial mismatches efficiently and easily so as to account for successful recognition of a word even with a distorted input (e.g., "barty" as "party").

The patterns of results obtained by introducing initial phoneme mispronunciations in behavioral experiments depend upon various factors, including the particular experimental paradigm, type of lexical activation tested (e.g., semantic vs. phonological), degree of phonological mismatch, and length of the mispronounced word (Frauenfelder et al., 2001; Spinelli, Segui, & Radeau, 2001). Studies (Connine et al., 1993; Marslen-Wilson & Zwitserlood, 1989) that focused on the semantic activation produced by initially mismatching inputs in a semantic priming paradigm showed significant priming only when the nonword prime was phonologically very close to the target word (one distinctive feature difference). Other studies have evaluated the effect of initial mismatch upon phonological rather than semantic activation. Connine, Titone, Deelman, and Blasko (1997) used the phoneme monitoring task to examine the amount of activation produced by nonwords that varied in their degree of initial mismatch from a target word and showed graded, decreasing detection latencies for nonwords that were phonologically more similar to words. Interestingly, even the nonwords that deviated from the intended words by several (at least five) phonetic features produced some lexical activation relative to the unrelated nonword control.

Phonological mismatch is an effective manipulation to investigate the phonological mapping process of spoken word processing for words in isolation. Spoken word recognition in sentences, however, requires listeners to continuously evaluate multiple sources of information as the input signals accumulate over time (Li, 1996). There is evidence that semantic context plays a role in the on-line recognition of spoken words (Zwitserlood, 1989). However, because interpretation of reaction time results generally entails inferences from measures taken at the end point of processing—the vocal output or the button press—rather than on-line, it is difficult for those measures to reveal the more dynamic aspects of spoken sentence processing. Recent brain imaging techniques are appealing because they provide more on-line measures of cognitive processing. Specifically, event-related potentials (ERPs) provide a temporal resolution on the order of milliseconds, allowing measurements as the process of interest unfolds so that both early and late processes can be monitored. Thus, ERPs provide an ideal method to investigate issues pertaining to both the dimensions of (phonological, semantic) mapping and time-course of spoken word recognition in context.

The pre-eminent ERP measure in the study of language processing and the underlying mechanisms has been a negative component that typically peaks about 400 ms after stimulus onset, the N400 (Kutas & Hillyard, 1980). It is generally accepted that the N400 component is sensitive to semantic integration of words, and arises in both the visual and auditory modalities (Brown, Hagoort, & Chwilla, 2000). The N400 tends to be smaller in amplitude when the elicited word is preceded by a congruous sentence frame (e.g., "He spread the warm bread with *butter*"), or a related single word, than if the eliciting word is incongruous or unrelated to the preceding context (e.g., "He spread the warm bread with *socks*") (Kutas & Hillyard, 1980). Several findings have clearly demonstrated that the N400 amplitude varies as a function of how easily a word can be integrated into a sentence context (King & Kutas, 1995; Kutas, 1997). For example, Connine, Blasko, and Wang (1994) compared ERPs to sentence-final words in high constraint sentence contexts (e.g., "The king wore a golden *crown*") with those in low constraint sentence context (e.g., "The woman talked about the *frogs*"). A negative shift was observed for final words in low constraint sentences relative to final words in high constraint sentences.

A key interest among researchers is the conditions under which sentence context can influence lexical access for spoken words. Several studies have investigated this issue by manipulating the initial phoneme of sentence-final words so that they matched or mismatched the onset of the expected word (Connolly & Phillips, 1994; Connolly, Phillips, Stewart, & Brake, 1992; Connolly, Stewart, & Phillips, 1990; Van den Brink, Brown, & Hagoort, 2001). Connolly and Phillips (1994) manipulated the degree of constraint of sentence context on the final words as well as their mismatch with the (expected) initial phonemes. Under the anomalous final word condition, a late negative peak (the N400) arose when the anomalous final word had the same initial phonemes (e.g., "The gambler had a streak of bad *luggage*") as the expected word (*luck*), but both an early (N200) and a late (N400) negative peak were found for the final word when the onset differed (e.g., "The dog chased the cat up the *queen*") from the expected word (e.g., *tree*). The early effect (N200) was interpreted to reflect acoustic/phonological processing of the sentence-final word, and the N400 amplitude was claimed to be modulated by semantic expectancy.

Using a similar design and a technique that included a pause (500 ms) between final word and sentence frame, Van Petten and her colleagues (1999) examined phonemic analysis in spoken sentence context. Their experimental items consisted of high and low constraint sentences (e.g., "It was a pleasant surprise to find that the car repair bill was only seventeen…") that ended either with (a) a cohort congruous word (e.g., *dollars*), (b) a word that rhymed with the congruous word (rhyme incongruous word, e.g.,

scholars), (c) a word that shared the same initial phonemes as the congruous word (cohort incongruous word, e.g., *dolphins*), or (d) a fully unrelated word (plain incongruous word, e.g., *burdens*). Results revealed that the sentence-final words in all three incongruous conditions elicited a significantly larger N400 than the congruous ending words. Moreover, the onset of the N400 effect differed among the three incongruous conditions. In the plain incongruous and rhyme incongruous conditions, the onset of the N400 effect preceded the identification point of the sentence-final word as determined by the gating procedure (see Grosjean, 1980). Relative to these two conditions, the onset of the N400 effect in the cohort incongruous condition was found to be delayed by some 200 ms. The outcome suggested that the onset of the N400 effect mirrored the moment at which the acoustic input first diverged from the semantically defined expectation. In this study, Van Petten, Coulson, Rubin, Plante, and Parks (1999) concluded that the N400 effect reflected semantic processing of the auditory signals and that semantic processing could begin on partial and incomplete phonological information about words.

The ERP technique as applied to the experimental paradigm used by Van Petten and her colleagues (1999) could be considered a manipulation on the position of the mismatch between the presented terminal word and the expected word in spoken sentence context. Their results are compatible with the Cohort model, because the initial mismatch phonemes (rhyme incongruous and plain incongruous) showed an N400 effect different from that of the initial match phonemes (cohort incongruous) and there was no difference between the two initial mismatch conditions. The present study employed the ERP methodology and exploited a special property of the Chinese language to evaluate tolerance in spoken word recognition, to position and then degrees of phonological distortion of words in contexts.

The structure of the Chinese language differs from that of English in several potentially informative ways. Perhaps most important, the Chinese language consists of a limited set of, approximately 1200 syllables. A syllable in Chinese consists of a tone as well as an onset and a rhyme. More important, syllable structure in Chinese is restricted to only a few patterns (CV, CCV, CVC, and CCVC), and most words consist of only one or two syllables. As a result, in Chinese, many words differ by a single phoneme. For example, many syllables or words, such as /shao/, /she/, /shi/ and /shou/, differ from /sha/ only at the final phoneme. Similarly, the syllables or words, such as /cha/, /da/, /ba/, and /la/ differ from /sha/ only at the initial phoneme. Given the extensive similarity among words, it is possible to create experimental materials where two-syllable word pairs differ by a single feature or a phoneme.

Finally, 70% of Chinese words were two-syllable compound words. Each syllable in a compound word is phonologically and semantically defined, and usually corresponds to a morpheme and a character. For example, the two-syllable compound word */dian4/-/yuan2/* corresponds to the written word 电源, and means *electric power*. In Chinese, it is possible to find words like 水源 (means *water source*) that differ with respect to the first character (and syllable) but share the second character. Similarly, there are also words like 电池 (means *battery*) that share the first character (and syllable) but differ with respect to the second.

In the present study, we investigated spoken word recognition by recording brain activity while participants listen to sentences. In particular, we focused on the mapping process for sentence terminal words in spoken Chinese sentences and evaluated the influence of sentence context. In Experiment 1, we adapted the experimental paradigm of Van Petten et al. (1999), to the Chinese language so as to establish some basic knowledge about the ERP waveform of spoken word processing in Chinese sentence contexts. All of the sentence materials ended with two-syllable words. Crucially, the position of mismatch was manipulated by altering the first, second or both syllables of the sentence terminal words. All words appeared in congruous and high constraint sentence contexts. In Experiment 2, the degree of mismatch between the terminal word and sentence context was manipulated by altering few or many phonetic features of the onset or the whole first syllable of the terminal words that appeared in high or low constraint sentences. That is, we varied sentence constraint to evaluate the interaction between input information and sentence context.

As a whole, the present study was designed to investigate the effect of phonological mismatch (mismatch between syllables and mismatch at the level of onsets within a syllable), contextual constraint and the interaction between them using Chinese materials. The special contribution afforded by the Chinese language derives from the prevalence of one-syllable words, the high degree of homophony among those words and the tendency to build compound words from those homophonic syllables. As a result the particular combination of syllables may have special importance because it helps to resolve homophony in the course of auditory recognition. The more specific questions we addressed are: (1) Is the N400 shift sensitive to the *position* of mismatch between the presented disyllabic terminal word and the expected (congruous) one in Chinese sentence contexts? A difference has been observed in English, but it has been difficult to demonstrate any N400 activation when mismatch is early and match is later in the word. (2) Is the negative shift sensitive to the number of phonetic features that mismatch in the first syllable of a two-syllable terminal word in a Chinese sentence? (3) Does mismatch interact with the constraints imposed by a sentential context? The unique contribution of the ERP evidence provided by Experiment 1 derives from the potential to introduce lexically valid alternatives that vary systematically from the congruous word with respect to position of phonological mismatch.

2. Experiment 1

We examined the effect on the N400 of phonological mismatch in the first or second syllable of an auditorily presented two-syllable Chinese word presented in a sentence context. We used the extensive homophony that exists in Chinese so that syllabic substitutions to the word that is congruous with the sentence context create other legal two-syllable words.

2.1. Method

2.1.1. Materials

Experimental materials consisted of spoken sentence frames (e.g., *"The sound in the radio became weaker and weaker. It seems that I must buy several new sets of..."*) paired with terminal words that formed semantically congruous or incongruous endings. There were four critical conditions. (1) The congruous sentence condition ended with cohort congruous words that fit with the sentence frame (e.g., */dian4/-/chi2/*, *battery)*. In addition, there were three types of incongruous terminal words: (2) the cohort incongruous condition, in which the first syllable of the final word was same as that of the cohort congruous word, but the second syllable was different (e.g. */dian4/-/lu2/*, *electric stove*); (3) the rhyme incongruous condition, in which the second syllable was same as the cohort congruous word, but the first syllable was different (e.g. */shui3/-/chi2/*, *water pool*); (4) the plain incongruous condition, in which both syllables were different from the cohort congruous word (e.g. */bing4/-/tai4/*, *illness*). The terminal words in the three incongruous conditions were balanced in terms of the number of homophones of the first syllable, as well as the number of words including the first syllable. To establish the degree of constraint of the sentence frames, a group of 18 college students, who did not participate in the following ERP experiments, were asked to rate the predictability of 248 sentence frames for the terminal word completions using a seven-point scale. A set of 192 sentences with high predictability (Mean = 6.33, $SD = 0.48$) was selected as the stimuli in Experiment 1.

The sentence frames were spoken and recorded without their terminal words and were then digitized and edited to yield one audio file for each sentence frame. Sentence frames averaged 19.5 characters and 4 s in duration. The terminal words were recorded separately and each word was edited to generate one audio file. The duration of the terminal words (first syllables) in the congruous condition, cohort incongruous condition, rhyme incongruous condition and plain incongruous condition were respectively: 566 (285), 579 (294), 570 (290), and 721(367) ms. [1]

Four versions of the stimuli were presented to 4 groups of subjects and were completely counterbalanced, so that each subject heard each sentence frame with one of the terminal words. Across subjects, each sentence frame was paired with a cohort congruous word, a cohort incongruous word, a rhyme incongruous word, and a plain incongruous word. In order to balance the number of congruous and incongruous sentence presented during a session, filler sentences were added and all of them had congruous terminal words. Accordingly, for each subject, the stimulus list included one version of each of the 192 critical sentences (of which one fourth were congruous) and 105 filler sentences all of which were congruous to produce a total of 153 congruous sentences. The 297 sentences were split into three blocks. Participants viewed 14 practice sentences (8 for the critical sentences equally distributed among 4 categories and 6 for filler sentences with congruous terminal words).

2.1.2. Participants

Sixteen subjects (7 male; all right-handed) served as paid volunteers. All were native Chinese speakers who reported normal hearing, normal or corrected to normal vision, and no history of neurological disorder. Their age varied from 17 to 25.

2.1.3. Procedure

Each individual participated in three blocks lasting about 18 min each. In the experiment, the subjects were tested in a dimly illuminated sound-attenuating booth. They were seated comfortably and instructed to move as little as possible. A fixation point was presented at the center of the computer screen at all times. Subjects were required to keep their eyes fixated on that point. Each trial consisted of an auditory sentence frame, 500 ms of silence, and an auditory terminal word, followed by a visual target presented 2 s after the offset of the terminal spoken word. The inter-trial interval was 2 s. The task for the participants was to listen to the auditory sentence with full attention and judge whether the visual target word had appeared in the sentence that they had just heard. They indicated their response by pressing the "yes" or "no" response-key. Among all sentences trials, there were 147 sentences in which the visual target appeared in the sentence frame. In the remaining 150 sentence trials, the visual target did not appear in the sentence. Instructions for the judgment task encouraged the participants to listen carefully to the auditory sentences. The results from the judgment task were not recorded.

2.1.4. Data acquisition and analysis

NeuroScan ERP Workstation software (made in USA) was used in the present experiment. Electroencephalographic (EEG) data were recorded using a 32-channel Quick-cap with tin electrodes, referenced to the link of the left and right mastoids. A careful skin preparation and an electrode gel produced impedance of less than 5 kΩ. From bipolar montages, the lateral vertical electrooculogram (VEOG) was recorded by electrodes above and below the left eye, to detect blink artifacts. A band pass of

[1] With the exception of the plain incongruous condition, the durations of terminal words in the different conditions did not differ.

0.05–100 Hz was used to continuously digitize the recording at a sampling rate of 500 Hz. EOG artifact was automatically corrected by Neuro Scan software (Semlitsch, Anderer, Schuster, & Preslich, 1986) and other artifacts (exceeding 80 μV) were rejected off-line. The epoch was 1100 ms including 100 ms baseline.

2.2. Results and discussion

From the grand mean ERP, N1, P2, and N400 components were observed in the present experiment (Fig. 1), an outcome which is consistent with the results of Van Petten et al. (1999). The scalp distribution of N1 was extensive, whereas P2 was only apparent in the anterior electrodes. These two components are typical auditory ERP responses to stimuli with an abrupt acoustic onset. In the previous experiments, which used continuous speech materials, the early N1 and P2 were much reduced because clear physical boundaries between the words of the sentences are lacking. In Experiment 1, the primary interest was in the N400 including its sensitivity to sentence congruity and phonological mismatch as a function of position. The N400 component observed in Experiment 1 had a slow negative shift and an extensive scalp distribution. Congruent with other reports, the N400 was largest at the central and parietal sites. Based upon a preliminary analysis, electrode site Pz showed the largest N400 effect. Further statistical analyses focused on the waveform recorded from the channel Pz (Fig. 2). First, the peak latency of N400 was measured. Then we used the peak latency to define a temporal region extending 150 ms forward and backward respectively. We computed the average amplitude of this 300 ms time-window as a function of the amplitude of N400 in each condition. Furthermore, we used time-window methods to compare the differences of N400 effects in the three incongruous conditions. That is, we examined the difference between ERP waveforms in the cohort congruous and each incongruous condition in each 100 ms window—extending from 200 to 800 ms.

2.2.1. Statistical analysis

2.2.1.1. N400 peak latency. The main effect of congruity condition was significant, [$F(3,45) = 31.104$, $p < .001$] (Table 1). Post hoc comparisons with a Newman–Keuls test showed that the peak latency of the N400 in the cohort incongruous condition was later than in the other three conditions ($p < .01$). The plain incongruous condition was later than either the cohort congruous or the rhyme incongruous condition ($p < .05$).

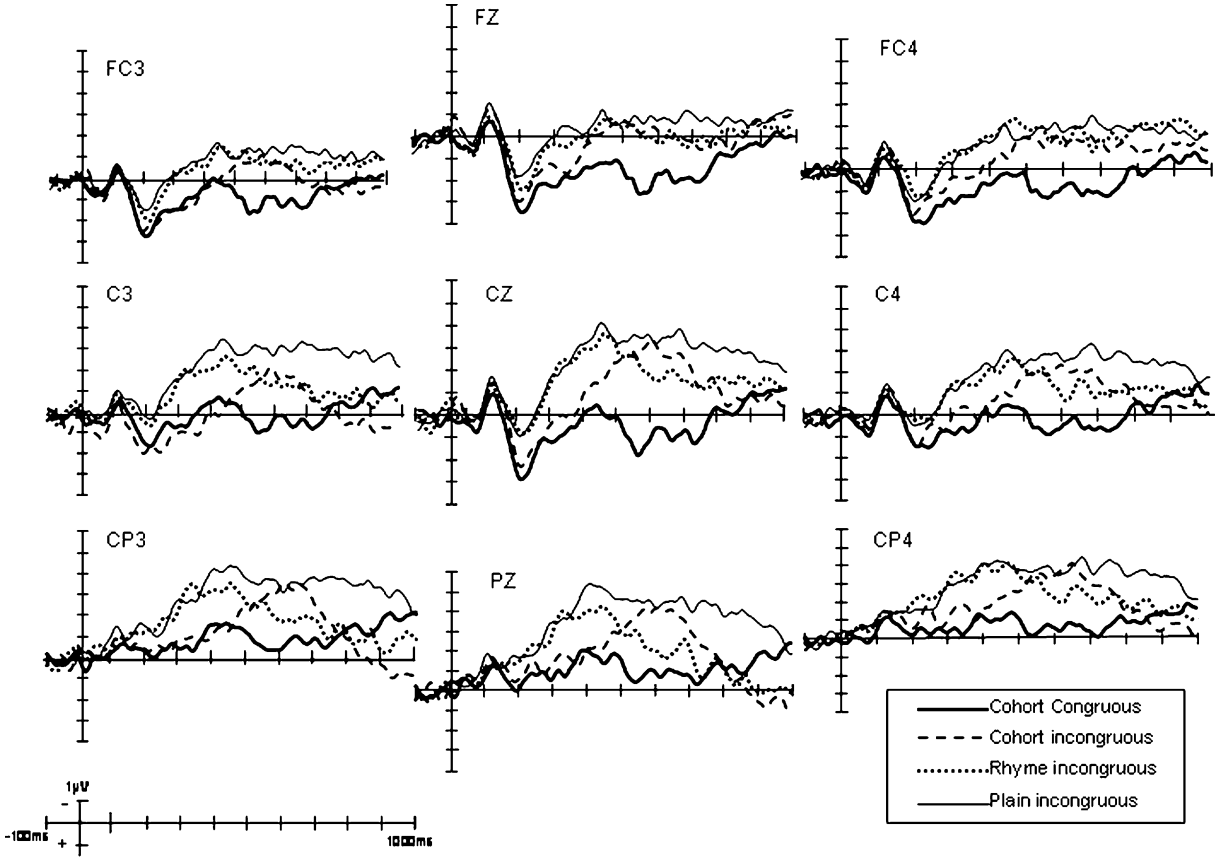

Fig. 1. Grand mean ERP of the four conditions in Experiment 1. Nine electrodes were selected as the representative. Time 0 is the onset of the spoken word. The bold solid line denotes the cohort congruous condition, e.g., */dian4 chi2/*; the dashed line denotes the cohort incongruous condition, e.g., */dian4 lu2/*; the dotted line denotes the rhyme incongruous condition, e.g., /shui3 chi2/; and the light solid line denotes the plain incongruous condition, e.g., */bing4 tai4/*.

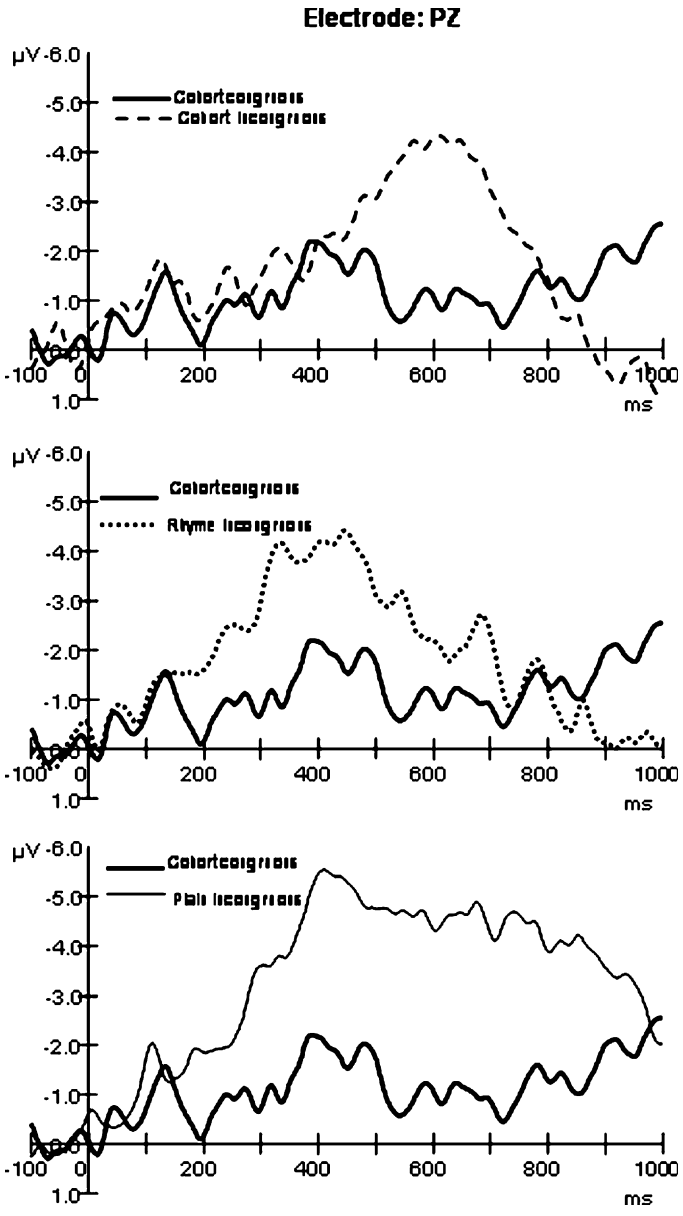

Fig. 2. Grand mean ERP of the four conditions at Pz electrode site in Experiment 1. Time 0 is the onset of the spoken word. Top: cohort congruous condition vs. cohort incongruous condition (dashed line, e.g., */dian4 lu2/*). Middle: cohort congruous condition vs. rhyme incongruous condition (dotted line, e.g., */shui3 chi2/*). Bottom: cohort congruous condition (bold solid line, e.g., */dian4 chi2/*) vs. plain incongruous condition (light solid line, e.g., */bing4 tai4/*).

2.2.1.2. N400 amplitude. The main effect of congruity was significant, [$F(3,45) = 9.07$, $p < .001$]. Post hoc comparisons with a Newman–Keuls test found a difference between the congruous and the three incongruous conditions ($p < .05$), but there was no significant difference among the three incongruous conditions ($p > .05$).

2.2.1.3. Time-window. The mean amplitude measure of 100 ms epochs was taken starting 200 ms after the word onset and extending to 800 ms. Table 2 shows that in the rhyme incongruous and plain incongruous conditions, the congruity effect, or the N400 effect, emerged from the 200-ms window, and extended to the 500-ms window in the rhyme incongruous condition, and at least to the 700-ms window in the plain incongruous condition. However, the N400 effect began later in the cohort incongruous condition. It emerged at the 500-ms window and only extended to the 600-ms window.

Consistent with other reports, the results of Experiment 1 showed that the three incongruous sentence completions elicited much larger N400 amplitudes than did the congruent sentence completions. This finding suggests that the integration of the incongruous completions met with difficulties. There was no difference in N400 amplitude among the three incongruous completions although they differed in terms of the latencies and durations of the N400 effect.

In the cohort incongruous condition, the N400 effect was delayed by some 300 ms as compared with the rhyme incongruous and plain incongruous conditions, while its duration was shorter than those in the other two incongruous conditions. One possible explanation for the late N400 is that the cohort incongruous words shared the first syllable with the expected words, so that the negative shift did not start until the mismatching second syllable came. The outcome suggests that ERPs are sensitive to a sequential property of the auditory signal. The shorter duration of N400 in the cohort incongruous condition compared with the rhyme incongruous and plain incongruous conditions can be explained as evidence that the early processing of spoken words with matching first syllables in context met with less difficulty in integration than did those with mismatching initial syllable, and that context effects begin very early in lexical processing—during the initial syllable of the critical word. In isolation, the pattern in the cohort incongruous condition in Chinese spoken word processing supports the Cohort model developed from English.

In the rhyme incongruous condition, the N400 effect began at the 200 ms time-window, but disappeared after the 500 ms time-window, a time course that is much shorter than that in the plain incongruous condition. This pattern is different from that of Van Petten and her colleagues' experiments in English (1999) who observed no difference between the rhyme incongruous condition and the plain incongruous condition. The early disappearance of the N400 effect and the decreased difficulty in integration caused by the initial mismatch in our plain incongruous condition may be related to the semantic properties of the matching final syllable. We proceed with caution, however,

Table 1
The amplitude and peak latency of N400 in Experiment 1 at Pz electrode site

	Cohort congruous	Cohort incongruous	Rhyme incongruous	Plain incongruous
Amplitude (μV)	−1.38	−3.53	−3.57	−4.70
Latency (ms)	407	596	415	466

Table 2
Time course of the sentence congruity effects in Experiment 1

Latency window (ms)	Cohort incongruous vs. cohort congruous		Rhyme incongruous vs. cohort congruous		Plain incongruous vs. cohort congruous	
	$F(1,15)$	p Value	$F(1,15)$	p Value	$F(1,15)$	p Value
200–300	0.419	>.05	6.898	.019*	8.565	.010*
300–400	0.390	>.05	11.860	.004*	12.404	.003*
400–500	1.167	>.05	8.444	.011*	15.111	.001*
500–600	10.432	.006*	8.307	.011*	18.838	.001*
600–700	9.655	.007*	3.720	>.05	19.093	.001*
700–800	3.114	>.05	0.458	>.05	22.634	.000*

* Means $p < .05$.

because in the plain incongruous condition, the N400 effect was evident in the 200 ms time-window, as in the rhyme incongruous condition, although the peak latency was relatively later. We suspect that the delayed peak latency might simply reflect the longer spoken word durations of items in the plain incongruous condition. The terminal words were about 150 ms longer than those in the other three conditions (about 80 ms longer for the first syllable).

In Experiment 1, we found that (a) the auditory system registered the mismatch information quickly; (b) the ERP method was very sensitive to the sequential aspects of the auditory signal; and that (c) context effects on lexical processing of terminal word began early. These findings are consistent with those in the English language literature. More interestingly, we also found that there were some differences between the pattern and time course of the rhyme incongruous condition and the plain incongruous condition. These results differ from those of previous English studies. We reserve discussion until after Experiment 2.

3. Experiment 2

In Experiment 1, we replicated several of the basic findings about ERP waveform for spoken word recognition in Chinese sentence context and demonstrated that the ERP paradigm can be used reliably to investigate Chinese spoken word recognition. In Experiment 2, we continued to explore the mapping process by manipulating the degree of mismatch between the presented auditory sound stimulus and the sentence congruous word. At the same time, we examined the role of sentence context on the mapping process by manipulating the degree of constraint.

3.1. Method

3.1.1. Materials

The stimulus materials consisted of spoken sentence frames paired with congruous terminal words and incongruous nonwords.

A separate group of 90 college students participated in a cloze pretest to establish the constraint information from sentence frames of potential congruous terminal words. Each individual saw 100 sentence frames without terminal words, and was directed to complete the frame. Each sentence was completed by a minimum of 22 participants. According to the answers provided by the participants, an appropriate terminal word for each sentence frame was selected. From the cloze probability of the terminal word and the number of cloze words supplied by participants, sentence frames were divided into high (cloze probability: 84.8%; number of different cloze completions: 2.24) and low (cloze probability: 10.6%; number of cloze completions: 9.78) constraint conditions. As in Experiment 1, sentence frames were spoken and recorded without their terminal words and were then digitized and edited to yield one audio file for each sentence frame.

Nonwords were generated by altering either the onset of the first syllable or the entire first syllable of each congruous terminal word. These incongruous nonwords fell into three conditions: (a) minimal-onset-mismatch nonwords (e.g., */chao1 hu1/*), generated by altering one or two linguistic features of the onset of the first syllable in the congruous words (e.g., */zhao1 hu1/*, "*greeting*"). The phonemes /zh/ and /ch/ differ only on one feature (aspiration: aspirated and nonaspirated respectively). (b) Maximal-onset-mismatch nonwords (e.g., */lao1 hu1/*), generated by altering two or more linguistic features of the onset, so as to become less similar to the congruous condition. /zh/ and /l/ are different on at least two features (manner of articulation: affricate and lateral respectively; place of articulation: back and central respectively) (Li & MacWhinney, 2002; Xing, Shu, & Li, 2002). (c) First-syllable-mismatch nonwords (e.g., */xing1 hu1/*), generated by altering the first syllable, so that among nonwords they have the least similarity with the congruous completion. Sentences averaged 19.1 characters and 4.4 s in duration. Terminal words and nonwords were recorded separately and their durations in each condition were approximately 588 ms.

The set of critical stimuli consisted of 320 sentences, of which half were high constraint sentence frames (e.g., *Today he is in happy mood, almost to every one he met give his...*) and half were of low constraint (e.g., *You'd better wait in the living room, because there are some...*). Four stimulus lists were constructed so that no participant heard a sentence frame or a terminal word more than once. Across subjects, each sentence frame was paired with its congruous word, its minimal-onset-mismatch nonword, its maximal-onset-mismatch nonword, and its

first-syllable-mismatch nonword. To balance the incidence of congruous and incongruous sentences, 80 filler sentences were constructed, in which sentence frames and final words were always congruous. All stimulus lists included the same set of 80 filler congruous sentences. In addition 10 sentences (4 for congruous sentences, and 6 for sentences ending with mispronounced terminal stimuli) were constructed as the practice items.

3.1.2. Participant

Sixteen subjects (7 male; all right-handed) served as paid volunteers. All were native Chinese speakers who reported normal hearing, normal or corrected to normal vision, and no history of neurological disorder. They did not participate in either Experiment 1 or in the cloze pretest. Their age varied from 17 to 28.

3.1.3. Procedure

Each individual participated in four blocks lasting about 18 min each. The procedure and the task were similar to that of Experiment 1. Half of the visual targets appeared in a congruous sentence frame and half in an incongruous frame.

3.1.4. Data acquisition

Electroencephalographic data were recorded using a 64-channel Quick-cap with tin electrodes, referenced to the right mastoid. Left mastoid data were recorded as a separate channel, and the data were referenced to an average mastoid for analysis. Other operations were the same as Experiment 1.

3.2. Results and discussion

As in Experiment 1, Experiment 2 produced clear N1, P2 and N400 components in four kinds of terminal stimuli (Fig. 3). The N400 elicited by the terminal stimuli in this experiment also showed a slow negative shift, which began at around 200 ms and was largest in the central-parietal area. Consistent with Experiment 1, electrode Pz was selected for the statistical analysis. Based upon the grand mean waveform (Fig. 4), the mean amplitude from 300 to 600 ms after the onset of the auditory completions was defined as the amplitude of the N400 from which to examine the difference across conditions. In addition, we used the time-window method to assess differences of onset and

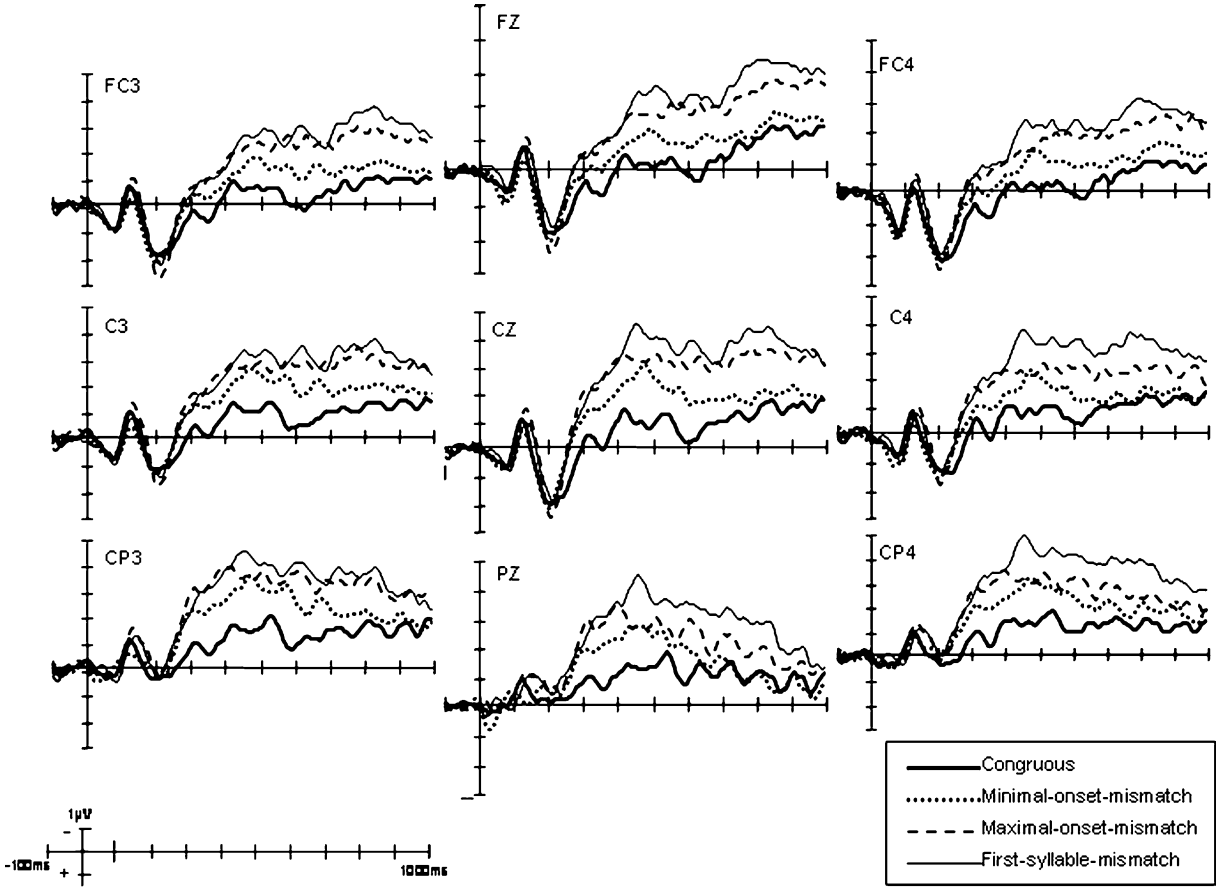

Fig. 3. Grand mean ERP of the four conditions in Experiment 2. As in Experiment 1, nine electrodes were selected as the representative. Time 0 is the onset of the spoken word. The bold solid line denotes the congruous condition, e.g., */zhao1 hu1/*; the dotted line denotes the minimal-onset- mismatch condition, e.g., */chao1 hu1/*; the dashed line denotes the maximal-onset-mismatch condition, e.g., */lao1 hu1/*; and the light solid line denotes the first-syllable-mismatch condition, e.g., */xing1 hu1/*.

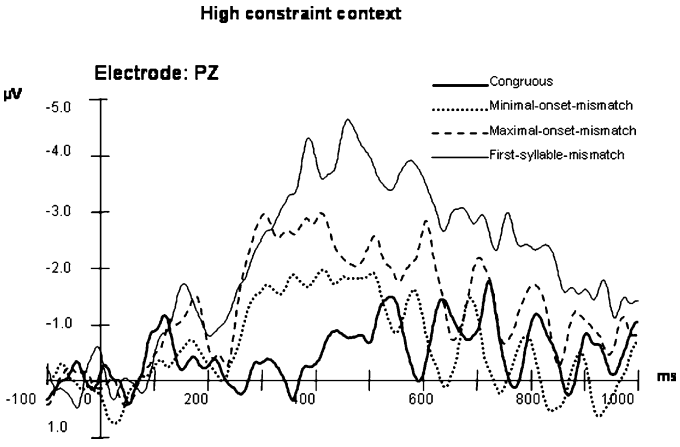

Fig. 4. Grand mean ERP of the four conditions at Pz electrode site in Experiment 2. The waveforms just for the high constraint sentences. Time 0 is the onset of the spoken word. The bold solid line denotes the congruous condition, e.g., */zhao1 hu1/*; the dotted line denotes the minimal-onset-mismatch condition, e.g., */chao1 hu1/*; the dashed line denotes the maximal-onset-mismatch condition, e.g., */lao1 hu1/*; and the light solid line denotes the first-syllable-mismatch condition, e.g., */xing1 hu1/*.

duration of the N400 effect among the three incongruous conditions. In the ANOVA results, F tests with more than one degree of freedom in the numerator were adjusted by means of the Greenhouse–Geisser correction.

3.2.1. Statistical analysis

3.2.1.1. N400 amplitude. A 2 (sentence constraint) by 4 (types of terminal words) ANOVA was conducted, showing main effects of both sentence constraint, [$F(1,15) = 7.92$, $p < .05$], and types of terminal stimuli, [$F(3,45) = 15.88$, $p < .001$]. The interaction of the two factors was marginally significant, [$F(3,45) = 2{,}77$, $p = .059$]. Post hoc comparisons with a Newman–Keulus test showed that all three incongruous conditions elicited increased N400 compared with the congruous condition. To further explore the effect of sentence constraint, we performed an omnibus ANOVA taking sentence constraint and type of incongruous terminal stimuli as the two within-subject factors.[2] There was a marginally statistically significant main effect of sentence constraint, [$F(1,15) = 3.98$, $p = .064$]; as well as a main effect of type of incongruous terminal stimuli, [$F(3,45) = 6.20$, $p < .05$]. The interaction of the two factors was also significant, [$F(3,45) = 3.93$, $p < .05$]. Further a simple-effect test showed that the N400 was the same size for the first-syllable-mismatch in both the high and low constraint sentences ($F < 1$). In the other two conditions they differed (Minimal-onset-mismatch: $F(1,15) = 3.27$, $p < .1$; maximal-onset- mismatch: $F(1,15) = 8.03$, $p < .05$). There is a graded effect in high constraint context ($F(2,30) = 10.63$, $p < .001$) and equally large N400s for all the incongruous endings in the low constraint context ($F < 1$) (see Table 3).

3.2.1.2. Time-window. As in Experiment 1, successive mean amplitude measures with a 100 ms epoch starting at 200 ms post-target and extending to 800 ms revealed the influence of sentence constraint and terminal stimuli types. In a *Condition (4) × constraint (2) × Window (6)* ANOVA, the 3-way interaction was marginally significant [$F(15,225) = 13.303$, $p = .071$], and the 2-way interaction of *condition (4) × window (6)* [$F(15,225) = 7.455$, $p < .001$] and *constraint (2) × window (6)* were both significant [$F(5,75) = 13.303$, $p < .001$]. Then the N400 effects in the three incongruous conditions were examined both in high and low constraint sentences at each separate 100 ms time-window from 200 to 800 ms. Table 4 shows that in high constraint sentences, N400 effect in the minimal-onset-mismatch condition was only significant in the 300 ms time-window; whereas N400 effect in the maximal-onset-mismatch and first-syllable-mismatch conditions both began from the 200 ms time-window, and were extended to 500 ms and 700 ms time-window respectively. However, in low constraint sentences, the congruity effect of the two onset-mismatch completions existed from the 300 ms to the 600 ms time-window, while syllable-mismatch condition extended the difference at least to the 700 ms time-window.

In Experiment 2, all three incongruous conditions elicited increased N400 compared with the congruous condition. This finding suggested that the auditory system could recognize the mismatch information in the smaller phonetic unit (onset), and that this mismatch caused difficulties in semantic integration. At the same time, the significant main effect of sentence constraint demonstrated the influence of sentence context on spoken word processing.

In the high constraint sentences, the amplitude of the N400 was decreased in the two onset-mismatch conditions as compared with the syllable-mismatch condition. This finding indicated that the nonword input facilitated the potential lexical candidates sharing the same rhyme and decreased the integrative difficulty. While there was no significant difference in the amplitude of N400 between the two onset-mismatch conditions, in the time-window analysis, the minimal-onset-mismatch showed a significantly shorter duration for the N400 effect than did the maximal-onset-mismatch. It suggested that the mismatch input with a minimal initial difference could activate the target word, and be integrated into the context more easily than the maximal-onset-mismatch input. The latter needed more affiliated information, and integration required a longer time.

In low constraint sentences, by contrast, few differences were apparent between the three incongruous nonword completions either in the amplitude or the duration of the N400 effect. The different pattern of results in the two kinds of sentence contexts demonstrates the interaction of phonetic information with context.

[2] In this analysis, the congruent condition is not included because cloze probability is confounded with sentence constraint. That is, in our experimental materials, the cloze probability is high for the "high constraint" congruent completion and low for the "low constraint" congruent condition.

Table 3
The amplitude of N400 in high and low constraint sentences in Experiment 2 (μV)

Sentence constraint	Congruous	Minimal-onset-mismatch	Maximal-onset- mismatch	First-syllable- mismatch
High	−0.57	−1.62	−2.42	−3.68
Low	−1.84	−3.10	−3.32	−3.49

Table 4
Time course of the sentence congruity effects in Experiment 2

Latency window (ms)	Minimal-onset vs. congruous		Maximal-onset vs. congruous		First-syllable vs. congruous		Incongruous vs. congruous	
	High	Low	High	Low	High	Low	High	Low
200–300	>0.05	>0.05	0.044*	>0.05	0.008*	>0.05	0.007*	>0.05
300–400	0.008*	0.016*	0.001*	0.004*	0.000*	0.007*	0.000*	0.001*
400–500	>0.05	0.001*	0.009*	0.013*	0.000*	0.008*	0.000*	0.000*
500–600	>0.05	0.008*	0.025*	0.018*	0.000*	0.006*	0.005*	0.000*
600–700	>0.05	0.036*	>0.05	0.013*	0.001*	0.004*	0.007*	0.000*
700–800	>0.05	>0.05	>0.05	>0.05	0.005*	0.017*	>0.05	0.003*

* Means $p < .05$.

In this experiment some constraints on auditory processing were observed, in particular the influence of lexical similarity and contextual constraint. We found that (a) the nonword completions of different degrees of mismatch with congruous words produced graded activation of the lexical representation, and (b) context served to increase the activation of the lexical representation and to facilitate the integration process.

4. General discussion

We investigated the effect of mismatch between initial or final syllables (Experiment 1) and mismatch at the level of onsets within a syllable (Experiment 2) in auditory word recognition with Chinese materials where, due to extensive homophonic among syllables and the prevalence of compound words composed of homophonic syllables the particular sequence of syllables may have special importance.

In Experiment 1, we manipulated the mismatch position of a phoneme within a disyllabic word presented in a sentence context, and found that incongruous sentence completions elicited a larger N400 component than did congruous completions, and that the effect of incongruity on waveforms began about 200 ms after the word onset. When we examined the position of mismatch, we observed that relative to the rhyme incongruous condition, the incongruity effect in the cohort incongruous condition was delayed by some 200 or 300 ms; as it appeared to start at the point when the acoustic input began to diverge from the expected word. In some respects, the present results in Chinese are consistent with the results in English by Van Petten et al. (1999) and others (Connolly & Phillips, 1994; Connolly, Phillips, & Forbes, 1995; Connolly et al., 1992, 1990). Evidently, in Chinese as well as in English, listeners form an expectation based on sentence context, and the N400 elicited by an incongruous word reflects the mismatch with this expectation (Van Petten et al., 1999). Earlier appearance of the N400 effect in the rhyme incongruous and plain incongruous conditions relative to the cohort incongruous condition, along with the shorter duration time of the N400 effect in the cohort incongruous condition are worthy of further discussion.

The time-window analysis revealed that the N400 difference between the rhyme incongruous and the cohort congruous condition disappeared about 200 ms earlier than that between the plain incongruous and the cohort congruous condition. In essence, the rhyme incongruous words elicited a shorter negative wave than the totally incongruous sentence completions. Although a difference between the congruity effect in the rhyme incongruous condition and in the plain incongruous condition has not been documented previously with the ERP measure in either Chinese or English, it is consistent with a pattern that can arise with an eye-tracking methodology (Allopenna, Magnuson, & Tanenhaus, 1998). In those experiments, subjects were asked to move one of the objects following a spoken instruction (e.g., Pick up the *beaker*, now put it below the diamond) and four conditions were compared: the target (*e.g., beaker*), a cohort distractor (e.g., *beetle*), a rhyme distractor (e.g., *speaker*) and an unrelated distractor (e.g., *carriage*). The results showed that subjects began to fixate on the target and cohort objects more often than on unrelated ones beginning about 200 ms after the onset of the target word, with the target diverging from the cohort at about 400 ms. Importantly, they also documented fixation to rhyme objects that began shortly later (about 300 ms). While fixation to the cohort began earlier and had a higher peak than that of the rhyme, like ERP in Chinese, eye-tracking results showed that both cohort and rhyme distractors competed for lexical activation.

Our results also contrast with an English result in a potentially revealing way. Specifically, Van Petten and her colleagues (1999) reported that the waveforms of the rhyme incongruous and the plain incongruous condition were almost overlapping in both amplitude and duration. This was not the case in our study. A likely explanation reflects characteristics of the Chinese language. In contrast to English where, with the exception of compounds, the

English syllables in a multiple syllable word (e.g., *dollar*) are usually only phonological units and are devoid of meaning (because they are not morphemes), in Mandarin Chinese, a syllable typically corresponds to one (or more) character(s) and morpheme(s). For example, the syllables in the disyllabic word (e.g., */dian4/-/chi2/*), battery are also morphemic units (e.g., *electric-pool*). To construct the materials of the rhyme incongruous and the cohort incongruous condition in Experiment 1, we changed one of the two syllables in the critical cohort congruous word and preserved the other syllable. In fact, post hoc analysis showed that when syllables matched, they shared the same characters (morpheme) about 70% of the time. For example, the critical word "电池 */dian4/-/chi2/*, *battery* or *electric-pool*" was changed to "水池 */shui3/-/chi2/*, *water-pool*" (rhyme incongruous condition) and "电炉 */dian4/-/lu2/*, *electric-stove*" (cohort incongruous condition), so that the matching syllables shared the same characters and morphemes. Therefore, we posit that the basis for the rhyming effect in Chinese that we observed in Experiment 1 is the influence of a shared morpheme. To elaborate, the preserved morpheme of the final syllable may provide a context that permits lexical candidates that do not begin with the same segments to become activated. The preserved morphemic context allowed the late waveform of the rhyme incongruous condition to diverge from that of the plain incongruous condition so that the N400 disappeared. This evidence of late as well as early N400 activation is more consistent with the TRACE than the Cohort model.

In the second experiment of the present ERP study, our primary finding was that in high but not in the low constraint contexts, the N400 varied as a function of the degree of phonological match (or mismatch). The amplitude of the N400 gradually increased across the minimal-onset-mismatch, the maximal-onset-mismatch and the first-syllable-mismatch conditions in the high constraint sentence contexts. Also the duration of the N400 tended to be shorter in the minimal-onset-mismatch than in the maximal-onset-mismatch condition, and shorter in both of the onset-mismatch conditions than in the first-syllable-mismatch condition. The pattern indicates that the duration of the N400 effect can be sensitive to the degree of phonetic mismatch between word predicted by the sentence context and the word-initial information.

It is often suggested that the N400 reflects the interruption of ongoing sentence processing by a semantically inappropriate word (Kutas & Hillyard, 1980). In English, few ERP studies have attempted to investigate this problem by including nonwords. However, in a behavioral study with a cross-modal priming task (Connine et al., 1993), the magnitude of priming also depended on the degree of similarity between a base word and an unrelated stimulus. The graded N400 effect, in our second auditory ERP experiment using nonwords in critical conditions, demonstrated a similar effect. Not only were there N400 responses to a nonword but also, in a highly constrained sentence context, the degree of similarity or goodness of fit of elements in the input to the anticipated spoken word recognition was important. Evidently, the goodness-of-fit of the acoustic–phonetic information of a spoken nonword presented in a sentence context can influence processing.

Crucially, phonological mismatch of nonwords to words depended on degree of sentence constraint. The results of Experiment 2 suggest that while the expectation about the upcoming target word may entail not only semantic, but also phonemic information, context plays an important role in forming those expectations. Only in high constraint sentences did differences in the degree of mismatch produce differences in the amplitude and the duration of the N400 effect. In fact, in the low constraint sentences no differences among the three incongruous conditions could be confirmed other than a robust N400 congruity effect. Evidently, degree of sentence constraint is key. The high constraint contexts influenced the activation level of a specific lexical representation and degree of integration difficulty reflected the degree of mismatch. In the low constraint context there was no intended word or template to guide the match process and the subsequent integration process.

In conclusion, the N400 is usually considered to be sensitive to lexical integration and elicited by a semantic mismatch between the meaning of a word and the semantic specification of its sentence context. The findings that all of the incongruous conditions resulted in a larger N400 than did the congruous conditions in our experiments is consistent with this characterization. In addition, our results showed that, the onset of the N400 diverged earlier in high (200–300 ms window) than in low constraint (300–400 ms window) conditions, and diverged earlier in the maximal-onset-mismatch and first-syllable-mismatch (200–300 ms) than in the minimal-onset-mismatch (300–400 ms) condition. Evidently, at least in a highly constrained context, phonetic and semantic processes can be coupled.

Acknowledgments

This research was supported by National Pandeng Project (95-special-09) and by grants from the Natural Science Foundation of China #60083005 and #60033020. We thank Bill Gehring for discussion of the ERP methodology, thank Julie Boland and Laurie Beth Feldman for their comments.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419–439.

Brown, C. M., Hagoort, P., & Chwilla, D. J. (2000). An event-related potential analysis of visual word priming effects. *Brain and Language, 72*, 158–190.

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition. *Journal of Memory and Language, 32*, 193–210.

Connine, C. M., Blasko, D. G., & Wang, J. (1994). Vertical similarity in spoken word recognition: Mutiple lexical activation, individual differ-

ences, and the role of sentence context. *Perception & Psychophysics, 56*, 624–636.

Connine, C. M., Titone, D., Deelman, T., & Blasko, D. G. (1997). Similarity mapping in spoken word recognition. *Journal of memory and language, 37*, 463–480.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience, 6*, 256–266.

Connolly, J. F., Phillips, N. A., & Forbes, K. A. K. (1995). The effects of phonological and semantic features of sentence-ending words on visual event-related potentials. *Electroencephalography and Clinical Neurophysiology, 94*, 276–287.

Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language, 43*, 1–18.

Connolly, J. F., Stewart, S. H., & Phillips, N. A. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and Language, 39*, 302–318.

Frauenfelder, U. H., Scholen, M., & Content, A. (2001). Bottom-up inhibition in lexical selection: Phonological mismatch effects in spoken word recognition. *Language and cognitive processes, 16*, 583–607.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics, 28*, 267–283.

King, J. W., & Kutas, M. (1995). A brain potential whose latency indexes the length and freauency of words. *Newsletter of the Center for Research in Language, 10*, 3–9.

Kutas, M. (1997). Views on how the electrical activity that the brain generates reflects the function of difference language structure. *Psychophysiology, 34*, 383–398.

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

Li, P. (1996). The temporal structure of spoken sentence comprehension in Chinese. *Perception & Psychophysics, 58*, 571–586.

Li, P., & MacWhinney, B. (2002). PatPho: A phonological pattern generator for neural networks. *Behavior Research Methods, Instruments, and Computers, 34*, 408–415.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. *Cognition, 25*, 71–102.

Marslen-Wilson, W. D., Moss, H., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1376–1392.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions during word recognition in continuous speech. *Cognitive Psychology, 10*, 29–63.

Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken word: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 576–585.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1–86.

Semlitsch, H. V. ;, Anderer, P. ;, Schuster, P. ;, & Preslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts applied to the P300 ERP. *Psychophysiology, 23*, 695–703.

Spinelli, E., Segui, J., & Radeau, M. (2001). Phonological priming in spoken word recognition with bisyllabic targets. *Language and cognitive processes, 16*, 367–392.

Tyler, L., & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception and Psychophysics, 34*, 409–420.

Xing, H. B., Shu, H., & Li, P. (2002). A self-organizing connectionist model of character acquisition in Chinese. In *Proceedings of the twenty-fourth annual conference of the cognitive science society*. Mahwah, NJ: Lawrence Erlbaum.

Van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience, 13*, 967–985.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 25*, 394–417.

Zwitserlood, P. (1989). The locus of the effects of sentential semantic context in spoken word processing. *Cognition, 32*, 25–64.