Neuroscience 158 (2009) 1416–1425

PERCEPTION OF PROSODIC HIERARCHICAL BOUNDARIES IN MANDARIN CHINESE SENTENCES

W. LI[a,b] AND Y. YANG[a]*

[a]*State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, 4A Datun Road, Chaoyang District, Beijing, China, 100101*

[b]*Graduate University of Chinese Academy of Sciences, 19A Yuquan Road, Shijingshan District, Beijing, China, 100049*

Abstract—The current study aimed at investigating the processing of prosodic hierarchical boundaries in Mandarin Chinese sentences using electroencephalography, mainly focused on the following questions: (1) whether prosodic boundaries at different levels could evoke the closure positive shift reflecting prosodic boundary perception; (2) what were the differences between them at latency, amplitude and topography; (3) whether this positive component was modified by the variations of acoustic cues (e.g. pause). Main results were: (1) As the previous studies indicated, intonational phrases elicited the closure positive shift as a marker of online speech structuring; (2) phonological phrases evoked the same positive effect with shorter onset latency and somewhat lower amplitude; (3) when the pauses in the vicinity of prosodic boundaries were entirely removed, the original latency difference between the two conditions disappeared, which clearly demonstrated the influence of pause on prosodic boundary processing; (4) prosodic word boundaries only induced amplitude variation waving around the baseline, which was more positive compared with the one elicited by syllable boundaries. The present results indicated that listeners were very sensitive to both intonational phrase boundaries and phonological phrase boundaries. © 2009 IBRO. Published by Elsevier Ltd. All rights reserved.

Key words: closure positive shift, syllable boundaries, prosodic word boundaries, phonological phrase boundaries, intonational phrase boundaries.

Auditory language processing depends upon various sources of information such as syntax, semantics, pragmatics and phonology, including prosody. Prosody comprises intonation, accentuation and rhythmic patterns, conveying both linguistic and affective information to listeners by variations in acoustic–phonetic parameters such as fundamental frequency (F0), intensity, duration, timbre and spectral characteristics.

Spoken language is organized into a hierarchy of prosodic constitutes according to a tree or grid structure. While researchers may have different views on how many and what levels are included in sentence prosodic hierarchy, intonational phrase, phonological phrase and prosodic word are commonly recognized as components of sentence prosody (Selkirk, 1980; Beckman and Pierrehumbert, 1986; Nespor and Vogel, 1986; Wightman et al., 1992). (Phonological phrase, the unit just below the intonational phrase, has many different names and definitions in the literature (Shattuck-Hufnagel and Turk, 1996, for a review). In this article, we used the universal term, phonological phrases, and we refer to the definition given by Nespor and Vogel (1986).) In comparison with Indo-European languages, the syllable is a more significant phonological unit of mandarin Chinese: it is the unit of lexical tones and a temporal unit. Thus, in the prosody studies of Mandarin, a four-tier hierarchy which includes intonational phrase, phonological phrase, prosodic word and syllable has been widely adopted (Cao, 2003; Lin, 2000; Wang et al., 2004; Yang, 1997). Intonational phrases are the largest phonological entity with phonetically definable boundaries into which utterance can be divided, and often correspond to whole sentences (or propositions within a sentence). Their boundaries are marked with pre-boundary lengthening, pause and pitch declination followed by pitch reset upon crossing the boundary (De Pijper and Sanderman, 1994; Shattuck-Hufnagel and Turk, 1996; Wang et al., 2004; Wightman et al., 1992). Among the boundaries of prosodic units, intonational phrase boundaries (IPBs) are most easily recognized by listeners. Phonological phrases, at a lower level of the prosodic hierarchy, often correspond but not necessarily to syntactic constituents (Nespor and Vogel, 1986; Shattuck-Hufnagel and Turk, 1996). From a phonetic point of view, phonological phrases are typically characterized by prefinal lengthening (Wang and Yang, 2002; Wang et al., 2004; Wightman et al., 1992), pitch reset of declination line and optional pause (De Pijper and Sanderman, 1994; Wang et al., 2004). Except for the syntactic and semantic relevance, the appearance of phonological phrase boundaries (PPBs) in a sentence is also influenced by factors such as speech rate and hesitations (filled or unfilled pauses), which might make them more difficult for listeners to detect (Ischebeck et al., 2008). Phonological phrase typically contains one or more prosodic words. A prosodic word consists of a content word, potentially grouped with some functional elements (Christophe et al., 2003), and marked by pre-boundary lengthening and pitch discontinuity of intonation counter (Lin, 2000; Wang and Yang, 2002; Wang et al., 2004).

Using behavioral methods, it has been shown that listeners are sensitive to boundaries of these prosodic units (De Pijper and Sanderman, 1994; Yang, 1998), and

*Corresponding author. Tel: +86-10-64888629; fax: +86-10-64872070.
E-mail address: yangyf@psych.ac.cn (Y. Yang).

Abbreviations: ANOVA, analysis of variance; COND, condition; CPS, closure positive shift; EEG, electroencephalogram; EOG, electrooculogram; ERPs, event-related potentials; F0, fundamental frequency; HEM, hemisphere; IPBs, intonational phrase boundaries; PPBs, phonological phrase boundaries; PWBs, prosodic word boundaries; REG, region; SBs, syllable boundaries; TW, time window.

0306-4522/09
doi:10.1016/j.neuroscience.2008.10.065

use them to disambiguate lexical ambiguities (Christophe et al., 2003, 2004; Isel et al., 2003; Salverda et al., 2003) and sentence level ambiguities (Beach, 1991; Clifton et al., 2002; Kjelgaard and Speer, 1999; Millotte et al., 2007, 2008; Nagel et al., 1996; Schafer et al., 2000; Speer et al., 1996). However, due to methodological reasons, the exact relationship between prosody and sentence processing is still not very clear, particularly the knowledge about the cognitive and neural basis of prosodic processing.

Due to its implicit on-line characteristics and its high time resolution, the employment of event related potential (ERPs) measures has joined the list of on-line methods as an additional approach to the study of language processing. In contrast to response times, ERPs patterns can more easily distinguish between different levels of linguistic processing. In the last two decades, the ERPs literature has mainly focused on the electrophysiological correlates of semantic and syntactic processing rather than prosodic and pragmatic processing. The first language-related component to be discovered was the N400 (Kutas and Hillyard, 1980), a negative component peaking around 400 ms post–critical word onset which reflects semantic processing and the integration of word meaning in sentence or discourse contexts. Syntactic processing difficulties due to violations as well as garden path effects and complex structures also elicited late centro-parietal positivities, referred to as P600 (SPS) components (Kaan et al., 2000). In addition, the P300, which was related to working memory (Donchin and Coles, 1988) and can be elicited by almost any "rare and relevant" stimulus, has also been widely studied by psychologists and neuroscientists. Other studies have also indicated that it reflects phasic activity of the neuromodulatory locus coeruleus–norepinephrine system (Nieuwenhuis et al., 2005).

Recently a number of studies have investigated the neurophysiological basis of prosody in various aspects, including the emotional function (Schirmer et al., 2002), the lexical function (Böcker et al., 1999; Friedrich et al., 2001) and the modality function of prosody (Astésano et al., 2004). A series of experiments were also conducted to investigate the structural function of prosody at the sentence level. Using sentences differing in their prosodic phrasing patterns caused by underlying syntactic structures, Steinhauer et al. (1999) found that a positive-going waveform, termed as the closure positive shift (CPS), was reliably elicited by IPBs. It seems to be triggered by the prefinal constituent lengthening, and lasts between approximately 500–1000 ms, with a maximal centro-parietal distribution (Steinhauer, 2003). This component was also observed for sentence materials that were deprived of semantic information, such as pseudoword sentences (Pannekamp et al., 2005), and for sentences with reduced or without segmental information, such as filtered speech materials (Steinhauer and Friederici, 2001) and hummed sentences (Pannekamp et al., 2005). The results of these studies have suggested that the CPS exclusively relies on pure prosodic information and has nothing to do with other segmental information (Pannekamp et al., 2005; Steinhauer, 2003). Furthermore, it has also been shown that the CPS is a reliable marker for prosodic phrasing during silent reading based on comma rules (Steinhauer and Friederici, 2001; Steinhauer, 2003).

For the past few years, the comparative study of music and language has been gaining a great deal of research interest (Marques et al., 2007; Patel, 2003; Schon et al., 2004). Like language, music is a human universal auditory system involving perceptually discrete elements organized into hierarchically structured sequences. They can serve as foils for each other in the study of brain mechanisms underling complex sound processing (Knösche et al., 2005; Patel, 2003). Just as expected, a CPS-like component has been observed in music phrasing (Knösche et al., 2005; Nan et al., 2006; Neuhaus et al., 2006). It has similar topography and amplitude to the CPS in speech but different latency and duration. Further studies have indicated that music phrase boundary processing entails not only the detection of phrase boundary cues, such as pause, but also the integration of information between the previous phrase and the upcoming one (mostly the initial note) (Knösche et al., 2005; Nan et al., 2006, 2008).

Ever since the CPS was found in 1999, researchers have been exploring the nature of this ERP component and its relation to acoustic cues. Although many valuable results have been obtained, many issues remain to be further clarified. The CPS has been demonstrated only at IPBs in spoken single sentences. It is well known that there are other levels (e.g. prosodic words and phonological phrases) in sentence prosody (Selkirk, 1980; Beckman and Pierrehumbert, 1986; Nespor and Vogel, 1986), and the acoustic–phonetic cues of their boundaries are similar but change systematically in quantities (Wang et al., 2004; Wightman et al., 1992; Yang, 1997). Moreover, behavioral studies revealed that untrained subjects can reliably differentiate the degree of these prosodic boundaries (De Pijper and Sanderman, 1994; Yang, 1998). On the basis of these findings, it is reasonable to expect that the CPS may also be evoked by prosodic boundaries at other levels. Hence, the goal of the present study is to investigate the cognitive processing of prosodic hierarchical boundaries, more specifically, to explore (1) whether prosodic boundaries at different levels could evoke the CPS, and (2) what the differences are between these CPS. It was expected that prosodic boundaries at different levels in sentences could all induce positive shifts, and that the higher prosodic boundaries could elicit stronger positive effect.

Mandarin Chinese differs significantly from most Indo-European languages in its syntactic and especially phonological systems. It involves a tonal system, which is acoustically realized by the pitch variation of syllables, and functionally distinguishes their meanings. As a tonal language, Chinese intonation is realized by the pitch variations of the sentence, which is determined by the conjuncture of syllable pitch contours and modulations dictated by sentence structures, such as information structure. In the model adopted by most Chinese scholars, intonation is described by two counters, the top and bottom lines. The bottom line consists of the lowest values of tone registers in the sentence, and reflects hierarchical structure of the prosodic constituents. The top line consists of highest values of tone registers, and reflects pitch accent locus. In the hierarchical structure of sentence prosody the pitch value of the bottom line for bigger rhythm constituent is lower than that for the smaller one (Shen, 1985). Besides, most words (more than 70%) in Chinese are disyllables, and there is usually no perceived pause between the two syllables. Therefore, the boundary between disyllables was taken as the baseline condition (COND) in this study.

EXPERIMENT 1
EXPERIMENTAL PROCEDURES

Participants. Twenty right-handed university students (nine men and 11 women, at the age of 18–25 years, with mean age of 22.4 years) participated in the experiment and were paid for their participation. All of them were native speakers of Mandarin Chinese without hearing or neurological disorders.

Stimuli. The stimulus materials contained 50 groups of sentences. In each group, there were four sentences (CONDs) of the same length, in which two critical syllables were embedded serving as pre-boundary and post-boundary syllables. The pre-boundary syllables had the same onset, rhyme and tone (namely homophone), the same for the post-boundary ones. The boundaries between the two critical syllables in the four sentences were boundaries between disyllables (syllable boundaries, SBs), prosodic word boundaries (PWBs), PPBs and IPBs respectively. The position of the two critical syllables in sentences was the same in each group but different among groups. In addition, the critical syllables never existed in other positions of the same group. As an example in the following group, the critical syllables are "xian1" and "hua1," which are indicated in italics.

(1a) 商店里的/*鲜花*/散发出/阵阵/浓郁的/芳香。(SBs)

Shang1dian4li3de0/*xian1hua1*/san4fa1chu1/zhen4zhen4/nong2yu4de0/fang1xiang1. In the store/the *flowers*/emit/bouts of/full-bodied/aroma.
The flowers in the store emit bouts of full-bodied aroma.

(1b) 小华/最好/*先*/*花*/点/时间/学习一下/弹/钢琴。(PWBs)

Xiao3hua2/zui4hao3/*xian1*/*hua1*/dian3/shi2jian1/xue2xi2yi2xia4/tan2/gang1qin2.
Xiaohua/had better/*first*/*spend*/some/time/learning/to play/the piano.
Xiaohua had better spend some time learning to play the piano first.

(1c) 养/这盆/水*仙*/*花*了/我/大量的/时间/和/精力。(PPBs)

Yang3/zhe4pen2/shui3*xian1*/*hua*le0/wo3/da4liang4de0/shi2jian1/he2/jing1li4.
Planting/this/*narcissus*/*costs*/me/a lot of/time/and/effort.
It costs me a lot of time and effort to plant this narcissus.

(1d) 想/保持/领*先*/*花*/时间/进行/练习/非常/必要。(IPBs)

Xiang3/bao3chi2/ling3*xian1*/*hua1*/shi2jian1/jin4xing2/lian4xi2/fei1chang2/bi4yao4.
If you want to/keep/*ahead*/, *taking*/time/to do/exercises/is very/necessary.
If you want to keep ahead, it is very necessary to take time to do exercises.

All of these sentences were produced by a male speaker of standard Chinese and recorded in a soundproof chamber, at a sampling rate of 22 kHz. For analyzing the tonal and durational properties of the prosodic boundaries, the duration of pre-boundary syllables (D1), the duration of pre-boundary syllables plus following silences (D2), and the minimal values of pitch for the pre-boundary syllables (F1) and post-boundary syllables (F2) in each COND were measured. Besides, the degree of pitch reset (F2 minus F1) was acquired. For the D1, a one-way analysis of variance (ANOVA) with COND (SBs, PWBs, PPBs and IPBs) run on duration showed a significant main effect ($F(3,196)=16.60$, $P<0.001$). For multiple comparisons, Bonferroni corrections were used and six independent tests were performed. In order to limit the risk of making at least one type I error to an overall value of 0.05, a stricter threshold of 0.05/6=0.008 was used instead of 0.05 (Abdi, 2007). This was the case for all the other multiple comparisons including the ERP data. Post hoc tests indicated that pair-wise comparisons were significant between SBs (M=212.72, S.D.=49.52) and PPBs (M=276.36, S.D.=59.29), $P<0.001$; SBs and IPBs (M=263.08, S.D.=59.68), $P<0.001$; PWBs (M=211.8, S.D.=63.93) and PPBs, $P<0.001$; PWBs and IPBs, $P<0.001$, except for SBs and PWBs, $P>0.008$, PPBs and IPBs comparison, $P>0.008$ (see Fig. 1A). For the D2, a one-way ANOVA with COND run on duration also showed a significant main effect ($F(3,196)=150.78$, $P<0.001$). Further post hoc tests indicated that pair-wise comparisons were significant between SBs (M=212.72, S.D.=49.52) and PPBs (M=340.86, S.D.=98.05), $P<0.001$; SBs and IPBs (M=542.24, S.D.=126.72), $P<0.001$; PWBs (M=211.80, S.D.=63.93) and PPBs, $P<0.001$; PWBs and IPBs, $P<0.001$; PPBs and IPBs, $P<0.001$, except for SBs and PWBs comparison, $P>0.008$ (see Fig. 1A). For the pitch, a one-way ANOVA with COND run on pitch reset showed a significant main effect ($F(3,196)=10.69$, $P<0.01$). Further post hoc tests indicated that pair-wise comparisons were significant between SBs (M=−6.50, S.D.=40.64) and PPBs (M=27.18, S.D.=35.90), $P<0.001$; SBs and IPBs (M=31.23, S.D.=31.72), $P<0.001$. However, they were not significant between SBs and PWBs (M=10.09, S.D.=40.58), $P>0.008$; PWBs and PPBs, $P>0.008$; PWBs and IPBs, $P>0.008$; PPBs and IPBs, $P>0.008$ (see Fig. 1B).

Procedure. The 200 experimental sentences were intermixed with 100 filler sentences and presented aurally in a pseudo-randomized order in four blocks of 75 trials. In each block, the sentences from the same COND were presented in no more than three consecutive trials. Each trial began with a fixation cross in

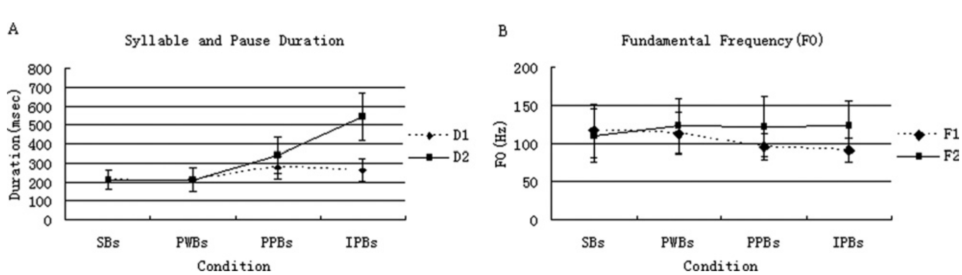

Fig. 1. Prosodic parameters. Left: the duration (ms) of pre-boundary syllable (D1) and the duration of pre-boundary syllable+silence (D2); right: F0 for pre-boundary syllable (F1) and post-boundary syllable (F2).

the center of the monitor in order to minimize eye movements which can cause artifacts in the electroencephalogram (EEG). After 300 ms, the sentence was presented aurally, while the cross remained on the screen. Subjects were instructed to listen carefully and comprehend the meaning of the sentences. The task was to answer yes or no to comprehension questions such as "Did Zhao Dan cry?" in 20% of the sentences as accurately as possible.

EEG recording. EEG was continuously recorded from 64 cap-mounted Ag/AgCl electrodes with a sampling rate of 500 Hz. The EEG data were re-referenced off-line to linked mastoid electrodes by subtracting from each sample of data recorded at each channel one-half of the activity recorded at the right mastoid. The electrooculogram (EOG) was recorded from electrodes placed at the outer canthus of each eye and from sites above and below the left eye. Impedances were kept below 5 kΩ. The EEG and EOG recordings were amplified with a high cutoff of 100 Hz. Participants were instructed before the beginning of the experiment to avoid eye blinking and other body movements during the presentation of the auditory sentences.

Data analysis. EEG data were processed with the software NeuroScan 4.3 (Neuroscan; Compumedics Limited, 30-40 Flockhart Street, Abbotsford Vic 3067, Australia). The EEG and EOG signals were screened off-line for eye movements and electrode drifting. The data were filtered with a low-pass filter of 40 Hz. Epochs comprised the 200 ms preceding and 1500 ms following the pre-boundary syllable onsets.

Grand average curves were first analyzed by visual inspection in order to identify time windows (TWs) of interest. PWBs only elicited amplitude variation waving around the baseline, and it was more prominent between 300 and 700 ms. The positive shifts were observed in the TW of 300–700 ms after the acoustic onset of the pre-boundary syllables for PPBs COND, thus, this TW was selected for analyzing. And for the same reason, the TW of 700–1100 ms was selected for analysis of the positive shifts elicited by IPBs. For each of the identified TW, a three-way ANOVA was performed with mean amplitude as dependent factor, with COND (SBs, PWBs, PPBs, IPBs), region (REG, anterior–posterior topography: anterior—F3, FZ, F4; central—C3, CZ, C4; posterior—P3, PZ, P4), and hemisphere (HEM, left-right topography: left—F3, C3, P3, midline—FZ, CZ, PZ, right—F4, C4, P4) as independent factors to test the general boundary effect. To further compare amplitude of the shifts elicited by PPBs and IPBs, another three-way ANOVA was conducted with mean amplitude from each TW as dependent factor, with COND (PPBs, IPBs), REG and HEM as independent factors. To establish the onset of the positive shifts, a series of onset analyses were conducted in consecutive mean amplitude latency bins of 10 ms wide (e.g. 100–110 ms, 110–120 ms, etc.) for PWBs, PPBs and IPBs compared with SBs. *P*-values were reported after Greenhouse-Geisser correction for non-sphericity.

RESULTS

As shown in Fig. 2, PPBs and IPBs elicited positive shifts separately starting at about 250 ms after the onset of pre-boundary syllables. We interpreted the positive shifts

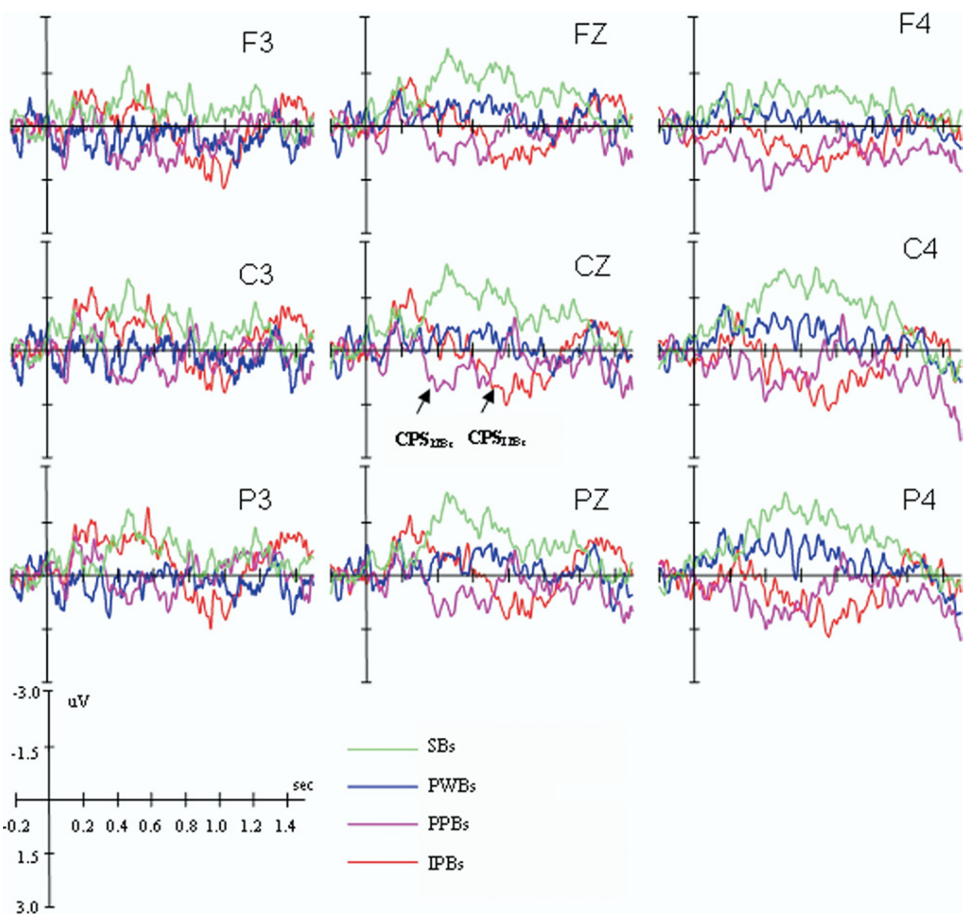

Fig. 2. Grand average waveforms over participants (n=20), time locked to the onset of pre-boundary syllable at selected channels for SBs, PWBs, PPBs and IPBs, four CONDs are superimposed. CPS_{PPBs} and CPS_{IPBs} represent the CPS elicited by PPBs and IPBs respectively.

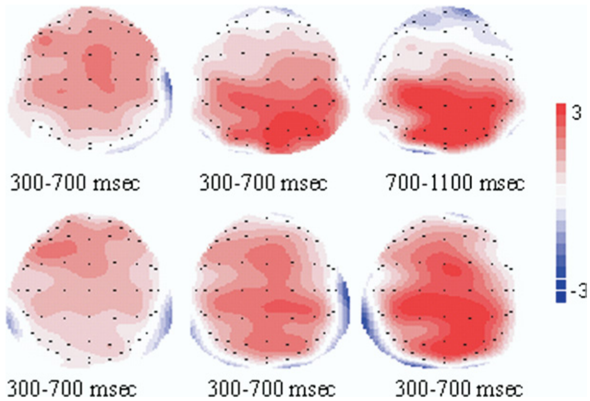

Fig. 3. Upper: Topography of the ERP effects for PWBs, PPBs and IPBs in order in experiment 1. Lower: Topography of the ERP effects for PWBs, PPBs′ and IPBs′ in order in experiment 2.

as the CPS, since their polarity, latency and topography (see Fig. 3: the upper one; the topographical distribution was more concentrated and stronger for the CPS elicited by IPBs) fit the standard characteristics of the CPS. However, PWBs only elicited amplitude variation waving around the baseline, which was more positive compared with SBs.

In the TW of 300–700 ms, the main effect of COND was observed ($F(3,57)=15.026$, $P<0.001$). Further post hoc tests indicated that pair-wise comparisons were significant between SBs (M=−1.326, S.E.=0.239) and PWBs (M=−0.295, S.E.=0.285), $P<0.008$; SBs and PPBs (M=0.729, S.E.=0.281), $P<0.001$; SBs and IPBs (M=−0.135, S.E.=0.260), $P<0.008$; PWBs and PPBs, $P<0.008$, while the other pair-wise comparisons were not significant, $P>0.008$. In addition, a COND×REG interaction ($F(6, 114)=8.342$, $P<0.001$) was shown in this TW. Simple effect analysis revealed that the effect of COND was significant at the anterior ($F(3, 57)=9.230$, $P<0.001$), the central ($F(3, 57)=15.80$, $P<0.001$) and the posterior electrodes ($F(3, 57)=14.010$, $P<0.001$). Results of multiple comparison for the three areas indicated that there was no significant difference between SBs and IPBs at the anterior electrodes, $P>0.008$, PWBs and IPBs at the central electrodes, $P>0.008$, PPBs and IPBs at the posterior electrodes, $P>0.008$, while the other pair-wise comparisons were significant, $P<0.008$. All effects were distributed bilaterally across HEMs.

In the TW of 700–1100 ms, the main effect of COND was observed ($F(3, 57)=13.098$, $P<0.001$). Further post hoc tests indicated that pair-wise comparisons were significant between SBs (M=−0.894, S.E.=0.256) and PWBs (M=−0.037, S.E.=0.256), $P<0.005$; SBs and PPBs (M=0.163, S.E.=0.290), $P<0.001$; SBs and IPBs (M=0.776, S.E.=0.184), $P<0.001$. Besides, the pair-wise comparisons approached significance between PWBs and IPBs, $P=0.009<0.01$, while the other pair-wise comparisons were not significant, $P>0.008$. No interactions were provable between the factors COND×HEM and COND×REG.

The CPS amplitude elicited by IPBs (M=0.776) seemed a little larger than that by PPBs (M=0.729), but the result of statistical analysis indicated that there was no significant difference ($F(1,19)=0.027$, $P>0.05$). The analysis of onset latency revealed that the CPS elicited by PPBs started in 320–330 ms latency range and 370–380 ms for IPBs followed by a long and uninterrupted series of bins with significant effect of boundaries. However, the deflection elicited by PWBs started in 410–420 ms latency range.

The present auditory experiment indicated that IPBs elicited positive shift reliably, which strongly resembled the CPS component as established by previous studies (Isel et al., 2005; Pannekamp et al., 2005; Steinhauer et al., 1999; Steinhauer, 2003). Most importantly, this ERP effect was also evoked whenever listeners detected PPBs. Both of them were associated with prosodic phrase processing, had a positive polarity, and displayed a bilateral centro-parietal scalp distribution with largest amplitudes at the posterior electrodes. Given all these characteristics, we identified the shifts elicited by PPBs and IPBs as the CPS. In contrast to previous studies (Steinhauer et al., 1999, 2001; Steinhauer, 2003; Pannekamp et al., 2005), we replicated a CPS effect for the processing of critical syllables located at the PPBs and IPBs using natural speech without any ambiguity. The replication of the CPS demonstrated the robustness of this physiological correlate. However, the ERP effect of PWBs was not so promising. PWBs in the current experiment did not induce the CPS, but only amplitude variation waving around the baseline. According to the acoustic analysis, the length of pre-boundary syllable for PWBs (M=211.800 ms) was significantly shorter than PPBs (M=276.360 ms) and IPBs (M=263.080 ms), and there was no pause insertion for PWBs. Theoretically the processing of PWBs should be faster than PPBs and IPBs, and accordingly show shorter onset latency relative to these two CONDs. However, the shifts elicited by PWBs started in 410–420 ms latency range, and might not be triggered by pre-boundary syllable. Besides, they showed frontal scalp distribution (see Fig. 3: the upper one), which further indicated that the shifts related to PWBs were not the CPS.

Special attention should also be paid to the point that the CPS evoked by the two CONDs manifested difference in visual detection and statistical analysis: the CPS elicited by PPBs showed shorter onset latency as compared with the one evoked by IPBs. Given that the waveforms were time locked to the onset of the pre-boundary syllable, and the difference in D2 (duration of pre-boundary syllable+length of pause) for PPBs and IPBs was significant, it was reasonable to suspect that the onset latency difference in the CPS for the two CONDs might reflect the increased duration of that pause in IPBs relative to PPBs COND. Besides, the CPS amplitude for IPBs was somewhat larger than the one for PPBs, which may also be partially attributable to the longer pause insertion of the former COND. In order to explore the contribution of the pause, the second experiment was conducted with pause removal in both PPBs and IPBs CONDs, which was similar to the third experiment in Steinhauer et al. (1999). Thus in the experiment 2, the entire pauses in both CONDs were carefully removed, preserving other prosodic boundary cues intact. The two resulting CONDs were referred to PPBs′ and IPBs′ respectively. If the original latency difference between PPBs and IPBs disappeared and the amplitude were somewhat lower in new CONDs, it would confirm the impact of pause on boundary processing.

EXPERIMENT 2
EXPERIMENTAL PROCEDURES

Participants. The participants were 19 right-handed university students (7 men and 12 women, at the age of 18–25 years, with the mean age of 21.8 years) with similar characteristics to those from experiment 1, and were paid for their participation.

Stimuli. The stimulus materials were the same as experiment 1, for the exception of CONDs PPBs and IPBs were replaced by PPBs′ and IPBs′.

Procedure and EEG recording. The procedure and EEG recording were identical to those of experiment 1.

Data analysis. The data were preprocessed using the same procedure as in experiment 1, with the same type of statistical analysis. The only difference was that the positive shift elicited by IPBs′ was observed in the TW of 300–700 ms after the acoustic onset of the pre-boundary syllables, thus, this TW was selected for analysis in both PPBs′ and IPBs′ CONDs.

RESULTS

As shown in Fig. 4, PPBs′ and IPBs′ elicited positive shifts starting at about 200 ms after the onset of pre-boundary syllables respectively. Due to the scalp topography (see Fig. 3: the lower one), the latency and the morphology of the positive-going ERP in the two CONDs, we also interpreted the deflection as the CPS. Besides, PWBs only induced amplitude variation more positive over centro-frontal electrodes compared with the one related to SBs.

In the TW of 300–700 ms, the main effect of COND was observed ($F(3, 54)=21.666$, $P<0.001$). Further post hoc tests indicated that pair-wise comparisons were significant between SBs (M=−1.219, S.E.=0.208) and PPBs′ (M=0.532, S.E.=0.228), $P<0.001$; SBs and IPBs′ (M=0.568, S.E.=0.182), $P<0.001$, PWBs (M=−0.674, S.E.=0.209) and PPBs′, $P<0.008$; PWBs and IPBs′, $P<0.008$. However, the pair-wise comparisons were not significant between SBs and PWBs, $P>0.008$, PPBs′ and IPBs′, $P>0.008$. A COND×REG interaction ($F(6, 108)=2.800$, $P<0.05$) was shown in this TW. Simple effect analysis revealed that the effect of COND was not significant at the anterior ($F(3, 54)=1.76$, $P>0.05$) and posterior areas ($F(3, 54)=0.59$, $P>0.05$), but significant at the cen-

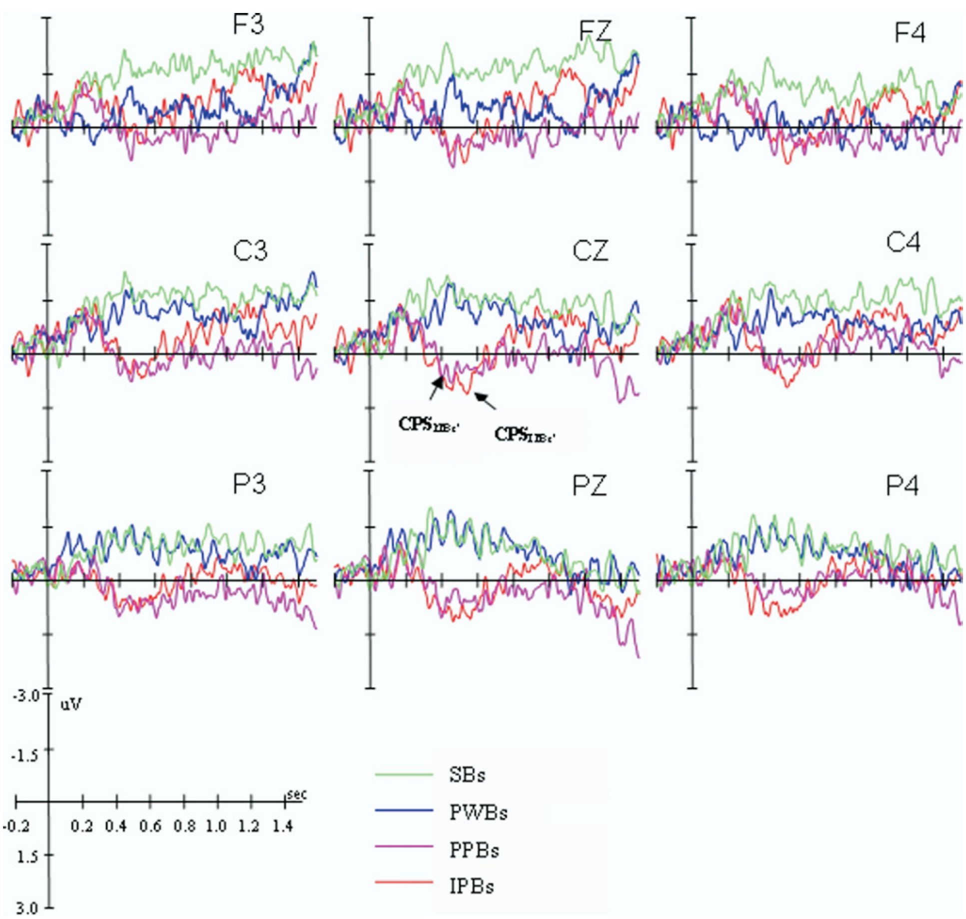

Fig. 4. Grand average waveforms over participants ($n=19$), time locked to the onset of pre-boundary syllable at selected channels for SBs, PWBs, PPBs′ and IPBs′, four CONDs are superimposed. $CPS_{PPBs'}$ and $CPS_{IPBs'}$ represent the CPS elicited by PPBs′ and IPBs′ respectively.

tral electrodes ($F(3, 54)=6.87$, $P<0.01$). Results of multiple comparison indicated that there was no significant difference between SBs and PWBs, $P>0.008$, PPBs′ and IPBs′, $P>0.008$, while other pair-wise comparisons were significant, $P<0.001$. In addition, a COND×HEM interaction was also shown ($F(6, 108)=3.573$, $P<0.01$). Simple effect analysis revealed that the effect of COND was not significant at the left ($F(3, 54)=2.40$, $P>0.05$) and midline ($F(3, 54)=0.57$, $P>0.05$), but significant at the right HEM ($F(3, 54)=5.89$, $P<0.01$). Results of multiple comparison indicated that there was no significant difference between SBs and PWBs, $P>0.008$, PPBs′ and IPBs′, $P>0.008$, while other pair-wise comparisons were significant, $P<0.008$.

The analysis of onset latency revealed that the CPS started in 320–330 ms time range for PPBs′, and 310–320 ms for IPBs′ followed by a long and uninterrupted series of bins with significant effect of boundaries. However, the deflection elicited by PWBs started in 310–320 ms time range and interrupted at about 340 ms.

ERP results for the new CONDs PPBs′ and IPBs′ confirmed that even without pause, prosodic boundaries were still perceived by the listeners, since we observed the CPS stably in these two CONDs. This finding clearly demonstrated that the original pause insertion was a dispensable cue for the hearers' detection of prosodic boundary in the presence of other prosodic parameters. It indicated that the CPS effect reflected the processing of prosodic boundary rather than the perception of a pause interrupting the speech input, which has been showed by other researchers (Nan et al., 2006; Steinhauer et al., 1999; Steinhauer, 2003). In addition, although the deflection induced by PWBs seemed a little positive in centro-frontal area (see Fig. 3: the lower one), it was not statistically different from the shift elicited by SBs. It seemed impossible to define it as the CPS.

It is important to note that pause is very essential for marking the level of prosodic boundaries. The removal of pause may have reduced the difference between IPBs and PPBs, and thereby modified the perception of IPBs to some extent. As indicated by this experiment, the original onset latency difference of the CPS elicited by PPBs and IPBs in experiment 1 was indeed mainly determined by the increased duration of that pause in IPBs as compared with PPBs. When the pauses were entirely removed, the CPS latency for IPBs′ was much shorter (started in 310–320 ms latency range for IPBs′, but 370–380 ms for IPBs), and almost equal to the one related to PPBs and PPBs′ (both started in 320–330 ms time range). It appears that the processing of prosodic boundaries was much quicker after pause removal, which was reflected by the shorter latency in the ERP.

Furthermore, the pauses maybe also influenced the amplitude of the positive shift: when the pauses were removed, the mean amplitude of the CPS was reduced from 0.729 μv to 0.532 μv for PPBs, and from 0.776 μv to 0. 568 μv for IPBs. To provide evidence statistically, we conducted another two ANOVAs with mean amplitude in each TW as dependent factor, with COND (PPBs and PPBs′, or IPBs and IPBs′), REG and HEM as independent factors. The two ANOVAs showed no main effect of COND (PPBs and PPBs′: $F(1, 18)=0.002$, $P>0.05$; IPBs and IPBs′: $F(1, 18)=0.707$, $P>0.05$), and no interactions were provable between the factors COND×HEM and COND×REG. Thus, although it seemed that the impression of phrasing was reduced by pause removal, the CPS amplitude was not significantly influenced.

To sum up, pause was not the decisive cue to elicit the CPS, but it can modify this ERP component effectively.

DISCUSSION

This research intends to investigate the electrophysiological responses to the processing of prosodic hierarchical boundaries in sentences. Many homophones with the same onset, rhythm and tone in Mandarin Chinese give us a good opportunity to provide answers to this question. By manipulating the boundary size between two critical syllables, prosodic hierarchical boundaries were realized in four sentences of each group at the same position. Using this kind of natural material, the current study replicated the result obtained in studies by Steinhauer and other researchers, showing that there were bilateral positive variations at central and parietal electrodes (the CPS) in correlation to IPBs. Second, and most importantly, it clearly showed that PPBs, as a lower level in the sentence prosody hierarchy, also elicited this positive deflection with similar temporal pattern of latency and scalp distributions. Furthermore, differences were obvious between the CPS evoked by these two CONDs, that is, the CPS elicited by IPBs showed longer onset latency and a little larger amplitude (although not significant statistically) compared with the one elicited by PPBs. However, when the pauses in the vicinity of boundaries were entirely removed, the original onset latency difference between PPBs and IPBs disappeared. As the lowest prosodic level investigated in the current study, PWBs did not induce the CPS, but only amplitude variation waving around the baseline.

Phonological phrases are at a lower level of the prosodic hierarchy, and may not necessarily correspond to syntactic boundaries. In addition, position and size of PPBs depend on several factors such as speech rate and hesitations (filled or unfilled pauses), which make them more difficult for listeners to detect (Ischebeck et al., 2008). However, phonological phrases are an important constituent in the sentence prosodic hierarchy and show similarities with intonational phrases in its acoustic correlates, such as the pitch reset of the bottom line of intonation counter and the prefinal lengthening. Furthermore, phonological phrases and intonational phrases have the shared function of segmenting sentences into smaller units. Research about developmental psychology has suggested that even newborn infants can perceive the cues correlate with PPBs (Christophe et al., 2001), and begin to react to the disruption of phonological phrases in whole sentences at about 9 months old (Gerken et al., 1994). Further studies have also shown that phonological phrases help infants and adults to segment spoken sentences into words

(Christophe et al., 2001, 2004; Gout et al., 2004; Millotte et al., 2007). These findings indicate that although PPBs are more difficult to detect in comparison to IPBs, listeners are very sensitive to them. The CPS occurs immediately when PPBs are perceived. It is probable that the CPS correlate with PPBs might have the same mechanism underlying the phrasing process, given the fact that the CPS elicited by PPBs and IPBs holds a similar timing pattern and topographical distribution.

As an independent level in sentence prosodic hierarchy, PWBs were expected to evoke the CPS effect like PPBs and IPBs. However, both auditory experiments showed the opposite results. Although the amplitude variation elicited by PWBs had a positive trend compared with SBs, it diverged in shape, temporal pattern and scalp distribution from the standard CPS. The reason why they cannot evoke the CPS may result from the fact that the prosodic words used in the experiments were not very typical. As we mentioned in the introduction, prosodic word consists of a content word, potentially grouped with some functional elements (Christophe et al., 2003). Nevertheless, in order to create proper stimulus complying with the purpose of our experiments, prosodic words in the current study only include a content word without any function word. These atypical prosodic words might lead to the weak acoustic makers of the PWBs as reflected in Fig. 1, and thereby the impression of phrasing was reduced. Descriptively, no difference existed in the length of pre-boundary syllable and pause between PWBs and SBs. Further, the statistical analysis showed that there was no significant difference between the two CONDs for the pitch reset (the difference between F2 and F1), although it was much higher across PWBs ($M=10.09$) as compared with SBs ($M=-6.50$). In principle, phrasing and phrase boundary detection rely on both temporal and spectral cues. However, the distinction of very fine-grained temporal features seems to play the most important role (Neuhaus et al., 2006). At least for the sentences used in the current study, listeners did not clearly detect the PWBs marked only by pitch movement, and accordingly the CPS could not be elicited in the listener's ERP. Nevertheless, it was also possible that PWBs cannot elicit the CPS in any CONDs. As some researchers have stated, the speech stream may be spontaneously perceived as a string of prosodic units, roughly corresponding to phonological phrases (Christophe et al., 1997, 2001). Prosodic words are not a constituent easy to detect without paying special attention to the prosodic information, so they cannot elicit stable brain response, which have been shown by the inconsistent results in the two auditory experiments. However, this is only our speculation. Whether the CPS can be elicited by PWBs in more typical CONDs is worthy of investigation in future.

In accordance with the cue-trading hypothesis, experiment 2 clearly showed that even after pause removal the remaining acoustic parameters (such as constituent lengthening) prove sufficient to mark the prosodic boundary in PPBs′ and IPBs′, since both CONDs elicited the CPS stably. The current data further indicated that the CPS was not simply related to the pause at the boundary but to an entire ensemble of prosodic cues, including the pre-boundary lengthening and changes in the F0 contour.

Although pause of prosodic boundary was not a necessary or sufficient factor to elicit the CPS, it can modify this ERP component effectively. For the onset latency difference between PPBs and IPBs in experiment 1, it was suspected that the longer pause duration in the later COND may be the most potential factor. In order to examine this possibility, we carefully removed the entire pauses of the both CONDs. As expected, the CPS indeed showed much shorter onset latency for new COND IPBs′ (started in 310–320 ms time range) as compared with IPBs (started in 370–380 ms time range) and PPBs (started in 320–330 ms time range). This result clearly showed the impact of pause on boundary processing. However, it was not the case for PPBs: the CPS elicited by PPBs started in the 320–330 ms time range regardless of the presence or absence of pause. This may be because no linear correlation exists between perceived boundary strength and acoustic cues. In other words, it is unlikely that once acoustic parameters vary in quantities, they will induce cognitive difference. A closer look at the acoustic analysis revealed that the pause for PPBs was only about 65 ms (279.16 ms for IPBs). It was conceivable that the pause of this length was not long enough to evoke the CPS latency difference. The present result may also be explained from another aspect. Pause is the acoustic correlate of strong prosodic boundary (Wang and Yang, 2002), and higher prosodic hierarchy is more dependent on pause to demonstrate its boundary. Thus, when pauses were removed, it had more influence on IPBs than on PPBs, which was reflected by the different modification on these two CONDs at onset latency.

The CPS is exclusively relying on pure prosodic information (Pannekamp et al., 2005), thus the phonological difference may be the most probable factor to explain the CPS amplitude difference existing between PPBs and IPBs. IPBs are ranked higher than PPBs in sentence prosodic hierarchy. The manifestation in acoustic phonetic markers was that IPBs have larger pitch reset crossing the boundary and much longer silence as compared with PPBs, although pre-boundary syllable lengthening was shorter in some way. The result from music processing has indicated that the CPS amplitude depends on classical phrase boundary markers, such as pause length, prefinal lengthening (Neuhaus et al., 2006). Although each of the acoustic cues was not a sufficient factor to elicit the CPS, the impression of “phrasedness” caused by these factors was much stronger for IPBs. From this perspective, the CPS amplitude reflects the extent to which prosodic representations are activated, which is in congruence with prior findings (Steinhauer and Friederici, 2001; Steinhauer, 2003).

The study of perception of phrase structure in music has shown that the CPS does not directly reflect the detection of the phrase boundary as such but rather memory- and attention-related processes, which are necessary for the transition from one phrase to the next (Knösche et al., 2005; Nan et al., 2006, 2008). As in music, prosodic boundary processing in language could also be seen as a complex process comprising several cognitive subcomponents, like the detection of prosodic boundary cues, the reorientation of attention, the integration of information among prosodic units, etc. It is suspected that neither the CPS elicited by PPBs nor the one elicited by IPBs should be viewed as monolithic, but more likely consisting of multiple subcomponents reflecting multiple sub-processes. Nevertheless, the present results cannot show which factor contributes more to the elicitation of the CPS. Furthermore, some researchers demonstrated that during a dialogue if the focus of attention has been directed toward a certain word, the CPS will follow this word (Hruska et al., 2001; Toepel et al., 2007). Taking the whole picture into consideration, we may conclude that the CPS is a complex component and the particular COND of its elicitation deserves further investigation.

CONCLUSION

The ERP data from the present study suggested that the CPS occurred immediately when prosodic phrase boundaries were perceived, that is, not only IPBs but also PPBs elicited the CPS component. The amplitude variation elicited by PWBs was only more positive compared with the deflection elicited by SBs. Although pause was not a necessary factor to elicit the CPS, it can modify its onset latency and amplitude.

Acknowledgments—Sponsorship: This research was supported by the National Natural Science Foundation of China (30370481) and the Foundation of State Key Laboratory of Brain and Cognitive Science (the cognitive significance of the closure positive shift and its related prosodic cues).

REFERENCES

Abdi H (2007) Bonferroni and Sidak corrections for multiple comparisons. In: Encyclopedia of measurement and statistics (Salkind NJ, ed), pp 1–9. Thousand Oaks, CA: Sage.
Astésano C, Besson M, Alter K (2004) Brain potentials during semantic and prosodic processing in French. Cogn Brain Res 18:172–184.
Beach CM (1991) The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. J Mem Lang 30:644–663.
Beckman M, Pierrehumbert J (1986) Intonational structure in Japanese and English. In: Phonology yearbook 3 (Ohala J, ed), pp 255–309. Great Britain.
Böcker KBE, Bastiaansen MCM, Vroomen J, Brunia CHM, De Gelder B (1999) An ERP correlate of metrical stress in spoken word recognition. Psychophysiology 36:706–720.
Cao J (2003) Prediction of prosodic organization based on grammatical information. J Chin Inf Proc 17(3):41–46.
Christophe A, Guasti MT, Nespor M, Dupoux E, van Ooyen B (1997) Reflections on phonological bootstrapping: Its role for lexical and syntactic acquisition. Lang Cogn Proc 12:585–612.
Christophe A, Mehler J, Sebastian-Galles N (2001) Perception of prosodic boundary correlates by newborn infants. Infancy 2:385–394.
Christophe A, Gout A, Peperkamp S, Morgan J (2003) Discovering words in a continuous speech stream: The role of prosody. J Phon 31:585–598.
Christophe A, Peperkamp S, Pallier C, Block E, Mehler J (2004) Phonological phrase boundaries constrain lexical access: I. Adult data. J Mem Lang 51:523–547.
Clifton C, Carlson K, Frazier L (2002) Informative prosodic boundaries. Lang Speech 45:87–114.
De Pijper JR, Sanderman AA (1994) On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. J Acoust Soc Am 96:2037–2047.
Donchin E, Coles M (1988) Is the P300 component a manifestation of context updating? Behav Brain Sci 11:357–374.
Friedrich CK, Alter K, Kotz SA (2001) An electrophysiological response to different pitch contours in words. Neuroreport 12:3189–3191.
Gerken L, Jusczyk PW, Mandel DR (1994) When prosody fails to cue syntactic structure: 9-month-olds' sensitivity to phonological versus syntactic phrases. Cognition 51(3):237–265.
Gout A, Christophe A, Morgan J (2004) Phonological phrase boundaries constrain lexical access: II. Infant data. J Mem Lang 51:548–567.
Hruska C, Alter K, Steinhauer K, Steube A (2001) Misleading dialogs: Human's brain reaction to prosodic information. In: Orality and gestures. Multi models of interactions and behaviors in the communication (Cave C, Guaitella I, Santi S, eds), pp 425–430. Paris: L'Hartmattan.
Isel F, Gunter TC, Friederici AF (2003) Prosody-assisted head-driven access to spoken German compounds. J Exp Psychol Learn Mem Cogn 29:277–288.
Isel F, Alter K, Friederici AD (2005) Influence of prosodic information on the processing of split particles: ERPs evidence from spoken German. J Cogn Neurosci 17:154–167.
Ischebeck AK, Friederici AD, Alter K (2008) Processing prosodic boundaries in natural and hummed speech: An fMRI study. Cereb Cortex 24(18):541–552.
Kaan E, Harris A, Gibson E, Holcomb P (2000) The P600 as an index of syntactic integration difficulty. Lang Cogn Proc 15:159–201.
Kjelgaard MM, Speer SR (1999) Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. J Mem Lang 40:153–194.
Knösche TR, Neuhaus C, Haueisen J, Alter K, Maess B, Witte OW, Friederici AD (2005) The perception of phrase structure in music. Hum Brain Mapp 24:259–273.
Kutas M, Hillyard SA (1980) Reading senseless sentences: Brain potentials reflect semantic incongruency. Science 207:203–205.
Lin M (2000) Breaks and prosodic phrases in the utterances of standard Chinese. Contemp Linguist 2:210–217.
Marques C, Moreno S, Castro SL, Besson M (2007) Musicians detect pitch violation in a foreign language better than non-musicians: behavioral and electrophysiological evidence. J Cogn Neurosci 19:1453–1463.
Millotte S, Wales R, Christophe A (2007) Phrasal prosody disambiguates syntax. Lang Cogn Proc 22:898–909.
Millotte S, Rene A, Wales R, Christophe A (2008) Phonological phrase boundaries constrain the online analysis of spoken sentences. J Exp Psychol Learn Mem Cogn 34(4):874–885.
Nagel HN, Shapiro LP, Tuller B, Nawy R (1996) Prosodic influences on the resolution of temporary ambiguity during on-line sentence processing. J Psycholinguist Res 25:319–344.
Nan Y, Knösche TR, Friederici AD (2006) The perception of musical phrase structure: A cross-cultural ERP study. Brain Res 1094:179–191.
Nan Y, Knösche TR, Zysset S, Friederici AD (2008) Cross-cultural music phrase processing: an fMRI study. Hum Brain Mapp 29:312–328.
Neuhaus C, Knösche TR, Friederici AD (2006) Effects of musical expertise and boundary markers on phrase perception in music. J Cogn Neurosci 18:1–22.
Nespor M, Vogel I (1986) Prosodic phonology. Dordrecht, Holland: Foris.
Nieuwenhuis S, Aston-Jones G, Cohen JD (2005) Decision making, the P3, and the locus coeruleus-norepinephrine system. Psychol Bull 131(4):510–532.
Pannekamp A, Toepel U, Alter K, Hahne A, Friederici AD (2005) Prosody-driven sentence processing: an event-related brain potential study. J Cogn Neurosci 17:407–421.

Patel AD (2003) Language, music, syntax and the brain. Nat Neurosci 6:674–681.
Salverda AP, Dahan D, McQueen JM (2003) The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. Cognition 90:51–89.
Schafer AJ, Speer SR, Warren P, White SD (2000) International disambiguation in sentence production and comprehension. J Psycholinguist Res 29:169–182.
Schirmer A, Kotz SA, Friederici AD (2002) Sex differentiates the role of emotional prosody during word processing. Cogn Brain Res 14:228–233.
Schon D, Magne C, Besson M (2004) The music of speech: music training facilitates pitch processing in both music and language. Psychophysiology 41:341–349.
Selkirk E (1980) Prosodic domains in phonology: Sanskrit revisited. In: Juncture (Aronoff M, Kean M-K, eds). Saratoga, CA: Anma Libri.
Shattuck-Hufnagel S, Turk AE (1996) A prosody tutorial for investigations of auditory sentence processing. J Psycholinguist Res 25(2):193–246.
Shen J (1985) Pitch range and intonation of Chinese tone. In: Experimental phonetics of Chinese [in Chinese] (Ling D, Wang LJ, eds), pp 75–107. Beijing: Pecking University Press.
Speer SR, Kjelgaard MM, Dobroth KM (1996) The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. J Psycholinguist Res 25:249–271.
Steinhauer K, Alter K, Friederici AD (1999) Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nat Neurosci 2:191–196.
Steinhauer K, Friederici AD (2001) Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. J Psycholinguist Res 30(3):267–295.
Steinhauer K (2003) Electrophysiological correlates of prosody and punctuation. Brain Lang 86:142–164.
Toepel U, Pannekamp A, Alter K (2007) Catching the news: Processing strategies in listening to dialogs as measured by ERPs. Behav Brain Funct 3(53):1–13.
Wang B, Yang Y (2002) Acoustic correlates of hierarchical prosodic boundary in mandarin. Proceedings of Speech Prosody, Aix en Provence, 11–13.
Wang B, Lü S, Yang Y (2004) Acoustic analysis on prosodic hierarchical boundaries of Chinese. Acta Acoust 29:29–36.
Wightman CW, Shattuck-Hufnagel S, Ostendorf M, Price PJ (1992) Segmental durations in the vicinity of prosodic phrase boundaries. J Acoust Soc Am 91:1707–1717.
Yang Y (1997) Prosodic cues to syntactic boundaries. Acta Acoust 22:414–421.
Yang Y (1998) Perception of prosodic structures of sentences. Acta Acoust 23:163–169.

(Accepted 17 October 2008)
(Available online 24 November 2008)