Cogn Neurodyn (2017) 11:467–481
DOI 10.1007/s11571-017-9442-4

RESEARCH ARTICLE

Attention is shaped by semantic level of event-structure during speech comprehension: an electroencephalogram study

Xiaoqing Li[1] · Yuping Zhang[1,2] · Lin Li[1,2] · Haiyan Zhao[1,2] · Xiufang Du[3]

Received: 25 October 2016 / Revised: 4 May 2017 / Accepted: 31 May 2017 / Published online: 17 June 2017

Abstract The present electroencephalogram study used an attention probe paradigm to investigate how semantic and acoustic structures constrain temporal attention during speech comprehension. Spoken sentences were used as stimuli, with each one containing a four-character critical phrase, of which the third character was the target character. We manipulated not only the semantic relationship between the target character and the immediately preceding two characters, but also the presence/absence of a pitch accent on the first character. In addition, an attention probe was either presented concurrently with the target character or not. The results showed that the N1 effect evoked by the attention probe was of larger amplitude and started earlier (enhanced attention) when the target character and the preceding two characters belonged to the same semantic event than when they spanned a semantic-event boundary, and this effect occurred only in the un-accented conditions. The results suggest that, during speech comprehension, the semantic level of event-structure can constrain attention allocation along the temporal dimension, and reverse the attention attenuation effect of prediction; meanwhile, the semantic and acoustic levels of event-structure interact with each other immediately to modulate auditory-temporal attention. The results were discussed with regard to the predictive coding account of attention.

Keywords Speech processing · Temporally selective attention · Event-structure · Predictive coding theory

Introduction

Speech signals convey a large amount of information and unfold rapidly in time, presenting significant challenges to auditory perception and comprehension. During speech comprehension, listeners need not only to select a target speech from the multiple simultaneous speech streams, but also to determine which time points of the selected speech stream should be processed in detail. Considerable research has demonstrated that the time sequences of speech signals are not processed equally, and appropriate allocation of attention along the temporal dimension (namely, temporal attention) facilitates speech perception or comprehension (e.g., Astheimer and Sanders 2009; Hruska et al. 2000; Magne et al. 2005; Wöstmann et al. 2016). With regard to temporal attention in speech processing, previous studies have focused mainly on how acoustic cues, such as prosody, modulate attention allocation. There is still no clear picture of how attention is modulated by the semantic relationship between elements of the speech sequences, and how the semantic and acoustic levels of information interact with each other in guiding attention. Thus, the present study aimed to investigate how the semantic and acoustic relationship of speech sequences modulates temporal attention while speech signals unfold in time.

✉ Xiaoqing Li
lixq@psych.ac.cn

✉ Xiufang Du
dxflxc@163.com

[1] CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, No. 16 Lincui Road, Chaoyang District, Beijing 100101, People's Republic of China

[2] University of Chinese Academy of Sciences, Beijing, People's Republic of China

[3] School of Psychology, Shandong Normal University, No. 88, East Wenhua Road, Jinan 250014, Shandong Province, People's Republic of China

In the field of psycholinguistics, quite a lot of research has investigated how listeners' selective attention is guided by the acoustic cues of speech signals. It has found that, salient acoustic cues, such as accentuation or word onsets, tend to attract more attention resources during speech processing (Astheimer and Sanders 2009; Cutler 1976; Li and Ren 2012; Li et al. 2014; Sanford et al. 2006). Accentuation, or pitch accent, is one type of prosodic information in the speech signal, which reflects the relative prominence of a particular syllable, word, or phrase in an utterance realized mainly by modulations of pitch (Shattuck-Hufnagel and Turk 1996). Using phoneme-monitoring task, Cutler (1976) found heightened attention (as indicated by faster phoneme monitoring responses) to a word that received a pitch accent. Li and colleagues' study also demonstrated that semantically incongruent words elicited a larger N400 than the semantically congruent words when the corresponding words were accented, but not when they were de-accented, indicating that listeners allocate more attention to the accented information and engage in deeper processing (Li and Ren 2012). Meanwhile, Astheimer and Sanders found that listeners direct more attention to acoustically salient word onsets, as suggested by the fact that the linguistic attention probe 'ba' presented at the word onset elicited a larger N1 than probes presented at other time points (Astheimer and Sanders 2009). Furthermore, during auditory processing, selective attention not only tends to be attracted by transient and salient acoustic cues but also is synchronized with the relatively long-term acoustic fluctuation of auditory stimuli, as indicated by an MEG (magnetoencephalography) study (e.g., Wöstmann et al. 2016). In this MEG study, participants attended to spoken digits presented at a rate of 0.67 Hz to one ear and ignored tightly synchronized distracting digits presented to the other ear. The hemispheric lateralization of alpha power in auditory cortical regions, which reveals the amount of attention allocated, is modulated in tune with the speech rate of spoken digits (Wöstmann et al. 2016). Overall, the above studies suggest that as the speech signals unfold in time, listeners' attention changes dynamically in pace with the short- or long-distance acoustic variations of speech signals.

The studies mentioned above mainly focus on the effect of the lower-level acoustic variations on selective attention during speech processing. Less is known about how the higher-level semantic information modulates temporal attention. The constituents of speech signals are organized along the temporal dimension, and their structure in times is critical. The dynamic attending theory (Jones 1976; Jones and Boltz 1989) has already proposed that, different from visual stimuli which are grouped into objects in two-dimensional or three-dimensional space, the constituents of speech signals are grouped into events, namely, patterns in time; most importantly, during speech processing, an event-structure along the temporal dimension shapes attention allocation. Speech sequences are grouped into events not only at the acoustic-sensory level but also at the higher knowledge or semantic level. For example, a study conducted by Billig and colleagues show that the lexical knowledge can influence how many streams of speech a listener perceives (Billig et al. 2013). An interesting question arising here is that, during speech processing, whether and how attention along the temporal dimension is modulated by the semantic level of event-structure, e.g., by the semantic relationship between the adjacent speech sequences.

An ERP study, which used the same attention probe paradigm as in the study conducted by Astheimer and Sanders (2009), has investigated how semantic prediction modulates listener's attention during speech comprehension (Li et al. 2014). In this ERP study, each pair of sentences had the same critical word (namely, a two-character word) (e.g., *ROSES, mei-gui* in Chinese) but different sentence contexts (e.g., *On Valentines Day, Xiao Li bought…* vs. *On holidays, Xiao Li bought…*), hence the same critical word being strongly or weakly predictable. The researchers also manipulated the presence/absence of an attention probe 'ba' on the critical word. The results revealed that the latency of the probe-related N1 effect was shortened for weakly predictable words as compared with strongly predictable ones, indicating that more attentional resources were allocated to unpredictable words during speech processing (Li et al. 2014). Some fMRI or MEG studies have examined how semantic predictability influences early perceptual processing and have revealed that predicted stimuli evoke reduced neural responses in the sensory cortex (Alink et al. 2010; den Ouden et al. 2010; Todorovic et al. 2011; Sohoglu et al. 2012). This sensory attenuation of predicted signals is in line with Li and colleagues' (Li et al. 2014) finding that listeners direct less attention to predicted words. Yet, in the study conducted by Li and colleagues, both the strongly- and weakly predictable words are complete and independent lexical-words, and the semantic relationship between the critical word (e.g., *ROSES*) and the immediately preceding word (e.g., *bought*) is exactly the same in the strong- and weak-prediction conditions; the degree of predictability is caused by the overall preceding context meaning. Besides the overall sentence context, the semantic relationship between the immediately adjacent constituents is also an important source of semantic structure. For example, in Mandarin Chinese, although the majority of words (namely, the basic semantic unit that can't be separated further) consist of two characters, some words consist of three or four characters. Therefore, when the listeners hear a character during speech comprehension, this character either is the onset of a new word (that means, sets a word boundary) or belongs to the same word with one, two, or even three characters preceding it (i.e., in a within-word position). It is still unclear whether and how selective attention is modulated by the semantic relationship between the adjacent character-based speech signals as they unfold in time.

Recently, the predictive coding theory has become a highly influential theory of perceptual processing. It proposes that attention operates to optimize the precision of perceptual inference, causing prediction errors (or, equivalently, sensory data) to be weighted (Knill and Pouget 2004; Friston 2005, 2009; Feldman and Friston 2010; Hesselmann et al. 2009; Hohwy 2012; Rao 2005; Saada et al. 2014). Both studies related to general sensory processing and those related to speech processing have demonstrated that prediction reduces sensory activities (e.g., Alink et al. 2010; den Ouden et al. 2010; Todorovic et al. 2011; Sohoglu et al. 2012), and thus have provided supporting evidence for the predictive coding theory. However, another line of studies has found that prediction sometimes seems to enhance rather than reduce sensory activities (e.g., Chaumon et al. 2008; Doherty et al. 2005; Koch and Poggio 1999; Rauss et al. 2011). A further study (Kok et al. 2012) explained the above opposite effects by showing that sensory data (namely, prediction error) is weighted according to how informative it is (i.e., according to the current predictions and tasks). Specifically, this study found that when the visual target stimuli are closely related to the current task (in the task-related visual hemisphere), the neural response in early visual cortex is larger in amplitude for the predicted stimuli compared with the unpredicted ones, since the predicted stimuli are highly informative for the current task; in contrast, when the visual target stimuli are not related to the current task (in the task-irrelated visual hemisphere), the neural response to the predicted stimuli is reduced in early visual cortex. The authors claimed that both the sensory-reduction and sensory-enhancement effects are in line with the predictive coding theory wherein prediction and attention operate to improve the precision of perceptual inference (Kok et al. 2012).

How does the semantic level of event-structure modulate selective attention during speech processing? For speech signals, when the adjacent characters belong to the same basic semantic event (such as, a lexical word), the character emerging later in time is usually strongly predictable as compared with the character at the onset boundary of a new semantic event. Therefore, as to how a semantic event structure constrains attention allocation along the speech sequences, there are two possibilities. The first one is that, during speech processing, the character with a higher level of predictability reduces attention, regardless of its semantic relationship with the immediately preceding characters. Accordingly, attention would be reduced at the character that is within a basic semantic event (compared with that spans a semantic-event boundary), as the character is usually strongly predictable in the within-event condition. That is, we would observe the same effect as in the study conducted by Li and colleagues (c.f. Li et al. 2014). Another possibility is that the semantic relationship between the adjacent speech sequences plays an important role in attention allocation, and may reverse the prediction-related sensory/attention reduction effect. Specifically, attention enhancement, instead of attention reduction, might be observed in the within-semantic-event condition (compared with the between-semantic-event condition), as the target character in the within-semantic-event condition is important for the understanding of the yet complete semantic event, hence its sensory information being weighted.

In addition, the studies mentioned above have already demonstrated that, during speech processing, accentuation, as a salient acoustic cue, tends to attract more attention resources (e.g., Cutler 1976; Li and Ren 2012). More importantly, this bottom-up acoustic cue can modulate the effect of semantic predictability on temporal selective attention, suggesting that attention reduction for the strongly predictable word was observed over the unaccented words but not over accented words (Li et al. 2014). For speech signals, once a moment is accented and consequently more salient, the subsequent moment following the pitch accent usually becomes less salient (indicated by pitch or/and intensity decreases), which is called post-focus compression (namely, post-accentuation compression) (Wang 2012). We still do not know whether attention allocation is affected by the acoustic relationship, such as post-accentuation compression. Moreover, it is still unclear, as speech signals unfold in time, whether and how the semantic relationship interacts with the post-accentuation-compression acoustic relationship in the process of selective attention.

Therefore, the aim of the present study is to further investigate whether and how the semantic level of event structure modulates attention allocation during speech comprehension, and whether and how semantic and acoustic levels of event-structure interact with each other in modulating temporal selective attention.

To study selective attention in the auditory-temporal domain, EEG and an attention probe paradigm were used in the present study. Whether selected on the basis of location or of time, attended transient auditory stimuli elicit an N1 that increases in amplitude or decreases in latency (Astheimer and Sanders 2009; Hink and Hillyard 1976; Hillyard et al. 1973; Näätänen and Winkler 1999; Folyi et al. 2012; Lagemann et al. 2010) relative to the unattended one. However, not all portions of the unfolding speech signal carry abrupt acoustic changes. Therefore, we used an auditory-temporal variant of the Posner probe paradigm (Posner 1980; Astheimer and Sanders 2009, 2011), in which an auditory probe was superimposed on the different time points of the speech signal (see Li et al. 2014 for detailed description of this paradigm). We subtracted the ERP waveforms elicited by the critical moment without a probe from those elicited by the same moment with a probe, which isolated the N1 effect triggered by the attention probe and canceled out the ERP effects elicited by the target moment itself. Based on previous studies, the N1 elicited by the probe is considered as a correlate of focal attention, with the enhancement of N1 amplitude or/and shortening of N1 latency reflecting more attention allocated to the corresponding time point.

Mandarin Chinese is an ideal language that can be used to examine the effect of semantic event-structure on attention allocation. In Mandarin Chinese, a single character is not usually used independently as a word, and the majority of words (around 87.8%) are compound words that consist of more than one character (e.g., two-character words, three-character words, or four-character words) (A frequency dictionary of Modern Chinese, Beijing Language Institute 1986). In the present study, the experimental materials were isolated Chinese spoken sentences, each of which included a critical phrase. The critical phrase always consisted of four characters, and the third character was the target character where we measured attention. We manipulated both the acoustic level and the semantic level of an event structure. The critical phrase consisted of one Chinese four-character idiom or two two-character words; consequently, the target character and the preceding two characters belonged to the same semantic event in the case of idiom, but spanned a semantic event boundary in the case of two two-character words (Semantic structure: within-event vs. between-event). Moreover, the first character of the critical phrase was either accented or un-accented (Accentuation: accented vs. un-accented); therefore, relative to the un-accented condition, the first character was more acoustically salient in the accented condition; the target character (namely, the third character), however, was less acoustically salient in the accented condition due to the effect of post-focus compression (Wang 2012). In addition, a linguistic attention probe ‘ba’ was either presented concurrently with the target character or not (Probe: with-probe vs. without-probe). To measure attention directed to the target character, we subtracted the ERP waveforms elicited by the target characters without a probe from those elicited by the same characters with a probe, which isolated the N1 effect triggered by the attention probe. It is important to note that, before the target character appeared, the speech signals had already diverged between the accented and un-accented conditions. By comparing the difference ERP waveforms (namely, the probe-related N1 effect), instead of the original ERP waveforms, we could not only measure the attention directed to the target character but also avoid the potential confounding effect caused by the differences in the preceding baseline window.

We predicted that if the semantic level of event-structure modulates attention allocation and reverses the traditional sensory/attention attenuation effect of prediction, the attention allocated to the target moment will be enhanced in the within-event condition as compared with the between-event condition, reflecting as an earlier or larger probe-related N1 effect in the within-event condition. In contrast, if the high predictability of the target character reduces attention in spite of the relatively closer semantic relationship between this character and the immediately preceding characters, the attention directed to the target moment will be reduced in the within-event condition (compared with the between-event condition), reflecting as an earlier or smaller probe-related N1 effect in the within-event condition. In addition, by examining the two-way interaction between Accentuation and Semantic structure, we would know how acoustic and semantic level of event-structures interact with each other in modulating temporal selective attention.

Methods

Ethics statement

All methods were carried out in accordance with relevant guidelines and regulations. All experimental protocols were approved by the Ethics Committee of Institute of Psychology, Chinese Academy of Sciences. All participants were over 18 years of age and gave written informed consent. They were notified that their participation was completely voluntary and that they can secede at any time.

Participants

Twenty-four right-handed university students (10 males), all of whom were native Mandarin Chinese speakers, participated in this experiment. The mean age was 24 years (range 20–27). None reported any medical, neurological, or psychiatric illness, and all gave informed consent. The data of 4 participants (two male) were removed from analysis because of excessive artifacts.

Experimental material

In the present study, 140 pairs of spoken Mandarin Chinese sentences were constructed, with each sentence including a critical phrase. All of the critical phrases consist of four characters, and the third character is the target character where we would measure attention. First of all, we manipulated the semantic event-structure of the critical phrase. That is, each pair of sentence had the same sentence frame, but different critical phrases: the critical phrase consisted of one Chinese four-character idiom or two two-character words (see Table 1). A Chinese four-character idiom is a kind of concise phrase that carries meanings much more than the sum of the four characters. It has been in use for a long time and has a fixed structure. Neither the characters nor their order could be changed. Some of the four-character Chinese idioms have developed into a kind of compound words, since their constituents can't be used independently in modern Chinese[34]. For example, the Chinese idiom '*大张旗鼓*' means '*someone does something with a great fanfare or on a grand scale'*. The first two characters *大张* do not constitute a real word in modern Chinese and do not convey a specific meaning. Only when *大张* are combined with *旗鼓* (namely, *大张旗鼓*), can it mean *on a grant scale*. Another example is *情不自禁*, which means *can't help oneself or can't control one's feelings*. The character *情*, *自*, or *禁* from *情不自禁* is seldom used alone in modern Chinese. More importantly, neither the first two characters *情不* nor the last two characters *自禁* can constitute a real word, and the four characters *情不自禁* must be used as a whole. In the present study, all of the four-character idioms were selected under the requirement that the first two characters in the idiom did not constitute a word and must not be used independently, hence not conveying specific meaning. Meanwhile, in Mandarin Chinese, a single character is seldom used independently as a word. Given the characteristics of the idioms used in the present study and the features of Mandarin Chinese, the four-character idioms here can be considered as a kind of compound words (Sun 1990), which must be used as a whole and can't be further separated. That is, when the listeners just hear the first two characters, they will know these characters do not constitute a word and have to wait for further incoming language signals since a word is a basic semantic unit of language. Therefore, in the present study, the four characters of the critical phrase were of different semantic levels of event-structure in the 'one idiom' and 'two two-character words' conditions. The target character (namely, the third character) and the preceding two characters belonged to the same semantic event in the idiom condition (within-event), but spanned a semantic event boundary in the two two-character words condition (between-event).

In addition, we manipulated the acoustic structure of the critical phrases. The 140 pairs of sentences were spoken by a female speaker and recorded at a sampling rate of 22,050 Hz. The first character of the four-character critical phrase was either accented or un-accented. Therefore, the first character was more acoustically salient in the accented condition than in the un-accented condition; in contrast, the target character (namely, the third character) was less acoustically salient in the accented condition than in the un-accented condition, due to the effect of post-focus compression (Wang 2012). In addition, the target character was added either with or without a linguistic attention probe. The linguistic attention probe was created by extracting a 50 ms excerpt of the narrator pronouncing of the syllable ''ba'' that was spoken with a light tone. This probe was added to 50 ms after the acoustic onset of the target characters. The probe had an intensity of 48 dB in accented condition and an intensity of 51 dB in the un-accented condition, since the intensity of the target characters in the accented condition was almost 3 dB lower than that in the un-accented condition (see the next paragraphs for the acoustic analysis of the critical phrases). Taken together, this resulted in a full factorial design with all combinations of the factors, namely, Semantic structure (within-event vs. between-event), Accentuation (accented

Table 1 Illustrations for the experimental materials used in the present study

Conditions	Example sentences
Within-event; accented	After hearing the story, the students "could not help themselves" and cried 同学们听完故事后都"*情*不自禁"哭了起来
Within-event; un-accented	After hearing the story, the students "could not help themselves" and cried 同学们听完故事后都"情不自禁"哭了起来
Between-event; accented	After hearing the story, the students "blamed themselves" and cried 同学们听完故事后都"*谴*责自己"哭了起来
Between-event; un-accented	After hearing the story, the students "blamed themselves" and cried 同学们听完故事后都"谴责自己"哭了起来

Quotes indicate the critical phrase. As seen in the Chinese version of example sentences, the bold and italic indicates the presence of the pitch accent on the first character of the critical phrase; the underline indicates the target character whether we measure attention vs. un-accented), and Probe (with-probe vs. without-probe) (see Table 1 for example sentences).

When constructing the experimental sentences, we also controlled the following confounding factors. First, all of the critical phrases were not in the sentence-final position. Second, for each pair of sentences, the third character of the critical phrases was exactly the same in both the within- and between-event conditions. Third, in Mandarin Chinese, the word '*shi…*' can be used as syntactic focus-marker, which means '*it is…that*'. For the present materials, except for pitch accent on the first character, there was no other marker of focus, such as '*shi…*', in the sentences.

To confirm that the idioms used in the present study are the smallest semantic event or unit, we conducted a word-independence pre-test by presenting the first two characters of the idioms (namely, in the within-event condition) and the first two-character word in the between-event condition. 20 participants who didn't attend the EEG experiment and other pre-tests were instructed to mark the possibility that the two characters could be used independently as a word on a 7-point scale (from −3 to 3). The larger the score is, the larger the possibility is. The mean scores for the within-event condition and between-event condition were −0.96 (STDEV = 0.94) and 2.65 (STDEV = 0.55) respectively. The paired-T test revealed that the rating score in the within-event condition was significantly smaller than that in the between-event condition ($t_{(139)} = -36.53$, $p < .001$). Meanwhile, the rating score in the within-event condition was also significantly smaller than 0 ($t_{(139)} = -10.25$, $p < .001$). The result indicates that the first two characters in the idioms are relatively unlikely to be used independently.

To examine the predictability of the target character in the sentence context, we conducted a cloze probability test by presenting the sentence frames until the first two characters of the critical phrases. 24 participants who didn't attend the EEG experiment and other pre-tests were instructed to fill in the first event that came to their mind and made the sentence meaningful. The close probabilities of the target character were 0.84 (STDEV = 0.23) and 0.10 (STDEV = 0.20) for the within-event and between-event conditions respectively. The paired-T test revealed that the cloze probability of the target characters in the within-event condition was significantly higher than that in the between-event condition ($t_{(139)} = 31.84$, $p < .001$).

To confirm that the critical phrases were congruent in both within- and between-event conditions, we conducted a congruency pre-test by presenting written sentences up until the critical phrases. 24 participants who didn't attend the EEG experiment and other pre-tests were instructed to mark the semantic congruence of the last phrase, namely the last four characters, in each sentence on a 5-point scale (from 1 to 5). The larger the score was, the more congruent the last phrase was. The mean scores for the within- and between-event conditions were 3.74 (STDEV = 0.69) and 3.87 (STDEV = 0.69) respectively. The paired-T test revealed that the semantic congruence of the critical phrases in the within-event condition was not different from that in the between-event condition ($t_{(139)} = -1.44$, $p = .152$).

To ensure that our speaker had succeeded in correctly accenting the first character, ANOVAs were performed on the corresponding acoustic measurements, with Semantic structure (within-event vs. between-event) and Accentuation (accented vs. un-accented) as independent factors. Mandarin Chinese is a tone language and the pitch correlate of the lexical tone is not a single point but a pitch contour, which is called pitch register. Previous studies have found that, in Mandarin Chinese, a focus accent is realized by lengthening the syllable duration and by expanding the pitch range of the pitch register, with the latter mainly resulting from raising of the pitch maximum (e.g., Chen 2006; Jia et al. 2006; Wang et al. 2002). Therefore, the dependent factors for the ANOVAs here were pitch maximum, pitch range, and syllable duration. Although intensity is not a reliable acoustic parameter for accentuation, we still added intensity as an additional dependent factor. The results of the ANOVAs (with duration, intensity, pitch maximum, or pitch range of the first character as dependent factors) revealed a significant main effect of Accentuation ($F_{(1,\ 139)} = 1318.29$, $p < .001$; $F_{(1,\ 139)} = 655.79$, $p < .0001$; $F_{(1,\ 139)} = 360.40$, $p < .001$; $F_{(1,\ 139)} = 133.92$, $p < .001$ for duration, intensity, pitch maximum, and pitch range respectively), indicating that there were significant increases in syllable duration, intensity, pitch maximum, and pitch range expansion from un-accented characters to accented characters. Importantly, the two-way interaction between Accentuation and Semantic structure failed to reach significance (all ps > .3). The acoustic measurements of the first characters confirmed that the first character in the experimental sentences was spoken with the intended accentuation pattern (see Fig. 1a and c).

$$\text{Pitch maximum} = 12\log_2(\text{Maximum Pitch}/100).$$

$$\text{Pitch range} = 12\log_2(\text{Maximum Pitch}/\text{Minimum Pitch}).$$

We also measured the acoustic parameters (duration, intensity, pitch maximum, and pitch range) of the target character, namely the third character of the critical phrase. Although the ANOVAs with duration or pitch range as dependent factors found neither main effects of Semantic structure (or Accentuation) nor interaction between them (with the smallest p value being 0.135), those with intensity and pitch maximum as dependent factors found that the character following the accented character reduced both

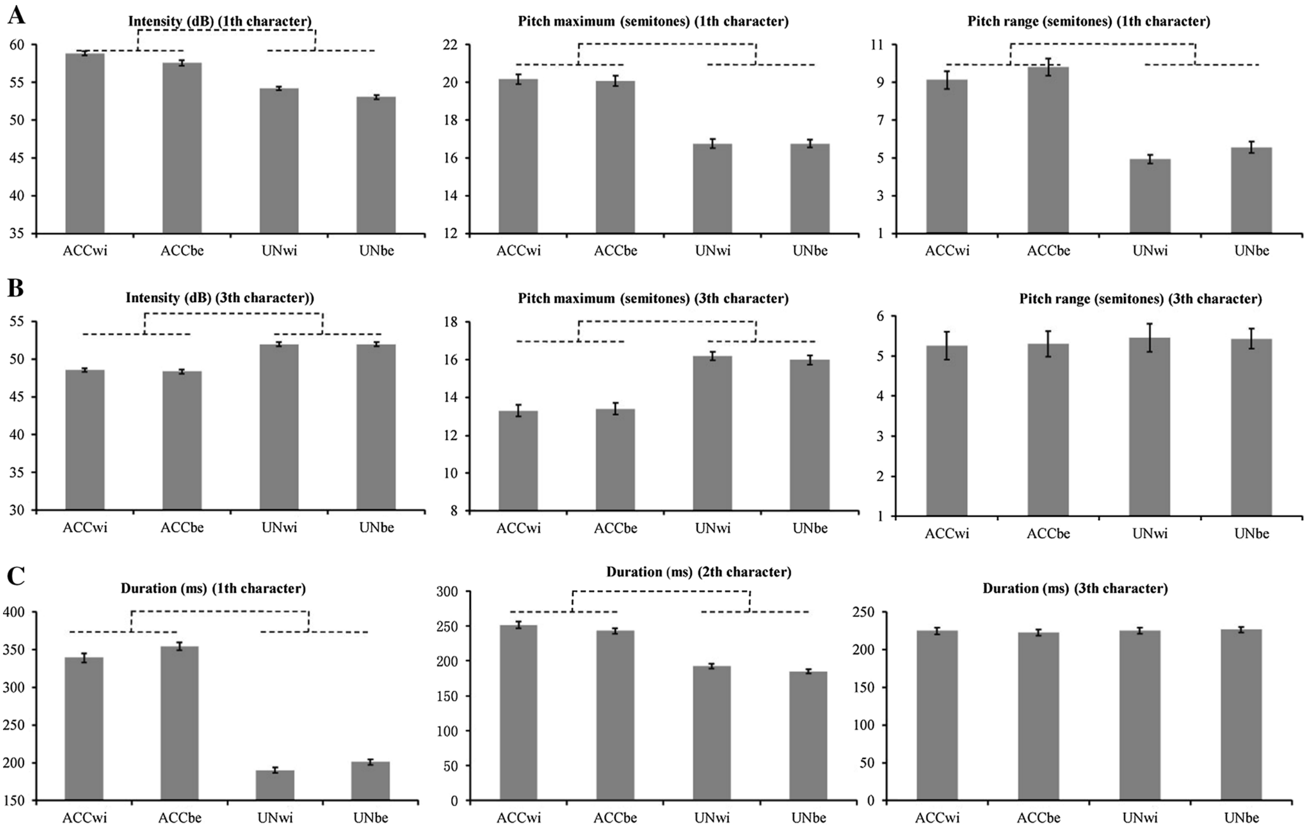

Fig. 1 The acoustic parameters (duration, intensity, pitch maximum, and pitch range) of the critical phrases in the four experimental conditions: ACCwi indicates 'within-event, accented'; ACCbe indicates 'between-event, accented'; UNwi indicates 'within-event, un-accented'; UNbe indicates 'between-event, un-accented'. **a** the acoustic parameters of the first character of the critical phrase; **b** the acoustic parameters of the third character (namely, the target character) of the critical phrase; **a** the duration of the first (*left*), second (*middle*), and third (*right*) characters of the critical phrase

pitch maximum ($F_{(1,\ 139)} = 118.53$, $p < .001$) and intensity ($F_{(1,\ 139)} = 444.23$, $p < .001$) as compared with that following the un-accented character (see Fig. 1b, c). That is, the target character was less salient in the accented condition than in the un-accented condition.

The Fig. 1c describes in detail the duration of the target character (the third character) and the immediately preceding two characters of the critical phrase. The ANOVA with the duration of the second character as dependent factor showed that, the main effect of Accentuation reached significance ($F_{(1,\ 139)} = 512.34$, $p < .001$, longer duration in the accented condition relative to the un-accented condition), whereas both the main effect of Semantic structure and the interaction between the two factors didn't reach significance (with the smallest p value being 0.085), which is in line with the result of the ANOVA conducted over the duration of the first character. In contrast, the ANOVA with the duration of the third character found neither the main effect of Semantic structure (or Accentuation) nor interaction between them.

To confirm that the appropriateness of the presence of pitch accent on the first character was the same between the between- and within-event conditions, 24 participants who didn't attend the EEG experiment and other pre-tests were instructed to mark whether the sentences were spoken appropriately on a 7-point scale (from 1 to 7). The larger the score was, the more appropriately the sentence was spoken. The mean scores for the within- and between-event conditions were 4.97 (STDEV = 0.42) and 4.89 (STDEV = 0.44) respectively. The ANOVA with Semantic structure and Accentuation as independent factors revealed that neither the main effect of Semantic structure nor the two-way Semantic structure × Accentuation interaction reached significance (with smallest p value being 0.09). The rating results indicated that the sentences in the within-event condition were spoken as appropriately as the sentences in the between-event condition.

The experimental materials (140 sets, with each set including 8 versions of sentences) were grouped into 4 lists of 280 sentences according to the Latin square procedure based on the four experimental conditions (combination of Accentuation and Probe). That is, the within-event sentence and the between-event sentence coming from the same stimuli set were included in the same list, since they had different sentence-level meanings. Consequently, the same target character was presented twice in each list, because the sentences in the within- and between-event conditions had the same target characters. In each list, there were an equal number of sentences (35 sentences) for each of the eight experimental conditions and additional 120 filler sentences. Subjects were divided into 4 groups, with each group listening to only one list of materials. Meanwhile, the whole list of 400 sentences (280 experimental sentences and 120 filler sentences) was divided into four blocks, with the first and the second presentations of the same target characters being separated by one block; and the order of the two presentations of the same target characters was counterbalanced between subjects.

Experimental procedure

After the electrodes were positioned, subjects were asked to listen to each sentence for comprehension. Meanwhile, their EEG signals were recorded. After finishing 40 percent (namely, 160 sentences) of the overall sentences (400 sentences: 280 experimental sentences and 120 filler sentences) in each list, the subjects were asked to judge the correctness of a question sentence regarding the meaning of the sentence just heard. Each trial consisted of a 300 ms auditory warning tone, followed by 700 ms of silence and then the target sentence. To inform subjects of when to fixate and sit still for EEG recording, an asterisk was displayed from 500 ms before the onset of the sentence to 1000 ms after its offset. After a short practice session that consisted of 10 sentences, the trials were presented in four blocks of approximately 13 min each, separated by brief resting periods.

EEG acquisition

EEG was recorded (0.05–100 Hz, sampling rate 500 Hz) from 64 Ag/AgCl electrodes mounted in an elastic cap, with an on-line reference linked to the left mastoid and off-line algebraic re-reference linked to the left and right mastoids. EEG and EOG data were amplified with AC amplifiers (Synamps, Neuroscan Inc.). Vertical eye movements were monitored via a supra- to sub-orbital bipolar montage. A right-to-left canthal bipolar montage was used to monitor horizontal eye movements. All electrode impedance levels (EEG and EOG) were kept below 5 kΩ.

ERP analysis

We analyzed the ERPs time-locked to the target character (namely, the third character) in the eight conditions to examine how attention is allocated at the target character. The raw EEG data were first corrected for eye-blink artifacts using the ocular artifact reduction algorithm in the Neuroscan v. 4.3 software package. Then, the EEG data was filtered with a band-pass filter 0.1-40 Hz. Subsequently, the filtered data were divided into epochs ranging from 100 ms before the acoustic onset of the target characters to 1000 ms after the acoustic onset of these characters. A time window of 100 ms preceding the onset of the target characters was used for baseline correction. Trials contaminated by eye movements, muscle artifacts, electrode drifting, amplifier saturation, or other artifacts were identified with a semiautomatic artifact rejection (automatic criterion: signal amplitude exceeding ±75 μV, followed by a manual check). Trials containing the abovementioned artifacts were rejected (about 13% overall). Rejected trials were evenly distributed among conditions. Finally, averages were computed for each participant, each condition, and at each electrode site before grand averages were calculated across all participants.

Statistical analysis

For statistical analysis, the cluster-based random permutation test implemented in the Fieldtrip (http://fieldtrip.fcdonders.nl) software package (Maris and Oostenveld 2007) was used. This non-parametric statistical procedure optimally handles the multiple-comparisons problem. With ERP amplitude as the dependent factor, the cluster-based random permutation test was performed within 0–400 ms post-character onset (in a step of 2 ms) over 60 electrodes (PO7 and PO8 were deleted, as the electrode PO8 was linked to the scalp position of right mastoid during online EEG acquisition, and consequently was used as offline re-reference). For every data point (electrode by time) of two conditions, a simple dependent-samples *t* test was performed. All adjacent data points exceeding a preset significance level ($p < 0.05$) were grouped into clusters. Cluster-level statistics were calculated by taking the sum of the *t*-values within every cluster. The significance probability of the clusters was calculated by means of the so-called Monte Carlo method with 1000 random draws.

We first analyzed the main effect of Attention Probe to examine whether an N1 effect had been elicited by the attention probe. That is, the four conditions with an attention probe were combined together (With-probe), and the other four conditions without an attention probe were combined together (Without-probe). We calculated the ERPs in With-probe and Without-probe conditions separately, and the cluster-based random permutation test was applied to these two conditions (With-probe vs. Without-probe) with amplitude (0–400 ms in a step of 2 ms) as dependent factor. If the N1 effect evoked by the attention probe reached significance, we subtracted the ERP waveforms elicited by the target characters without a probe from those elicited by the same characters with a probe. This resulted in four conditions of difference waveforms (ACCwi: within-event and accented; UNwi: within-event and un-accented; ACCbe: between-event and accented; UNbe: between-event and un-accented). Then, further permutation tests were performed with amplitude of the difference waveforms (0–400 ms in a step of 2 ms) as a dependent factor.

To examine the two-way Semantic structure × Accentuation interaction, the cluster-based random permutation tests were performed to compare the amplitude of the two difference-difference waves (ACCwi-minus-UNwi vs. ACCbe-minus-UNbe), with significant difference indicating the presence of two-way Semantic structure × Accentuation interaction. If this two-way interaction reached significance or marginal significance, we further examined the simple effects of Semantic structure at the different levels of Accentuation (namely, 'ACCwi vs. ACCbe' and 'UNwi vs. UNbe'). Otherwise, the main effect of Semantic structure would be analyzed. To examine this main effect, we calculated ERPs in the 'Within-event' (WITHIN: 'ACCwi' combined with 'UNwi') and "Between-event" (BETWEEN: 'ACCbe' combined with 'UNbe') conditions, and the permutation test was applied to these two conditions (WITHIN vs. BETWEEN).

In addition, the permutation tests were also conducted over the four pairs of 'with-probe' and 'without-probe' conditions (with-probe vs. without-probe) independently to examine the onset latency of the attention probe effect in each of the four experimental conditions (Accentuation by Semantic structure).

Furthermore, traditional ANOVA (analyses of variance) analysis was conducted within the window latency probe-related N1, namely 190–250 ms after the acoustic onset of the target character (140–200 ms post-probe onset). The dependent factor was the mean amplitude obtained from 60 ms-wide time window around the peak of the probe-related N1 effect, namely 190–250 ms post-character onset (140–200 ms post-probe onset). Because we did not know in advance the scalp distribution of the potential N1 effect, analyses of variance were conducted based on the selection of electrodes that represent the three midline areas (midline-anterior 'FZ/FCZ', midline-central 'CZ/CPZ', and midline-posterior 'PZ/POZ') and the six lateral areas (left-anterior 'F5/F3/FC3', right-anterior 'F4/F6/FC4', left-central 'C5/C3/CP3', right-anterior 'C4/C6/CP4', left-posterior 'P5/P3/PO3', and right-posterior 'P4/P6/PO4'). Moreover, in order to test more clearly the potential hemispheric lateralization of the potential N1 effects, ANOVAs were performed separately for the lateral and midline electrodes. To examine whether the N1 effect was elicited by the attention probe, 1-ANOVAs were conducted based on the original ERP waveforms time-locked to the target character in the eight conditions: for the statistical analysis over the midline electrodes, the independent factors were Semantic structure (within-event vs. between-event), Accentuation (accented vs. un-accented), and Anteriority (anterior: FZ/FCZ, central: CZ/CPZ, and posterior: PZ/POZ); for the lateral electrodes, the mean amplitude values were entered into statistics analysis with Hemisphere (left vs. right) as an additional factor and lateral electrodes (F5/F3/FC3; F4/F6/FC4; C5/C3/CP3; C4/C6/CP4; P5/P3/PO3; P4/P6/PO4) nested under Hemisphere. Then, to examine how Semantic structure and Accentuation modulated the probe-related N1 effects, 2-ANOVAs were performed based on the difference waveforms time-locked to the target character: over the midline electrodes, the independent factors were Semantic structure, Accentuation, and Anteriority; over the lateral electrodes, the independent factors were Semantic structure, Accentuation, Anteriority, and Hemisphere.

When the degree of freedom in the numerator was larger than one, the Greenhouse-Geisser correction was applied.

Results

Results of the cluster-based random permutation test

Within the window latency of target character, the target characters with an attention probe elicited a larger N1 than those without an attention probe (within 170–270 ms post-character onset, namely, 120–230 ms post-probe onset, $p <$ .001) (see Fig. 2).

The statistical analysis performed on the four experimental conditions (ACCwi, UNwi, ACCbe, and UNbe) of difference waveforms (with-probe_minus_without-probe) resulted in a significant interaction between Accentuation and Semantic structure (180–260 ms post-character onset, $p <$ 045). Further simple analysis demonstrated that, in the case of un-accentuation, the within-event condition elicited a larger probe-related N1 effect than the between-event condition (180–260 ms post-character onset, $p =$ .014); in contrast, in the case of accentuation, there was no significant difference between the within-event and between-event conditions (with the smallest p value being 0.500) (see Fig. 3).

The permutation test performed over the four pairs (with-prove vs. without-probe) of original ERP waveforms demonstrated that: the probe-related N1 effect started from around 140–150 ms after the post-character onset in the 'within-event and un-accented' condition, but started around 180–190 ms in the other three conditions (see Fig. 4).

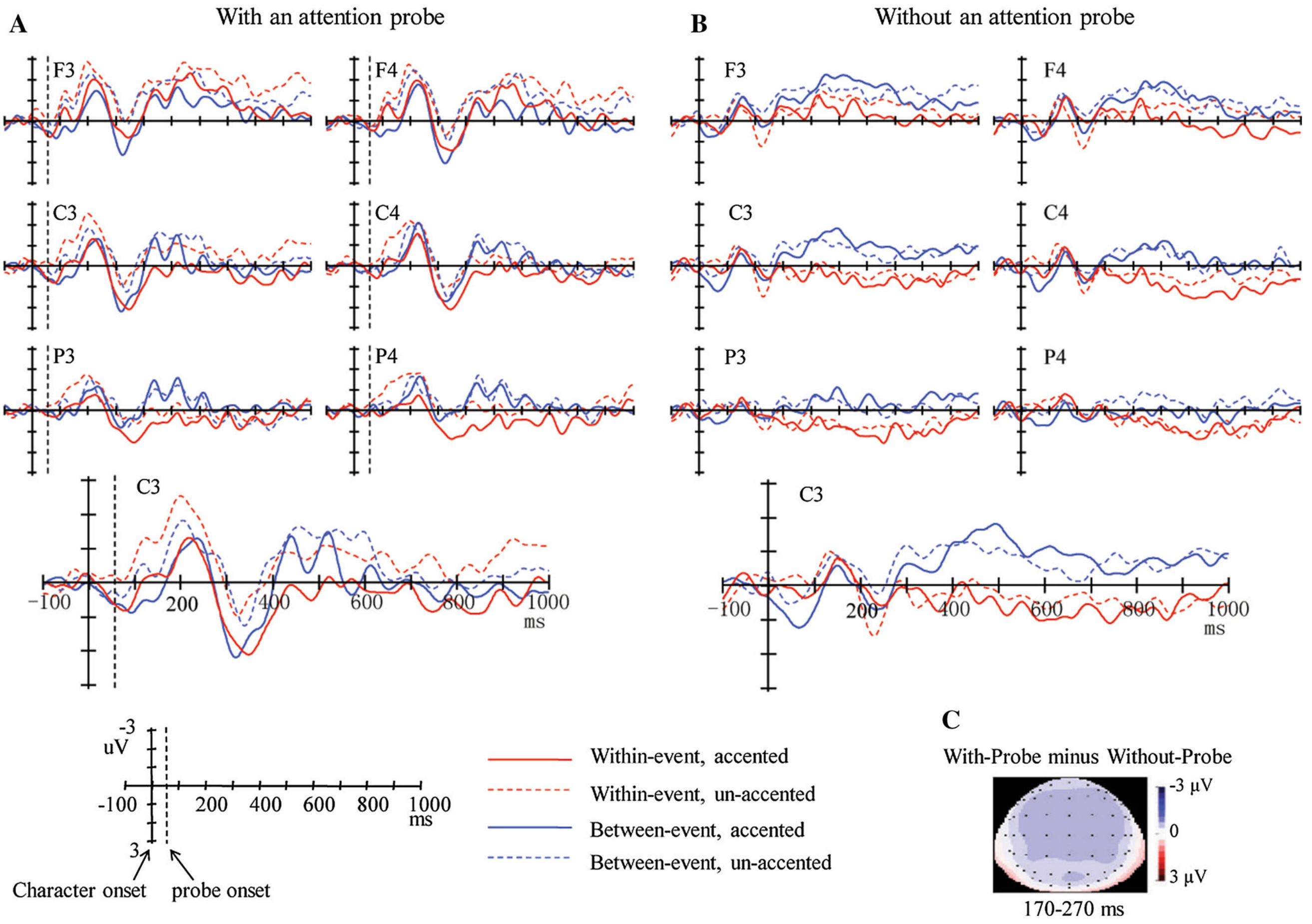

Fig. 2 Grand-average ERPs time-locked to the target character (namely, the third character). **a** Grand-average ERPs time-locked to the target characters that were added with a linguistic attention probe in the four experimental conditions; **b** Grand-average ERPs time-locked to the target characters that were not added with a linguistic attention probe in the four experimental conditions; **c** Topography of the attention-probe effect within 170–270 ms post-character onset

Results of the ANOVAs

First, the 1-ANOVAs conducted over the eight conditions of original waveforms revealed a significant main effect of Probe ($F_{midline}$(1,19) = 27.87, η2=.537, p < .001; $F_{lateral}$ (1,19) = 36.94, η2 = .660, p < .001), indicating that the target character with a probe elicited a larger N1 than that without a probe.

Then, the 2-ANOVAs performed over the four conditions of difference waveforms revealed a significant two-way Semantic structure × Anteriority interaction over the midline electrodes ($F_{midline}$(2,38) = 6.73, η2 = .262, p = .011). Further simple analysis showed that the within-event condition evoked a larger N1 than the between-event condition over the frontal electrodes ($F_{midline}$(2,38) = 8.81, η2 = .462, p = .008).

Importantly, the 2-ANOVAs demonstrated that the two-way Semantic structure × Accentuation interaction ($F_{midline}$ (1,19) = 11.30, η2 = .373, p = .003; $F_{lateral}$(1,19) = 11.75, η2 = .382, p = .003) reached significance. Further simple analysis found that, relative to the between-event condition, the N1 effect was of larger amplitude in the within-event condition when the target character followed an un-accented character ($F_{midline}$(1,19) = 13.23, η2 = .692, p = .002; $F_{lateral}$(1,19) = 8.76, η2 = .459, p = .004), but not when it followed an accented character ($F_{midline}$(1,19) = .58, η2 = .026, p = .456; $F_{lateral}$(1,19) = .62, η2 = .029, p = .339) (see Fig. 3). In addition, for the simple analysis, there was no significant interaction between semantic structure and Anteriority/Hemisphere, indicating that the enhanced N1 effect (within-event vs. between-event) observed in the case of un-accentuation should have a wide scalp distribution.

Discussion

With the help of EEG techniques and the attention probe paradigm, this experiment has investigated how the semantic level of event-structure modulates attention

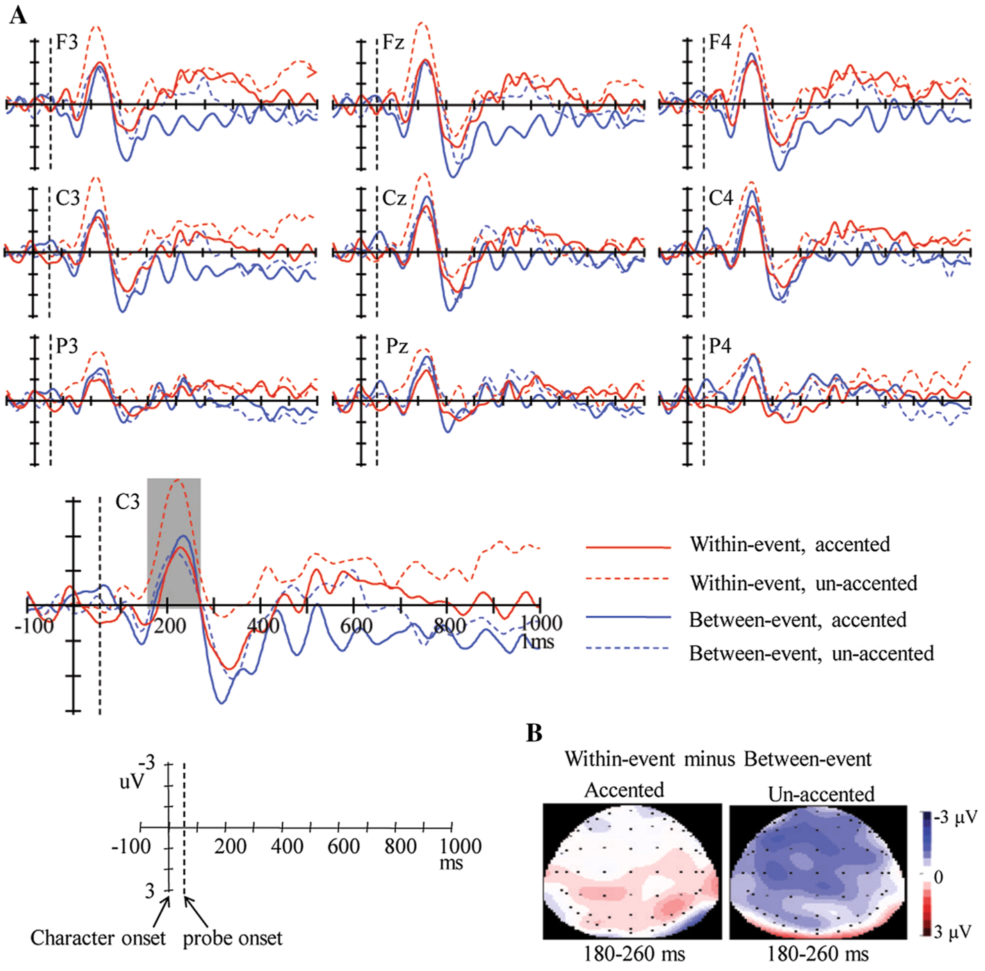

Fig. 3 Difference ERPs waveforms and topographies of the probe-related N1 effect time-locked to the target character. **a** Difference ERPs waveforms (with-probe minus without-probe) in the four experimental conditions. **b** Topographies of the Semantic structure effects at different levels of Accentuation, based on the difference ERP waveforms

allocation while speech signals unfold in time, and how semantic and acoustic structures interact with each other during temporal attention. We manipulated not only the semantic relationship between the current target character and the immediately preceding characters, but also the presence/absence of pitch accent on the preceding characters. The results revealed that the N1 effect evoked by the attention probe was enhanced when the current target character and the immediately preceding characters belonged to the same semantic event rather than when they spanned a semantic event boundary. This N1 enhancement effect occurred in the un-accented condition, but disappeared when the target character followed a pitch accent. These results are discussed in more detail below.

The most important aim of the present study is to examine whether and how the semantic level of event-structure shapes attention allocation while speech signals unfold in time. The results of the present study show that when there is no salient acoustic cue, such as the pitch accent, the probe-related N1 effect (with-probe vs. without-probe) is of larger mean amplitude (within 190–250 ms post-character onset) and starts earlier (starting around 140–150 ms post-character onset) when the current character and the immediately preceding characters belong to the same semantic-event than when they span a semantic-event boundary (starting around 180–190 ms post-character onset). This larger and faster probe-related N1 effect observed in the within-event condition indicates that listeners direct more attentional resources to the within semantic-event moment than to the between semantic-event moment while speech signals unfolded in time. Another possible explanation of the N1 enhancement effect is that it is caused by low-level acoustic differences, and consequently can't be linked to cognitive process, such as attention allocation. However, this acoustic-difference explanation of the N1 enhancement is untenable due to the following reasons. First, both the critical target character and the probe 'ba' matched on acoustic parameters (such as intensity, duration, pitch maximum, and pitch range) between the within- and between-event conditions. Second, the ANOVAs performed over the four without-probe conditions (with mean amplitude within 140-190 ms post-character onset as dependent factor) revealed neither significant main effect of Accentuation/Semantic structure (all

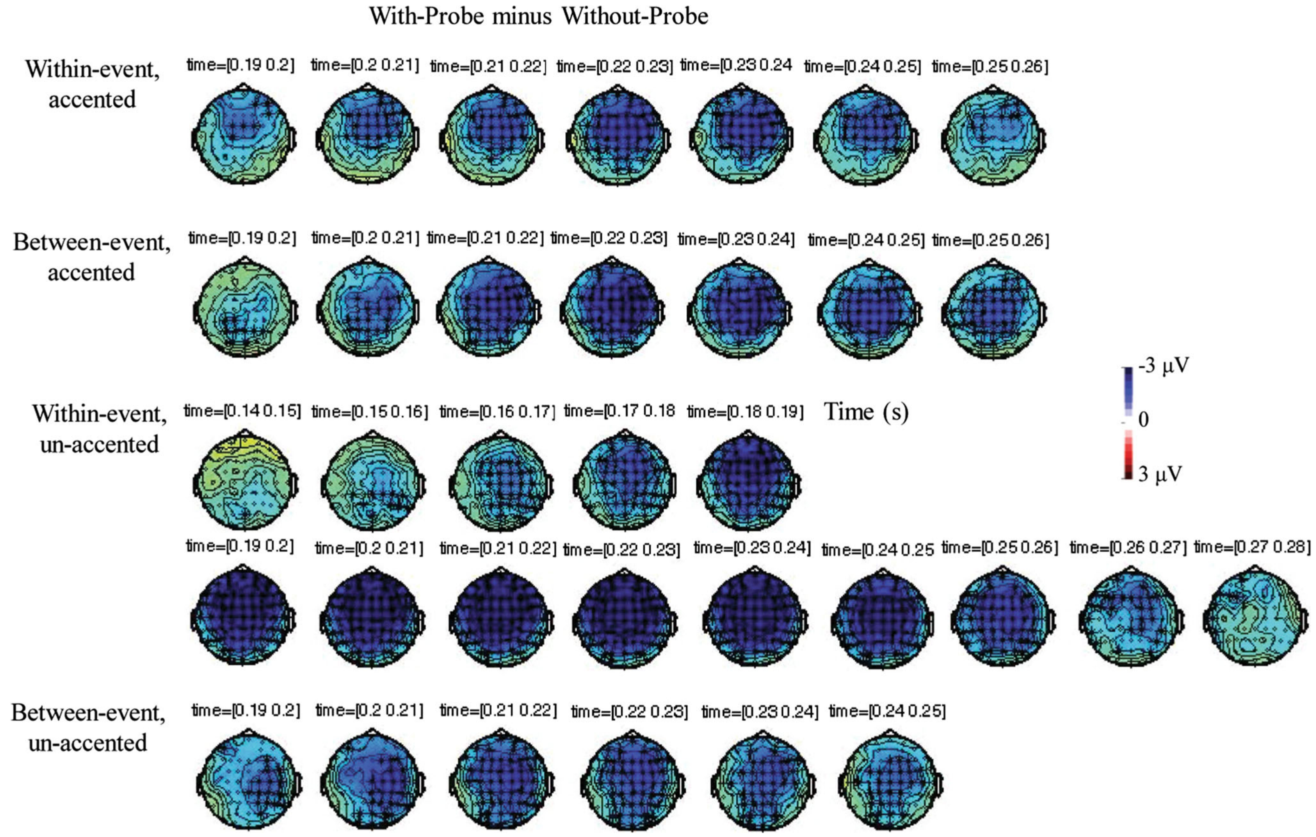

Fig. 4 The results of the cluster-based random permutation test to compare the attention-probe effect (with-probe vs. without-probe) in the four experimental conditions. The *dots* and the *asterisks* in the topography indicate the electrodes that were included in the permutation test; the *asterisks* indicate the electrodes over that the difference between the with-probe and without-probe conditions reach significance. Although the cluster-based permutation random tests were conducted in the step of 2 ms, the figures were shown in a step of 10 ms due to the limited space

ps > .880) nor significant interaction between these two factors ($p = .450$) (suggesting that the speech signal itself didn't evoke significant ERP effect within the relatively early window latency), whereas the probe-related N1 effect has already shown significant difference between the within- and between-event conditions around 140–190 ms post-character onset (as indicated by its starting latency). Third, in previous studies, the shortening of N1 latency or increases of N1 amplitude have already been correlated with allocation of more attentional resources (Astheimer and Sanders 2009, 2011; Haab et al. 2011; Hillyard et al. 1973; Hink and Hillyard 1976; Näätänen and Winkler 1999; Folyi et al. 2012; Lagemann et al. 2010; Obleser and Kotz 2011). Therefore, the extraneous acoustic-difference explanation of the N1 enhancement effect is farfetched. Furthermore, someone may also argue that the probe-related N1 enhancement reflects the degree of attention to the separated perceptual stream of 'ba', rather than to the ongoing speech itself. In the study conducted by Li and colleagues (Li et al. 2014), the focus accentuation of the critical word (bi-syllabic word) was realized by raising the pitch maximum, and consequently the accented word had a higher pitch maximum than the un-accented one. Therefore, the probe 'ba' added to the accented word was less salient than the same probe added to the unaccented word. If the N1 effect was driven by attention to the simultaneous stream of 'ba', probe added to the accented word would elicit a smaller delayed N1 (relative to that added to the un-accented word), which is inconsistent with the N1 enhancement effect observed in the accented condition of that study (Li et al. 2014). Moreover, in a recent study (Zhao and Li 2016), all of the critical words were monosyllabic word, and two kinds of critical words were used: high-tone/Falling-tone words (whose accentuation is realized by raising of the pitch maximum) and low-tone words (whose accentuation is realized by lowering of the pitch maximum). Therefore, the attention probe 'ba' added to the accented syllable (relative to that added to the un-accented syllable) was less salient for high-tone/Falling-tone words, but was more salient for low-tone words, whereas the speech stream itself was constantly more salient in the accented condition than in the un-accented condition. The ERP results showed that the probe-related-N1 effect is enhanced at the accented syllable (relative to the un-accented syllable), regardless of whether the accentuation is realized by raising or lowering of the pitch maximum (Zhao and Li 2016). This result provides further evidence for the assumption that the probe 'ba'-related N1 effect reflects the attention directed to the speech itself, rather than the attention directed to the simultaneous probe 'ba'. In sum, in the present study, a more plausible interpretation of the shortened latency and increased amplitude of the probe-related N1 effect in the within-event condition (compared with between-event condition) is that, during speech comprehension, a listener tends to direct more attention to the within semantic-event moment than to the between semantic event moment.

Will semantic and acoustic levels of event-structure interact with each other in modulating attention during speech comprehension? The present result showed that the attention enhancement effect (within semantic-event vs. between semantic-event) was observed only when there is no abrupt and larger acoustic variation within the corresponding window latency of speech signals; however, when the target character was preceded by a salient pitch accent, this attention enhancement effect disappeared. The absence of the attention enhancement effect of semantic-structure over the post-accentuation character might result from the fact that attention is reduced when the target speech moment immediately follows a pitch accent than when there is no salient pitch variation, since the probe-related N1 effect started later in the 'within-event and accentuation' condition (starting around 180–190 ms post-character onset) than in the 'within-event and un-accentuation' condition (starting around 140–150 ms post-character onset). However, in the present study, the target characters differed in pitch maximum and duration between the un-accented and accented (following a pitch accent) conditions, and consequently it is not ideal to directly compare the probe-related N1 effects between the accented and un-accented conditions. Although the post-accentuation attention reduction effect is in line with the post-accentuation acoustic compression (Wang 2012), it still needs to be examined further in future studies. Anyway, the result of the present study suggests that, during speech processing, the effect of semantic-event structure on temporal attention is modulated by the absence/presence of the pitch accent at the immediately preceding moment. That is, the acoustic and semantic levels of event structures interact with each other to guide selective attention while speech signals unfold in time.

The results of the present study provide a new outlook for our understanding of the selective attention mechanism and the predictive coding theory. This theory proposes that attention operates to optimize the precision of perceptual inference, hence prediction errors (or, equivalently, sensory data) being weighted (Rao 2005; Friston 2009; Feldman and Friston 2010, Hesselmann et al. 2009; Hohwy 2012). Considerable studies have demonstrated that prediction reduces sensory activities or attention (e.g., Alink et al. 2010; den Ouden et al. 2010; Todorovic et al. 2011; Sohoglu et al. 2012; Li et al. 2014). In this study, the target character in the within-event condition (relative to that in the between-event condition) not only is in the middle of a semantic event but also has a higher level of predictability. Our results indicate that more attentional resources are directed to the within-event condition than to the between-event condition, hence prediction enhancing rather than reducing attention. This attention enhancing effect of prediction may at first glance seem inconsistent with the early study (Li et al. 2014) and incompatible with predictive coding theories (Knill and Pouget 2004; Friston 2005); however, some recent studies have already demonstrated that prediction sometimes enhances rather than reduces neural responses to task-relevant stimuli (e.g., Chaumon et al. 2008; Doherty et al. 2005; Koch and Poggio 1999; Rauss et al. 2011). For example, two recent studies found opposite effects of predictability of a visual stimulus on neural activity in early visual areas (Doherty et al. 2005; Alink et al. 2010), with prediction enhancing visual-sensory processing in the former but reducing sensory processing in the latter study. Notably, the stimulus was related to task in the former but unrelated to task in the latter study. A subsequent study (Kok et al. 2012) further found that when the visual target stimuli are closely related to the current task (in the task-related visual hemisphere), the neural response in early visual cortex is enhanced by prediction; in contrast, when the visual target stimuli are not related to the current task (in the task-unrelated visual hemisphere), the neural response in early visual cortex is reduced by prediction. In the present study, the strongly predictable target character in the within-event condition belongs to the same basic semantic unit with the preceding characters, and consequently its processing can help to understand the not yet complete lexical meaning, which leads to heightened weighting of sensory evidence and a reversal of the sensory/attention reduction effect of prediction. However, in the early study, the strongly predictable target words and its immediately preceding words belonged to different lexical semantic units, and therefore attention reduction effect of prediction was observed (Li et al. 2014). Overall, under the circumstances of the present study, the semantic level of event-structure (namely, the semantic relationship between the preceding and the following contents) indeed plays a role in modulating attention while speech signals unfold in time, which can, even under certain circumstance, reverse the attenuation effect of predictability on attention allocation. During speech processing, both the attention reduction and attention enhancement effects of prediction might be underlined by the same mechanisms, namely attention operating to optimize the precision of perceptual processing.

Conclusions

As mentioned in the introduction section, the dynamic attending theory has already proposed that event-structure of the temporal sequences modulates attention allocation during auditory stimuli processing (Jones 1976; Jones and Boltz 1989). A recent MEG study, with spoken digits as materials, indeed found that temporally selective attention is coordinated with the long-distance acoustic fluctuations (Wöstmann et al. 2016). The present study provides new insights into the field of auditory-temporal attention by showing that the semantic relationship, namely, the semantic grouping structure, of the speech sequences is also able to modulate selective attention, which leads to heightened weighting of sensory evidence and reversal of the attention attenuation effect of prediction. Meanwhile, the acoustic level and the semantic level of event-structure interact with each other immediately to shape attention allocation while speech signals unfold in time. The present results are in line with the account that attention boosts the precision of sensory inference (Rao 2005; Friston 2009; Feldman and Friston 2010; Hesselmann et al. 2009; Hohwy 2012).

Acknowledgements This research was supported by State Key Program of the National Natural Science Foundation of China (61433018) and by Beijing Advanced Innovation Center for Imaging Technology (BAICIT-2016018).

Author contributions Xiaoqing Li made contributions to the design of the experiment, data analysis, the interpretation of data, and article writing. Lin Li joined in editing the language of this article. Yuping Zhang and Xiufang Du were responsible for data acquisition. Haiyan Zhao made to contributions to data analysis.

Compliance with ethical standards

Conflict of interest The authors have declared that no conflict of interest exists.

References

A frequency dictionary of Modern Chinese (Beijing Language and Culture University Press, Beijing, 1986) (**in Chinese**)

Alink A, Schwiedrzik CM, Kohler A, Singer W, Muckli L (2010) Stimulus predictability reduces responses in primary visual cortex. J Neurosci 30(8):2960–2966

Astheimer LB, Sanders LD (2009) Listeners modulate temporally selective attention during natural speech processing. Biol Psychol 80:23–34

Astheimer LB, Sanders LD (2011) Predictability affects early perceptual processing of word onsets in continuous speech. Neuropsychologia 49:3512–3516

Billig AJ, Davis MH, Deeks JM, Monstrey J, Carlyon RP (2013) Lexical influences on auditory streaming. Curr Biol 23:1585–1589

Chaumon M, Drouet V, Tallon-Baudry C (2008) Unconscious associative memory affects visual processing before 100 ms. J Vis. 8(3):1–10

Chen Y (2006) Durational adjustment under corrective focus in Standard Chinese. J Phon 34:176–201

Cutler A (1976) Phoneme-monitoring reaction time as a function of preceding intonation contour. Percept Psychophys 20:55–60

den Ouden HE, Daunizeau J, Roiser J, Friston KJ, Stephan KE (2010) Striatal prediction error modulates cortical coupling. J Neurosci 30(9):3210–3219

Doherty JR, Rao A, Mesulam MM, Nobre AC (2005) Synergistic effect of combined temporal and spatial expectations on visual attention. J Neurosci 25(36):8259–8266

Feldman H, Friston KJ (2010) Attention, uncertainty, and free-energy. Front Hum Neurosci 4:1–24

Folyi T, Fehér B, Horváth J (2012) Stimulus-focused attention speeds up auditory processing. Int J Psychophysiol 84(2):155–163

Friston KJ (2005) A theory of cortical responses. Trans R Soc B Biol Sci 360(1456):815–836

Friston K (2009) The free-energy principle: a rough guide to the brain? Trends Cogn Sci 13:293–301

Haab L, Trenado C, Mariam M, Strauss DJ (2011) Neurofunctional model of large-scale correlates of selective attention governed by stimulus-novelty. Cogn Neurodyn 5:103–111

Hesselmann G, Sadaghiani S, Friston KJ, Kleinschmidt A (2009) Predictive coding or evidence accumulation? False inference and neuronal fluctuations. PLoS ONE 5(3):1–5

Hillyard SA, Hink RF, Schwent VL, Picton TW (1973) Electrical signs of selective attention in the human brain. Science 182:177–180

Hink RF, Hillyard SA (1976) Auditory evoked potentials during selective listening to dichotic speech messages. Percept Psychophys 20:236–242

Hohwy J (2012) Attention and conscious perception in the hypothesis testing brain. Front Psychol 3:1–14

Hruska C, Steinhauer K, Alter K, Steube A (2000) ERP effects of sentence accents and violations of the information structure. In: Poster presented at the 13th annual CUNY conference on human sentence processing, San Diego

Jia Y, Xiong Z, Li A (2006) Phonetic and phonological analysis of focal accents of disyllabic words in Standard Chinese. In Chinese Spoken Language Processing, Springer, Berlin Heidelberg, pp 55–66

Jones MR (1976) Time, our lost dimension: toward a new theory of perception, attention, and memory. Psychol Rev 83(5):323

Jones MR, Boltz M (1989) Dynamic attending and responses to time. Psychol Rev 96(3):459

Knill DC, Pouget A (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neurosci 27(12):712–719

Koch C, Poggio T (1999) Predicting the visual world: silence is golden. Nat Neurosci 2(1):9–10

Kok P, Rahnev D, Jehee JFM, Lau HC, de Lange FP (2012) Attention reverses the effect of prediction in silencing sensory signals. Cereb Cortex 22:2197–2206

Lagemann L, Okamoto H, Teismann H, Pantev C (2010) Bottom-up driven involuntary attention modulates auditory signal in noise processing. BMC Neurosci 11:156

Li X, Ren G (2012) How and when accentuation influences temporally selective attention and subsequent semantic processing during on-line spoken language comprehension: an ERP study. Neuropsychologia 50:1882–1894

Li X, Lu Y, Zhao H (2014) How and when predictability interacts with accentuation in temporally selective attention during speech comprehension. Neuropsychologia 64:71–84

Magne C, Astésano C, Lacheret-Dujour A, Morel M, Alter K, Besson M (2005) On-line processing of "pop-out" words in spoken French dialogues. J Cognit Neurosci 17(5):740–756

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG-and MEG-data. J Neurosci Methods 164:177–190

Näätänen R, Winkler I (1999) The concept of auditory stimulus representation in cognitive neuroscience. Psychol Bull 125:826

Obleser J, Kotz SA (2011) Multiple brain signatures of integration in the comprehension of degraded speech. Neuroimage 55:713–723

Posner MI (1980) Orienting of attention. Q J Exp Psychol 32:3–25

Rao RP (2005) Bayesian inference and attentional modulation in the visual cortex. NeuroReport 16:1843–1848

Rauss K, Schwartz S, Pourtois G (2011) Top-down effects on early visual processing in humans: a predictive coding framework. Neurosci Biobehav Rev 35:1237–1253

Saada M, Meng Q, Huang T (2014) A novel approach for pilot error detection using dynamic Bayesian networks. Cogn Neurodyn 8:227–238

Sanford AJ, Sanford AJ, Molle J, Emmott C (2006) Shallow processing and attention capture in written and spoken discourse. Discl Process 42:109–130

Shattuck-Hufnagel S, Turk AE (1996) A prosody tutorial for investigators of auditory sentence processing. J Psycholinguist Res 25:193–247

Sohoglu E, Peelle JE, Carlyon RP, Davis MH (2012) Predictive top-down integration of prior knowledge during speech perception. J Neurosci 32(25):8443–8453

Sun L (1990) What is the constituent part of idiom? J Liupanshui Norm Univ 1:24–26

Todorovic A, van Ede F, Maris E, de Lange FP (2011) Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. J Neurosci 31(25):9118–9123

Wang B (2012) Prosodic focus with and without post-focus compression: a typological divide within the same language family? Linguist Rev 29(1):131–147

Wang B, Lü S, Yang Y (2002) The pitch movement of stressed syllable in Chinese sentences. Acta Acust 27:234–240

Wöstmann M, Herrmann B, Maess B, Obleser J (2016) Lateralized alpha oscillations reflect attentional selection of speech in noise. PNAS 113(14):3873–3878

Zhao H, Li X (2016) How Accentuation influences the allocation of temporally selective attention during online spoken language comprehension. J Psychol Sci 39:13–21