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 sen-