



The cognitive and neural oscillatory mechanisms underlying the facilitating effect of rhythm regularity on speech comprehension



Xiaoqing Li^{a,c,*}, Ximing Shao^{a,b}, Jinyan Xia^{a,b}, Xiaoying Xu^d

^a CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China

^b University of Chinese Academy of Sciences, Beijing, China

^c Jiangsu Collaborative Innovation Center for Language Ability, Jiangsu Normal University, Xuzhou, China

^d School of Chinese Language and Literature, Beijing Normal University, Beijing, China

ARTICLE INFO

Keywords:

Rhythm regularity
Speech comprehension
Brain oscillations
Dynamic attending theory

ABSTRACT

Although a larger number of studies with Indo-European languages as materials have demonstrated the facilitating effect of rhythm regularity on speech comprehension, the internal mechanisms underlying this facilitating effect is still not completely clear. The present electroencephalograph study examined whether and how a rhythmical sentence context affects Mandarin Chinese speech comprehension, and the cognitive and neural oscillatory mechanisms underlying this rhythm regularity effect. Participants listened to Mandarin Chinese sentences that had a regular or irregular rhythm context and that contained critical nouns that were semantically congruent or incongruent given the preceding sentence contexts. The results revealed that, relative to congruent nouns, incongruent nouns elicited a larger N1 and a larger N400 when they were embedded in a regular-rhythm context, but elicited only a larger P600 when they were embedded in an irregular-rhythm context. This difference in the timing of neural responses to incongruent nouns suggests that a temporal rhythmical context speeds up speech comprehension. Moreover, as compared to the irregular-rhythm contexts, regular-rhythm contexts induced neural oscillatory power increases in the beta band immediately preceding the critical nouns and power increases in the alpha band immediately following these nouns. The beta-power enhancement effect induced by the preceding rhythmic context was negatively correlated with the N400/P600 enhancement effect evoked by the subsequent incongruent nouns. The above results suggest that the facilitating effect of rhythm regularity on speech comprehension relies at least in part on alpha-beta band neural oscillations, especially on beta activity; a temporal rhythmical context facilitates speech comprehension both by enhancing neural excitability associated with early acoustic-phonological processing and by reducing cognitive costs associated with later semantic resolution. The present results were also discussed with regard to the dynamic attending theory.

1. Introduction

One of the most fascinating abilities of the human brain is that it can recognize the temporally regular patterns in the dynamically changing auditory or visual stimuli. For example, the sounds/constituents of speech signals are organized in a regular or semi-regular pattern in time, such as the regular alternation of strong or weak syllables. This rhythm regularity, namely the regularity of the temporal organization of stimuli, leads the human brain to develop temporal predictions about when the next event will occur in the

* Corresponding author. CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, No. 16 Lincui Road, Chaoyang District, Beijing 100101, China.

E-mail address: lixq@psych.ac.cn (X. Li).

<https://doi.org/10.1016/j.jneuroling.2018.05.004>

Received 31 October 2017; Received in revised form 15 May 2018; Accepted 16 May 2018
Available online 23 May 2018

0911-6044/ © 2018 Elsevier Ltd. All rights reserved.

forthcoming signals (Dalla Bella, Białuńska, & Sowiński, 2013; Large & Jones, 1999). A larger number of studies have demonstrated that temporal predictions derived from rhythm regularity play an important role in auditory/visual perception and language comprehension as well (e.g., Rothermich, Schmidt-Kassow, & Kotz, 2012; Rothermich & Kotz, 2013). Neural oscillations, which reflect fluctuations of neuronal excitability cycles, have recently been linked to at least part of the neural mechanisms underlying auditory/visual processing (e.g., Fujioka, Trainor, Large, & Ross, 2012; Luo & Poeppel, 2007). However, how the human brain relies on the temporal rhythmical structure and the underlying neural oscillatory activities to comprehend speech remains to be explored. The present study aims to further probe into the effect of rhythm regularity on speech comprehension, and especially the neural oscillatory mechanisms underlying this rhythm effect.

In the field of psycholinguistics, some studies have examined the effect of rhythm regularity on speech processing. Although the syllables of natural speech are not produced at an objectively constant rate, the semi-regular rhythm inherent in speech can allow a listener to form perceptual level of temporal regularity (e.g., Fowler, 1979; Hoequist, 1983; Morton, Marcus, & Frankish, 1976). Studies with measurable behavioral responses have already demonstrated the beneficial effects of temporal regularity on language acquisition (Jusczyk, 1999; Nazzi & Ramus, 2003; Thiessen & Saffran, 2007) and word segmentation in continuous speech (Endress & Hauser, 2010; Lee & Todd, 2004; Mattys & Samuel, 1997). Recent event-related potential (ERP) studies further found that the processing of syntactic incongruities/ambiguities (Cason & Schön, 2012; Roncaglia-Denissen, Schmidt-Kassow, & Kotz, 2013; Schmidt-Kassow & Kotz, 2009a, 2009b), lexical semantics (Magne et al., 2007), and semantic incongruities (Rothermich et al., 2012) is facilitated when the sentence context carries a regular rhythm structure. For example, in languages such as German or English, rhythm regularity is achieved through the regular alternation of strong and weak syllables (i.e. the metric foot) in speech. An ERP study with German sentences as material directly compared the effect of rhythmically regular (namely, regular alternation of stressed and unstressed syllables) and irregular sentence contexts on lexical-semantic integration. The results showed that the N400 enhancement (semantically incongruent words vs. congruent words) effect was reduced in the regular-rhythm condition (compared to the irregular-rhythm condition) (Rothermich et al., 2012). The N400 component, which has a centro-parietal distribution and reaches maximum around 400 ms after the onset of the stimuli, has been considered to reflect the difficulty of semantic processing (Hagoort & Brown, 2000; Kutas & Van Petten, 1994). Therefore, the reduced N400 effect found by Rothermich and colleagues (Rothermich et al., 2012) indicates that a regular rhythm can reduce the cognitive costs caused by integrating semantically incongruent words (Rothermich et al., 2012). Different from German and English, the rhythm organization of French in spoken sentences is realized by regularly lengthening the last syllable of word or group of words (Bailly, 1989; Wenk & Wioland, 1982). Magne et al. (2007) examined the effect of rhythm regularity on lexical-semantic processing through manipulation of syllabic duration of French sentences. They found that rhythmically incongruous sentence-final words evoked an enhanced biphasic N400-P600; meanwhile, behavioral data showed that participants made more errors for semantically congruous but rhythmically incongruous words when they were attending to the semantic aspects of the sentence. The N400 effect, combined with the behavioral results, suggests that the rhythm structure of a sentence interferes with lexical access to word meaning (Magne et al., 2007). All of the above studies provide experimental evidences for the fact that the temporal rhythm structure carried by a sentence context plays an important role in language comprehension, with regular rhythm facilitating semantic processing.

The existing studies that examined the effect of rhythm regularity on sentence processing are mainly conducted with Indo-European languages as materials. Languages can be divided into stress-timed (i.e., stressed syllables are isochronous) and syllable-timed languages (i.e., all syllables are distributed isochronously), irrespective of physical or phonetical isochrony. As mentioned earlier, English and German are typical stress-timed languages, in which stressed and unstressed syllables are regularly distributed in sentences; French is commonly classified as an example of syllable-timed languages, in which every syllable is thought to take up roughly the same amount of time (e.g., Roach, 1982). Compared to the Indo-European languages, Mandarin Chinese has its own rhythm patterns. Mandarin Chinese is a type of tone language, and it belongs neither strictly to stress-timed nor strictly to syllable-timed languages. In Mandarin Chinese, the rhythm regularity of speech is characterized by regular distribution of phonological-foot (namely, the basic prosodic unit) types. Specifically, in natural spoken sentences or discourses, bisyllabic phonological foot occupies around 74 percent, with the remaining being mono- or trisyllabic (Feng, 1998; Wu, 2005). Regarding Mandarin Chinese speech processing, although some studies have found that a violation of phrase-level rhythm (e.g., 2syllables + 1syllable vs. 2syllables + 2syllables for a verb + noun phrase) impeded sentence comprehension (Luo & Zhou, 2010; Luo, Duan, & Zhou, 2015), we still do not know how the relatively longer-scale rhythm regularity (e.g., the regular distribution of phonological-foot types in a longer-scale sentence context) modulates sentence comprehension. Further studies are needed to explore whether and how the facilitating effect of longer-scale rhythm regularity on speech processing extends to languages such as Mandarin Chinese.

More importantly, although the existing studies have already observed the facilitating effect of rhythm regularity on language processing, the mechanism underlying this facilitating effect is only beginning to be understood. Dynamic attending theory provides a framework to explain how temporal rhythm structure in speech is used to modulate selective attention and speech processing (Jones & Boltz, 1989; Large & Jones, 1999). According to this theory, the rhythmic structure of speech can coordinate our internal neural oscillations with periodicities at each hierarchical level of an external temporal structure; consequently, it guides more attention directed to temporally predicted moments of subsequent rhythmic events, and even can directly modulate perceptual excitability (Jones & Boltz, 1989; Large & Jones, 1999). Some neuroimaging studies have provided support for dynamic attending theory by showing that the oscillatory activity in primary perceptual areas (e.g., V1 and A1) is entrained to the external rhythmic stimuli, and consequently cortical excitability is maximized at predicted points in time (e.g., Besle et al., 2011; Lakatos et al., 2009, 2008; Large & Jones, 1999; Wöstmann, Herrmann, Maess, & Obleser, 2016). Behavioral studies, with the help of the diffusion model, also demonstrated that temporal expectation generated from the rhythmic structure of visual stimuli can boost the perception of relevant events by directly modulating signal contrast and early signal encoding (Jepma, Wagenmakers, & Nieuwenhuis, 2012; Rohenkohl,

Cravo, Wyart, & Nobre, 2012). In sum, both the existing experimental evidences and dynamic attending theory suggest that temporal rhythm regularity can modulate neural excitability of perceptual brain areas and facilitate early perceptual processing of simple auditory/visual stimuli. As to the relatively complex cognitive processes, such as speech comprehension, despite the experimental evidences for the facilitating effect of rhythm regularity on later semantic/syntactic processing, there is still not a clear picture of whether and how early sensory-perceptual processing of incoming speech input is facilitated by rhythm regularity.

Dynamic attending theory proposes that the temporal rhythm regularity of speech modulates neural excitability and information processing by controlling internal neural oscillations in time. Then, a further interesting question arising here is which frequency bands of neural oscillatory activities underline the facilitating effect of rhythm regularity on speech processing. Beta oscillations arguably play a complementary function during top-down control processes, such as temporal expectations (e.g., Arnal, Doelling, & Poeppel, 2015; Donner, Siegel, Fries, & Engel, 2009; Fujioka et al., 2012; de Lange, Rahnev, Donner, & Lau, 2013; Lewis, Wang, & Bastiaansen, 2015); an increase in beta activity signals that the processing system is actively maintaining the current cognitive/neural set or preparing for expected sensory input and a decrease in beta activity signals that the current mode of processing is expected to change (for review see Lewis et al., 2015; Lewis, Schoffelen, Schriefers, & Bastiaansen, 2016). Ongoing neuronal oscillations in delta-theta band are considered to provide some of the relevant mechanisms to parse continuous speech into the necessary chunks (e.g., syllables) for processing (Ding, Melloni, Zhang, Tian, & Poeppel, 2016; Luo & Poeppel, 2007; Peelle, Gross, & Davis, 2013). Alpha activity (8–12 Hz) has been linked to the inhibiting and enhancing functions of attention (e.g., for reviews see Klimesch, 2012 and Jensen & Mazaheri, 2010) and to the general attention alertness state (e.g., Sadaghiani et al., 2010). Temporal expectation or segmentation (and consequently selective attention) might also be affected by the temporal regularity of a speech context through modulation of internal neural oscillations. However, as for speech comprehension, we still do not know which frequency bands of neural oscillations are relied on by the human brain to use the rhythmic context to facilitate speech processing (e.g., detection or resolution of incoming semantically incongruent words in continuous speech signal).

Therefore, the aim of the present study was to use the EEG (electroencephalograph) technique to examine the effect of rhythm regularity on Mandarin Chinese speech comprehension and the internal mechanisms underlying this effect. Specifically, this study aimed to further explore, during speech comprehension, whether and how the early-sensory and later-semantic stages of information processing are affected by the rhythm of spoken sentences, and what is the neural oscillatory mechanism underlying this rhythm regularity effect on speech comprehension.

Mandarin Chinese spoken sentences were used as experimental stimuli, with each sentence including a critical noun. The critical noun is either semantically congruent or incongruent given the semantic constraints of the preceding sentence context, and meanwhile the rhythm pattern of this context is either regular (100% bisyllabic phonological foot) or irregular (random distribution of mono-, bi, or trisyllabic phonological foot). To examine the early sensory-perceptual and late semantic stages of processing of incoming critical nouns, we focused on two types of ERP components: N1 and N400/P600. The N1 component, which is typically elicited around 100 ms after stimulus onset, is sensitive both to acoustic aspects of individual stimuli (Näätänen & Picton, 1987) and to phonological features (such as vowels, syllables, and word onsets) present in speech signal (Obleser, Lahiri, & Eulitz, 2004; Obleser, Elbert, Lahiri, & Eulitz, 2003; Parviainen, Helenius, & Salmelin, 2004; Sanders & Neville, 2003), and can therefore be taken as an index of early low-level acoustic-perceptual processing of language input. The N400 component, as mentioned earlier, reflects the difficulty of semantic processing (Hagoort & Brown, 2000; Kutas & Van Petten, 1994); the P600 component has a broad or central-parietal distribution and occurs within a time range from approximately 500 to 1000 ms post-critical stimulus onset, which is considered to reflect the cost of reanalysis process when required (Bornkessel, McElree, Schlesewsky, & Friederici, 2004; Schlesewsky & Bornkessel, 2006; Wang, Schlesewsky, Bickel, & Bornkessel-Schlesewsky, 2009; Wolff, Schlesewsky, Hirotani, & Bornkessel-Schlesewsky, 2008). We predicted that if the rhythm regularity modulates not only the later-semantic but also the early-perceptual processing of incoming speech input, incongruent nouns (relative to congruent nouns) will elicit an enhanced N400/P600 and an enhanced N1 as well, and these N400/P600 and N1 effects will change dynamically as a function of the rhythm regularity of a sentence context.

As to how the rhythmic structure of speech modulates the early-perceptual and later-semantic processing of upcoming words, there might be two possibilities that are not mutually exclusive. One possibility is that a regular rhythm context might lessen the cognitive costs associated with later semantic integration or resolution of incongruent words, as it is argued to be able to coordinate internal neural oscillations with external speech signal and consequently guide more attention to the temporally predicted moments of speech. If this hypothesis is true, the N400/P600 effect elicited by incongruent nouns will be reduced or/and occur earlier in the regular-rhythm context (compared to the irregular-rhythm context); moreover, the more likely a listener perceives the temporal regularity carried by the preceding context and uses this information to process speech, the smaller the N400/P600 effect elicited by incongruent nouns will be. The second possibility is that a regular rhythm context might facilitate the early sensory-perceptual encoding of incoming speech input and consequently boost early detection of incongruent words, since the temporally predicted stimuli has been found to tend to be aligned to an optimal state of sensory neural excitability. If this is true, incongruent nouns will evoke a larger N1 than congruent nouns, and this N1 effect will be enhanced, rather than reduced, in the regular-rhythm context (compared to the irregular-rhythm context), or this early N1 effect will occur only in the regular-rhythm context. These two possibilities are not mutually exclusive, as it might be that the more likely the listeners detect the incongruent speech input at the early processing stage, the better they can resolve this incongruence later.

In addition, by examining the neural oscillatory activities associated with rhythm regularity processing and their relationship with the early-perceptual and later-semantic processing of upcoming words, we will get some understanding of the oscillatory mechanisms by which rhythm regularity affect speech comprehension.

2. Method

2.1. Ethics statement

All participants provided written informed consent in accordance with the Declaration of Helsinki. The ethics committee of the Institute of Psychology at the Chinese Academy of Sciences approved this study, its participant-recruitment procedure, and its methodology.

2.2. Participants

Twenty right-handed university students (21–26 years old; 10 males) participated in this experiment. All of them were native speakers of Mandarin Chinese. None reported any medical, neurological, or psychiatric illness, and all gave informed consent. The data of 2 participants (1 male) were removed from analysis because of excessive artifacts.

2.3. Stimuli

Mandarin Chinese sentences were used as stimuli, which were produced by a female speaker and recorded at a sampling rate of 22050 Hz. The prosodic rhythm pattern of the sentences was manipulated. Mandarin Chinese naturally contains highly regular prosodic rhythms: spoken Chinese sentence/discourse is composed of around 74% of bisyllabic prosodic units that is called phonological foot (the remaining phonological foot being mono- or trisyllabic) (Feng, 1998). In the present study, 160 pairs of sentence frames were constructed, with each pair including one sentence frame having a regular prosodic rhythm structure (regular-rhythm) and another having an irregular prosodic rhythm structure (irregular-rhythm). Specifically, in the regular-rhythm condition, all of the words are bisyllabic words, and correspondingly the sentence frame is composed of 100% of bisyllabic phonological foot (e.g., *Da pian/yin xing/shu ye/bian cheng/huang se/shuo ming ...*) (*Most of the ginkgo leaves turning yellow indicates that ...*); in the irregular-rhythm condition, the mono-, bi-, and trisyllabic words are quasi-randomly distributed, and consequently the proportion of bisyllabic phonological foot in the sentence is dramatically lower than the natural rate 74% (e.g., *Da pian/yin xing shu/bian/jin huang se/shuo ming ...*) (*Most of the ginkgo leaves turning yellow indicates that ...*) (see Table 1 for example). Importantly, in each pair of sentence frames, the regular- and irregular-rhythm conditions have the same syntactic structure, the same lexical meaning, and the same sentence-level meaning; the last word (which is bisyllabic verb) of the sentence frame is also exactly the same between the regular- and irregular-rhythm conditions. In addition, to successfully manipulate the regular/irregular rhythm, the number of characters preceding the critical nouns is sometimes (in around half of the experimental materials) different between the regular- and irregular-

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

Conditions	Example sentences
Group A:	
ReCon	小红 来到 蔬菜 大棚 看见 黄瓜 又细 又长。 Xiaohong/arrive at/vegetable/greenhouse/find/ <u>cucumbers</u> /thin/long Xiaohong arrived at the vegetable greenhouse and found that the <u>cucumbers</u> were thin and long.
ReIncon	小红 来到 蔬菜 大棚 看见 鲨鱼 凶猛 无比。 Xiaohong/arrive at/vegetable/greenhouse/find/ <u>sharks</u> /ferocious/extremely Xiaohong arrived at the vegetable greenhouse and found that the <u>sharks</u> were extremely ferocious.
IrreCon	王小红 到 蔬菜棚 看见 黄瓜 又细 又长。 Wang Xiaohong/arrive at/vegetable greenhouse/find/ <u>cucumbers</u> /thing/long Wang Xiaohong arrived at the vegetable greenhouse and found that the <u>cucumbers</u> were thing and long.
IrreIncon	王小红 到 蔬菜棚 看见 鲨鱼 凶猛 无比。 Wang Xiaohong/arrive at/vegetable greenhouse/find/ <u>sharks</u> /ferocious/extremely Wang Xiaohong arrived at the vegetable greenhouse and found that the <u>sharks</u> were extremely ferocious.
Group B:	
ReCon	小红 潜入 海洋 深处 看见 鲨鱼 凶猛 无比。 Xiaohong/dive into/sea/deep/find/ <u>sharks</u> /ferocious/extremely Xiaohong dived into the deep see and found that the <u>sharks</u> were extremely ferocious.
ReIncon	小红 潜入 海洋 深处 看见 黄瓜 又细 又长。 Xiaohong/dive into/sea/deep/find/ <u>cucumbers</u> /thing/long Xiaohong dived into the deep see and found that the <u>cucumbers</u> were thing and long.
IrreCon	王小红 到 深海处 看见 鲨鱼 凶猛 无比。 Wang Xiaohong/dive into/deep see/find/ <u>sharks</u> /ferocious/extremely Wang Xiaohong dived into the deep see and found that the <u>sharks</u> were extremely ferocious.
IrreIncon	王小红 到 深海处 看见 黄瓜 又细 又长。 Wang Xiaohong/dive into/deep see/find/ <u>cucumbers</u> /thing/long Wang Xiaohong dived into the deep see and found that the <u>cucumbers</u> were thing and long.

Note. The underlined words are the critical nouns; the semantically congruent nouns in group A are the semantically incongruent nouns in group B, and vice versa. ReCon indicates ‘regular rhythm & semantic congruent’; ReIncon indicates ‘regular rhythm & semantic incongruent’; IrreCon indicates ‘irregular rhythm & semantic congruent’; IrreIncon indicates ‘irregular rhythm & semantic incongruent’.

rhythm conditions, with the former having one more or less character than the later (e.g., for the example sentences in Table 1, the sentence frame preceding the critical noun has 10 characters in the regular-rhythm condition, but 9 characters in the irregular-rhythm condition); although we do not know whether the small difference in the number of characters would be a confounding factor or not, to exclude the potential influences of the sentence-frame length, the number of characters was counterbalanced between the regular- and irregular-rhythm across all 160 sets of materials.

The sentence frame is immediately followed by a bisyllabic noun, which is the critical noun and not in the sentence-final position. We manipulated the semantic congruency of the critical nouns, with them being semantically congruent (e.g., *Most of the ginkgo leaves turning yellow indicates that the AUTUMN is coming*) or incongruent (e.g., *Most of the ginkgo leaves turning yellow indicates that the SPRING is coming*) given the preceding sentence context. To avoid the difference between the congruent and incongruent nouns becoming a confounding factor, the 160 sets of sentences are divided into two groups (group A with 80 sets; group B with 80 sets); the congruent nouns in group A are the incongruent nouns in group B, and vice versa.

Taken together, this resulted in a full factorial design with all combinations of the factors Rhythm (regular vs. irregular) and Semantic congruency (congruent vs. incongruent): ReCon, ‘regular rhythm & semantic congruent’; ReIncon, ‘regular rhythm & semantic incongruent’; IrreCon, ‘irregular rhythm & semantic congruent’; IrreIncon, ‘irregular rhythm & semantic incongruent’ (see Table 1 for example sentences in the four experimental conditions).

To validate the semantic congruency of the critical nouns, 24 participants, who did not participate in the EEG experiment, were instructed to rate the degree of semantic congruency of the critical noun in each written sentence context on a 7-point scale (from –3 to 3). The larger the score was, the more congruent the nouns were. ANOVAs (analyses of variance) were performed on the rating scores with Rhythm (regular vs. irregular) and Semantic congruency (congruent vs. incongruent) as independent factors. The results revealed a significant main effect of Semantic congruency ($F_{(1, 159)} = 4054.9, p < .0001$), indicating that the rating scores in the incongruent condition were significantly lower ($-1.88, SD = 0.70$) than those in the congruent condition ($2.14, SD = 0.68$) (see Table 2 for semantic-congruency values of the four conditions). This main effect was not qualified by a two-way Rhythm \times Semantic congruency interaction ($F_{(1, 159)} = 2.14, p = .145$). These results suggested that the manipulation of semantic congruency was successful.

We also controlled the potential confounding factors coming from the acoustic parameters of the critical nouns in the four conditions. During sentence recording, the four conditions of sentences from the same set (e.g., the four sentences in Group A or Group B in Table 1) were recorded together, then another set of sentences was recorded; meanwhile, the recording order of the four conditions was also counterbalanced across the 160 sets of materials. Then, ANOVAs were performed on duration, intensity, and mean-pitch values of the critical nouns, respectively, with Rhythm (regular vs. irregular) and Semantic congruency (congruent vs. incongruent) as independent factors. The results revealed neither a significant main effect of Rhythm ($F_{(1, 159)} = 3.60, p = .062; F_{(1, 159)} = 2.015, p = .156; F_{(1, 159)} = 0.001, p = .972$ for duration, intensity, and mean pitch respectively) nor a significant main effect of Semantic congruency ($F_{(1, 159)} = 0.263, p = .615; F_{(1, 159)} = 0.030, p = .861; F_{(1, 159)} = 0.008, p = .930$ for duration, intensity, and mean pitch respectively); the two-way Rhythm \times Semantic congruency interaction did not reach significance either ($F_{(1, 159)} = 0.077, p = .781; F_{(1, 159)} = 1.173, p = .282; F_{(1, 159)} = 0.222, p = .638$ for duration, intensity, and mean pitch respectively) (see Table 2). Overall, these results showed that the acoustic parameters of the critical nouns are not confounding factors.

Totally, there are 160 sets of experimental sentences, with each set including 4 versions. The experimental materials were grouped into 4 lists of sentences according to the Latin square procedure based on the four experimental conditions. In each list, there were an equal number of sentences (40 sentences) for each of the four experimental conditions and additional 90 filler sentences (totally 250 sentences); moreover, the same critical noun was only presented once in each list. To prevent the subjects from figuring out the experimental aims, the rhythm regularity of the 90 filler sentences lies somewhat in between the regular-rhythm and irregular-rhythm experimental conditions, with half of them including a semantically incongruent noun. Subjects were divided into 4 groups, with each group listening to only one list of materials.

2.4. Procedure

After the electrodes were positioned, subjects were asked to listen to each sentence for comprehension. Meanwhile, their EEG

Table 2
Acoustic parameters and semantic congruency of the critical nouns in the four experimental conditions.

Parameters	Mean (Standard Deviation)			
	ReCon	ReIncon	IrreCon	IrreIncon
Duration (ms)	546.76 (67.6)	549.72 (65.8)	543.54 (67.0)	545.33 (62.5)
Intensity (dB)	58.22 (2.5)	58.08 (2.4)	57.90 (2.6)	58.00 (2.5)
Pitch (Hz)	237.89 (31.2)	238.16 (32.4)	238.46 (31.1)	237.68 (30.5)
SemCon	2.14 (0.71)	-1.81 (0.68)	2.14 (0.68)	-1.95 (0.67)

Note. SemCon indicates the semantic-congruency values of the critical nouns. ReCon indicates ‘regular rhythm & semantic congruent’; ReIncon indicates ‘regular rhythm & semantic incongruent’; IrreCon indicates ‘irregular rhythm & semantic congruent’; IrreIncon indicates ‘irregular rhythm & semantic incongruent’.

signals were recorded. At the end of each of one third of the 250 sentences (namely, 84 sentences), 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 the target sentence. To inform subjects of when to fixate and sit still for EEG recording, an asterisk was displayed from 500 ms before 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 10 min each, separated by brief resting periods.

2.5. 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 and horizontal electro-oculogram was monitored. All electrode impedance levels (EEG and EOG) were kept below 5 k Ω .

2.6. ERP analysis

We focused on the ERPs and brain oscillatory activities time-locked to the acoustic onset of the critical nouns. For ERP analysis, 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) and filtered with a band-pass filter 0.1–40 Hz. Subsequently, the filtered data were divided into epochs ranging from 100 ms before the onset of the critical words to 1000 ms after the onset of the critical words. A time window of 100 ms preceding the onset of the critical words 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 $\pm 75 \mu\text{V}$, followed by a manual check). Trials containing the artifacts mentioned above were rejected (15% 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.

For statistical analysis, analyses of variance were conducted on a selection of midline electrodes (FZ, CZ, and PZ) and lateral electrodes (F5/F3/FC5/FC3; F4/F6/FC4/FC6; C5/C3/CP5/CP3; C4/C6/CP4/CP6; P5/P3/PO5/PO3; P4/P6/PO4/PO6) respectively. The dependent factor was the mean amplitude in the 110–170 ms (for N1 effect), 350–550 ms (for N400 effect), and 700–1000 ms (for later negativity effect or P600 effect) after the acoustic onset of the critical nouns. Over the midline electrodes, the independent factors were Rhythm (regular vs. irregular) and Semantic congruency (congruent vs. incongruent), and Anteriority (anterior: FZ, central: CZ, and posterior: PZ); for lateral electrodes, the mean amplitude values were entered into statistical analysis with Hemisphere (left vs. right) as an additional factor and lateral electrodes nested under Hemisphere. When the degree of freedom in the numerator was larger than one, the Greenhouse-Geisser correction was applied.

2.7. EEG time-frequency analysis

Event-related spectral perturbation (ERSP) was applied to characterize the oscillatory brain activities. ERSP represented the time-frequency representations (TFRs) averaged across single trials. We are mainly interested in the neural oscillation difference between the regular- and irregular-rhythm conditions before and around onset of the critical nouns. Therefore, the ReCon and ReIncon conditions were collapsed as Regular condition, and IrreCon and IrreIncon conditions were collapsed as Irregular condition. 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), filtered with a band-pass filter 1–80 Hz, and divided into epochs ranging from 2400 ms before the onset of the critical nouns to 700 ms after the onset of the critical nouns. Then, we deleted the residual ocular artifacts and muscle artifacts using the ‘reject data using ICA’ algorithm implemented in the statcond function of EEGLAB (EEGLAB 10.2.5.5b, <http://www.scn.ucsd.edu/eeglab>) toolbox. Subsequently, ERSP were calculated using EEGLAB running under Matlab 7.5 (MathWorks, Natick, MA, USA). The low and high frequency ranges were computed separately by using Morlet wavelet decomposition (Goupillaud, Grossmann, & Morlet, 1984) within the 3100-ms epoch in a step of around 50 ms. In the low frequency range, we used a series of 28 log-spaced frequencies ranging from 3 Hz to 30 Hz, with 2 cycles at the lowest frequency and 8 cycles at the highest frequency; in the high frequency range, a series of 56 log-spaced frequencies (ranging from 25 Hz to 80 Hz) was used, with 10 cycles at the lowest frequency and 20 cycles at the highest frequency. Power values were normalized with respect to a –2000 to –1600 ms pre-noun onset baseline and transformed into decibel scale ($10 \cdot \log_{10}$ of the signal), yielding the ERSP.

For statistical analysis, the cluster-based random permutation test implemented in the Fieldtrip (<http://fieldtrip.fcdonders.nl>) software package (Maris & Oostenveld, 2007) was used. This non-parametric statistical procedure optimally handles the multiple-comparisons problem. The permutation test was performed within 1900 ms window latency (from –1600 ms pre-noun onset to 300 ms post-noun onset) in a step of around 50 ms over 58 electrodes (AF7/AF8 and O1/O2 were deleted due to artifact). For every data point (electrode by time by frequency) (28 frequencies from 3 to 30 Hz for low frequency range; 56 frequencies from 25 to 80 Hz for high frequency range) of two conditions, a simple dependent-samples *t*-test is performed. All adjacent data points exceeding a preset significance level (0.05%) are grouped into clusters. Cluster-level statistics are 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.

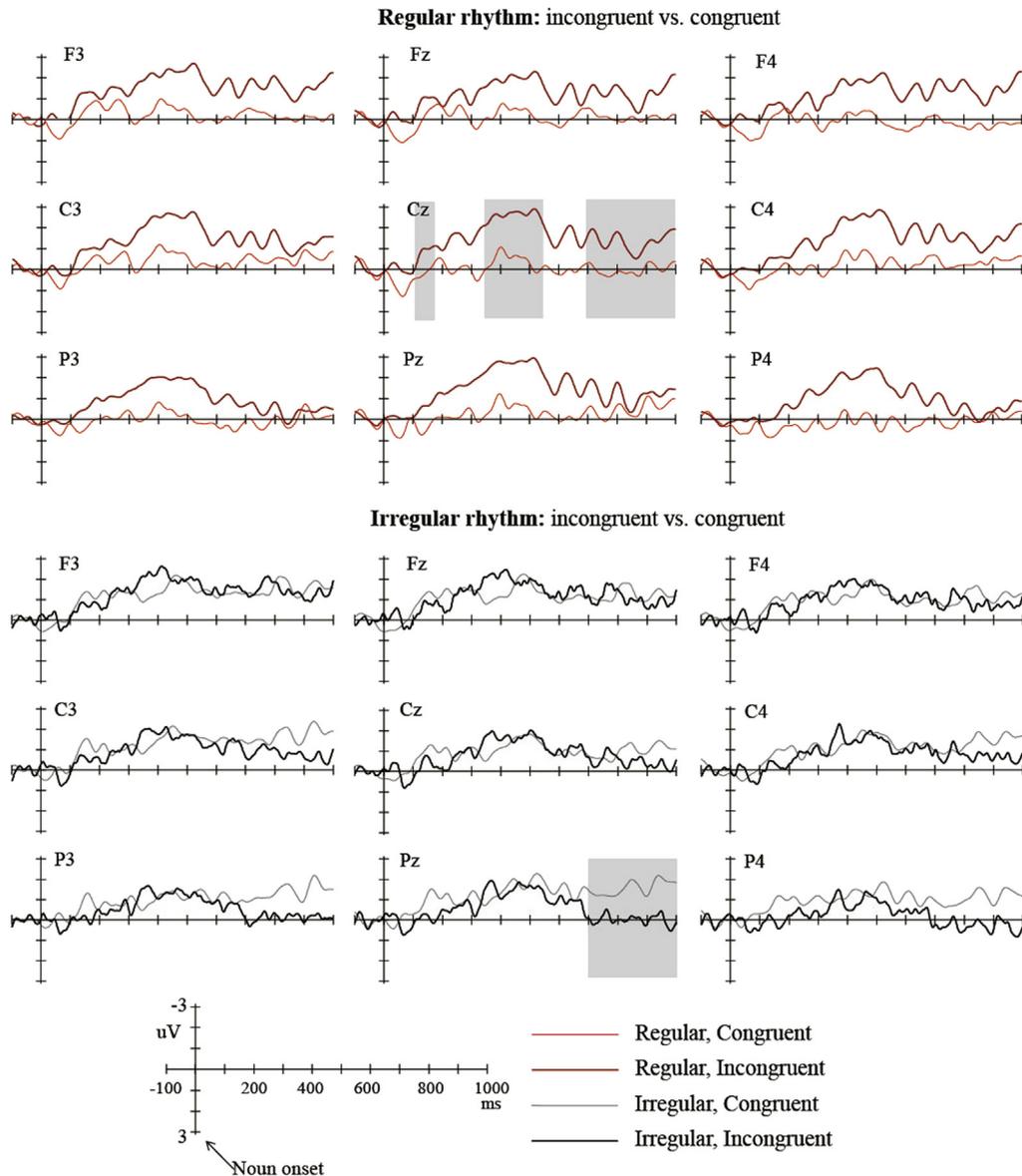


Fig. 1. Grand-average ERPs (time-locked to the critical nouns) evoked in the four experimental condition. The grey bars indicate the window latencies that show significant difference between the semantically incongruent and congruent nouns.

3. Results

3.1. Behavioral results

For all of the participants, the accuracy rate on the questions was higher than 85%, which indicated that the participants indeed read the sentence for comprehension during the online EEG recording.

3.2. ERP results (time-locked to the critical nouns)

Within the window latencies of 110–170 ms (see Fig. 1), the ANOVAs revealed a significant two-way interaction between Rhythm and Semantic congruency ($F_{midline}(1,17) = 7.12, p < .05, F_{lateral}(1,17) = 8.08, p < .01$). Further simple-analysis showed that semantically incongruent nouns elicited a significantly larger N1 (relative to semantically congruent nouns) when the preceding context carried a regular rhythm ($F_{midline}(1,17) = 5.36, p < .05, F_{lateral}(1,17) = 6.67, p < .05$) but not when the preceding rhythm context is irregular ($F_{midline}(1,17) = 2.36, p = .151, F_{lateral}(1,17) = 2.73, p = .117$).

The ANOVAs within 350–550 ms revealed a significant main effect of Semantic congruency ($F_{midline}(1,17) = 13.97, p < .005$,

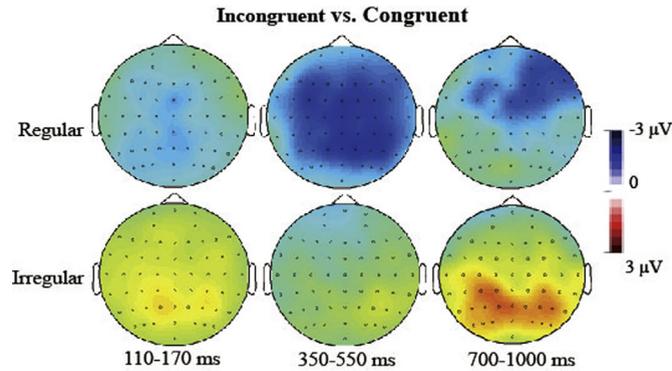


Fig. 2. Topographies of the ERP effects (incongruent vs. congruent) evoked in the case of regular rhythm or irregular rhythm.

$F_{lateral}(1,17) = 9.4, p < .01$), indicating that semantically incongruent nouns evoked a larger N400 than semantically congruent nouns. Furthermore, this main effect was qualified by a significant two-way interaction between Rhythm and Semantic congruency ($F_{midline}(1,17) = 4.28, p = .054, F_{lateral}(1,17) = 9.40, p < .01$), due to the fact that the enhanced N400 effect (incongruent vs. congruent) reached significance in the case of regular rhythm ($F_{midline}(1,17) = 13.73, p < .005, F_{lateral}(1,17) = 22.66, p < .0001$), but not in the case of irregular rhythm ($F_{midline}(1,17) = 0.67, p = .425, F_{lateral}(1,17) = 0.259, p = .622$).

Within the window latencies of 700–1000 ms, the ANOVAs still found a significant two-way interaction between Rhythm and Semantic congruency over the lateral electrodes ($F_{midline}(1,17) = 9.63, p < .01, F_{lateral}(1,17) = 11.43, p < .005$). Further analysis showed that when the sentence rhythm is regular, semantically incongruent nouns still evoked a larger negative deflection than semantically congruent nouns ($F_{midline}(1,17) = 9.78, p < .01, F_{lateral}(1,17) = 10.59, p < .005$); however, when the sentence rhythm is irregular, semantically incongruent nouns evoked a larger positive deflection (namely, P600) relative to semantically congruent nouns ($F_{midline}(1,17) = 2.46, p = .135, F_{lateral}(1,17) = 4.46, p < .05$) (see Figs. 1 and 2).

In summary, when the longer-scale rhythm context is regular, incongruent noun (compared to congruent nouns) elicited an enhanced N1 and an enhanced N400, with this enhanced negativity still being significant within a later time window; when the longer-scale rhythm context is irregular, incongruent noun (relative to congruent nouns) only evoked an enhanced P600.

3.3. Time-frequency results (time-locked to the critical nouns)

The permutation tests showed that, as compared to the irregular-rhythm condition, the regular-rhythm condition induced beta-band power increases ($p = .025$, around 17–24 Hz) just before the onset of the critical nouns (from 150 ms before the acoustic onset of the critical nouns to 50 ms after the acoustic onset of these nouns) and alpha-band power increases ($p = .025$, around 11–14 Hz) immediately following the onset of the critical nouns (from 50 ms to 150 ms post-noun onset) (see Fig. 3).

3.4. Correlation analysis

We further examined the relationship between the neural oscillatory activities induced by the preceding rhythmic context (namely, the beta power enhancement effect) and the ability to detect (or resolve) the subsequent incongruent nouns. For the correlation analysis, one factor is the averaged beta power increases (regular-rhythm vs. irregular-rhythm) over 18 frontal-central electrodes (F3/F1/F2/F4; FC3/FC1/FC2/FC4; C3/C1/C2/C4, CP3/CP1/CP2/CP4, FZ, CZ) around the onset of the critical nouns (from 150 ms pre-noun onset to 50 ms post-noun onset). Another factor is the ERP enhancement effects evoked by incongruent nouns as compared to congruent nouns ('ReIncon vs. ReCon' or 'IrreIncon vs. IrreCon'). Specifically, the ERP enhancement effects included in this correlation analysis are the N1-enhancement ('ReIncon vs. ReCon', averaged within 110–170 ms over frontal-central electrodes 'F3, FZ, FC3, C3, CZ, CP3, P3, PZ, PO3'), N400-enhancement ('ReIncon vs. ReCon', averaged within 350–550 ms over central-parietal electrodes 'C5/C3/C4/C6; CP5/CP3/CP4/CP6; P5/P3/P4/P6; PO5/PO3/PO4/PO6; CZ/PZ'), later-negativity-enhancement ('ReIncon vs. ReCon', averaged within 700–1000 ms over frontal-central electrodes 'AF4, F3, FZ, F2, F4, F6, FC3, FC2, FC4, FC6, CZ, C2, C4, CP4, CP6'), and P600-enhancement ('IrreIncon vs. IrreCon', averaged within 700–1000 ms over frontal-central electrodes 'CP5/CP6, CP3/CP4, P5/P6, P3/P4, P1/P2, PO5/PO6, PO3/PO4') effects, respectively. Two-tailed correlation analysis demonstrated that beta power increases around noun-onset were significantly correlated with the N400 enhancement effect ($r = 0.635, p < .005$) and the later-negativity enhancement effect ($r = 0.475, p = .048$) in the case of regular rhythm; moreover, beta power increases around noun-onset were also marginally significantly correlated with the P600 enhancement effect in the case of irregular rhythm ($r = -.463, p = .053$). That is, the larger the neural responses to rhythm regularity were, the smaller the later ERP responses (N400, later-negativity, or P600) to semantically incongruent nouns became (see Fig. 4).

4. Discussion

The present EEG study, with Mandarin Chinese spoken sentences as materials, examined the effect of rhythm regularity on speech

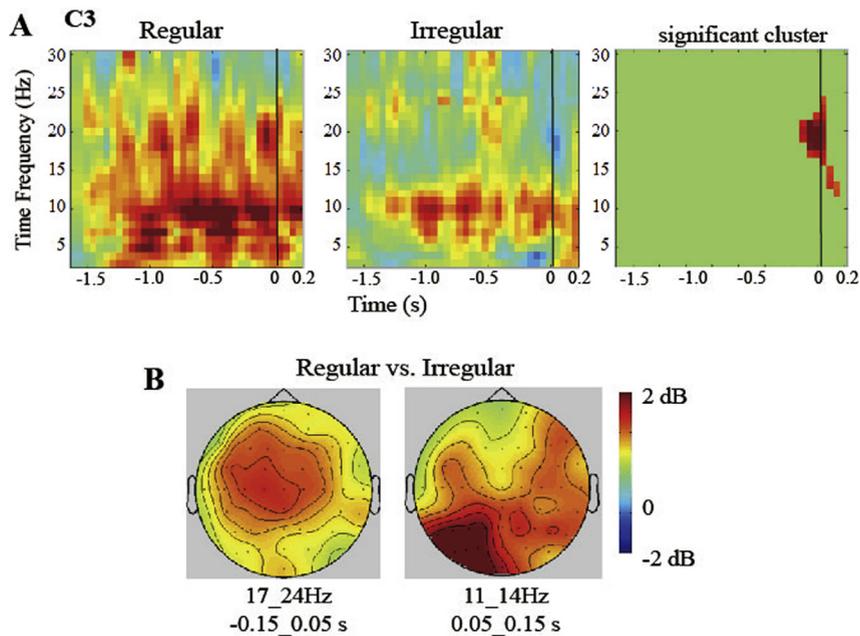


Fig. 3. Time-frequency analysis of electroencephalogram series in the regular- and irregular-rhythm conditions. The vertical line indicates the onset of the critical nouns. A: *Left*, event-related spectral perturbation (ERSP) from electrode C3 in the regular-rhythm condition; *Middle*, event-related spectral perturbation (ERSP) from electrode C3 in the irregular-rhythm condition; *Right*, the significant cluster that the permutation tests revealed (regular vs. irregular). B: Topographies of the rhythm regularity effects over beta-band power (17–24 Hz, from 0.15 s pre-noun onset to 0.05 s post-noun onset) and over alpha-band power (11–14 Hz, within 0.05–0.15 s post-noun onset).

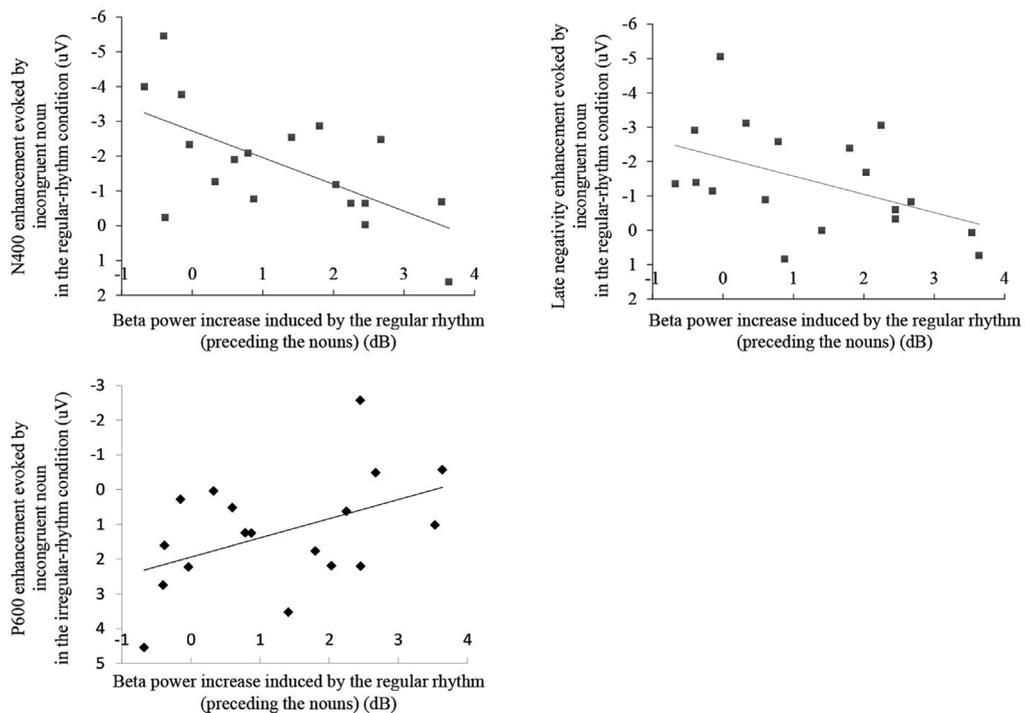


Fig. 4. Correlation between beta-power increases induced by rhythm regularity (regular vs. irregular rhythm) around the onset of the critical nouns (from 0.15 s pre-noun onset to 0.05 s post-noun onset) and later ERP responses (N400, later negativity, or P600) evoked by the incongruent nouns (compared to the congruent nouns).

comprehension and the neural oscillatory mechanisms underlying this effect. The results showed that, first, when the preceding context has a regular rhythm, as compared to congruent nouns, semantically incongruent nouns evoked an enhanced N1 (within 110–170 ms post-noun onset) and an enhanced N400 (within 350–550 ms), with this enhanced negativity still being significant within a later time window (namely, within 700–1000 ms); in contrast, when the preceding context has an irregular rhythm, semantically incongruent nouns only elicited an enhanced P600 within a later time window (namely, within 700–1000 ms post-noun onset). Secondly, as compared to the irregular-rhythm context, the regular-rhythm context induced power increases in the beta frequency band just preceding the onset of the critical nouns (from 150 ms pre-noun onset to 50 ms post-noun onset) and power increases in the alpha band immediately following the onset of these nouns (50–150 ms post-noun onset). Moreover, the beta-power enhancement effect induced by the preceding rhythmic context was significantly or marginally significantly correlated with the later ERP responses (N400, later negativity, or P600) to the subsequent incongruent nouns.

One aim of the present study is to examine whether and how rhythm regularity affects Mandarin Chinese speech processing. The present results showed that semantically incongruent nouns were processed differently in the contexts of regular and irregular rhythm. In the case of regular rhythm, incongruent nouns (relative to congruent nouns) evoked enhanced ERP responses already within the window latencies of N1 and N400; however, in the case of irregular rhythm, incongruent nouns did not elicit significant ERP responses until the window latency of P600. This earlier ERP response to words embedded in a regular-rhythm context (compared to an irregular-rhythm context) suggests that, during Mandarin Chinese speech comprehension, rhythm regularity can speed up the processing of incoming words. Previous Mandarin Chinese reading comprehension studies have already found that a violation of the rhythm structure of a verb + noun phrase (e.g., 2syllables + 1syllable) impeded semantic processing (Luo & Zhou, 2010; Luo et al., 2015). For the experimental materials used in the present study, although the phonological foot types (mono-, bi, or trisyllabic phonological foot) are semi-randomly distributed in the irregular rhythm condition, the phrase-level rhythm violation (e.g., 2syllables + 1syllable pattern of verb + noun phrase) never occurred. Moreover, the sentences used in the present study are all naturally produced spoken sentences, and the rhythm regularity is consequently achieved not through strict acoustic regularity (e.g., constant duration of each word) but by abstract phonological-foot regularity. The present results indicate that, during Mandarin Chinese speech comprehension, even though there is no phrase-level rhythm violation and no strict acoustic regularity, listeners still can extract rhythm regularity carried by a relatively longer-scale sentence context, and this kind of rhythm regularity is able to facilitate the processing of upcoming speech input.

Which stages of speech processing are affected by the relatively longer-scale rhythm regularity? In the present study, the semantically incongruent nouns elicited a larger N1 than the congruent one only in the case of regular rhythm. The N1 component has been found to be sensitive to acoustic aspects of individual stimulus (Näätänen & Picton, 1987) and to phonological features present in the speech signal as well (Obleser et al., 2004, 2003; Parviainen et al., 2004; Sanders & Neville, 2003). We argued that the N1 effect observed here is not driven by pure lexical differences or acoustic differences between the congruent and incongruent nouns, as the critical nouns of the present study are fully counterbalanced between the congruent and incongruent conditions, and the acoustic cues of these nouns also do not differ between these two conditions (see the method section for the detailed description of the acoustic parameters). Given the experimental design of the present study and the cognitive significance associated with the N1 component, a more plausible explanation of this N1 effect is that when the preceding context carries a regular rhythm, listeners can immediately detect the mismatch between the acoustic/phonological expectation generated from the preceding semantic context and the actual acoustic features extracted from the bottom-up signals. That is, the contextually incongruent words can be detected at the early stage of acoustic-phonological processing when they are imbedded in a rhythmical sentence context. In addition, the larger N400 (or P600) elicited by incongruent nouns suggest that the later semantic integration or reanalysis process is more difficult for incongruent nouns, as the N400 and P600 components have been considered to be related to the difficulty of semantic processing (e.g., Hagoort & Brown, 2000; Kutas & Van Petten, 1994) and the cost of reanalysis (e.g., Wang et al., 2009; Wolff et al., 2008) respectively. Specifically, in the context of a regular rhythm, the listeners have tried to integrate the incongruent nouns into the preceding context already with the window latency of N400; however, in the context of an irregular rhythm, the listeners do not detect and reanalyze the incongruent nouns until the window latency of P600. Overall, the present results suggest that the rhythm regularity of sentence context affects speech comprehension by modulating both the later semantic stage of processing and the early acoustic-phonological stage of processing as well.

Another important aim of the present study is to examine the neural oscillatory mechanisms underlying the facilitating effect of rhythm regularity on speech comprehension. The present results observe power increases (regular vs. irregular) in the beta band immediately preceding the onset of the critical nouns (starting already from 150 ms before the acoustic onset of these nouns) and power increases in the alpha band immediately after the onset of these nouns. In the present study, the critical nouns and the word immediately preceding these nouns are exactly the same in the regular and irregular rhythm conditions; moreover, the acoustic cues of these words showed no significant difference between these two conditions. Therefore, alpha-beta power increases observed here are not driven by low-level acoustic differences, but instead have their own cognitive significance. Power increases in the beta frequency band have been found to be related to top-down control process, such as maintaining the current cognitive/neural set or neural preparation for temporally expected stimuli (e.g., Fujioka et al., 2012; Gulberti et al., 2015; Kilavik, Zaepffel, Brovelli, MacKay, & Riehle, 2013; Lewis et al., 2015; Merchant, Grahn, Trainor, Rohrmeier, & Fitch, 2015; Patel & Iversen, 2014; Teki, 2014; de Lange et al., 2013). Given the reasons mentioned above, beta power increases observed here suggest that the human brain might have been entrained to the rhythmic temporal structure of the sentence context, and is maintaining the current cognitive process (e.g., segmenting the speech stream in a two-syllable by two-syllable manner) to be better prepared for the processing of upcoming speech input. In line with beta power increases, the subsequent alpha power increases suggest that attention alertness states (or neural excitability) of the lower perceptual systems might increase right after the presence of a critical noun, as power increases of

alpha-band neural oscillations have already been associated with the general attention alertness state (e.g., Sadaghiani et al., 2010) and specifically with the inhibiting functions of attention (e.g., for reviews see Klimesch, 2012 and Jensen & Mazaheri, 2010).

The alpha-beta band of power increases induced by the rhythmical context indeed exerted an influence on the processing of upcoming words in speech. First, probably due to the better neural preparation and increases of attention alert state in the regular rhythm context (as indicated by alpha-beta power increases), upcoming incongruent nouns can be detected at the early acoustic-perceptual stage of processing only in the rhythmical context. Secondly, the preceding beta-band oscillatory synchronization has a more direct relationship with the later semantic processing of the following speech input. The larger the beta power increase induced by preceding rhythmic context is, the smaller the later ERP responses (N400 or P600) evoked by the subsequent incongruent nouns become. The negative correlation between the preceding beta-power-increase effect and the following N400/P600-enhancement effect suggests that it is more possible that the more top-down processing the human brain relies on the regular rhythm context to conduct, the smaller the cognitive costs caused by resolving or integrating the semantically incongruent words become. Overall, the facilitating effect of rhythm regularity on speech processing is closely related to the alpha-beta band of neural oscillations, especially to beta activity. A rhythmical speech context (and the underlying neural oscillatory activities) facilitates the processing of upcoming speech input by increasing sensory neuron excitability associated with early acoustic-perceptual stage (hence facilitating early detection of incongruent words) and by decreasing cognitive costs associated with later semantic processing (hence facilitating semantic resolution or reanalysis of incongruent words).

As mentioned in the introduction, the existing studies have demonstrated that, during Indo-European language comprehension, rhythmic context realized by regular distribution of syllable stress or regular lengthening of syllable duration can facilitate syntactic or semantic processing (Cason & Schön, 2012; Magne et al., 2007; Roncaglia-Denissen et al., 2013; Rothermich et al., 2012; Schmidt-Kassow & Kotz, 2009a, 2009b). The present study extends the facilitating effect of rhythm regularity on speech comprehension to Mandarin Chinese processing, in which rhythm regularity is realized by regular distribution of phonological foot types. Moreover, the present study provides further insights on the cognitive mechanisms underlying this facilitating effect by showing that rhythm regularity facilitates both the later semantic processing and early sensory processing of incoming speech input.

The above differences in the effect of rhythm regularity on speech processing might be related to the specific speech rhythm properties. In the early studies with Indo-European language as materials, speech rhythm is defined as the regularity of surface prosodic prominence (e.g., syllable stress in English; Rothermich et al., 2012) or surface timing (e.g., syllable time in French; Magne et al., 2007). In contrast, in the present study, the rhythmical structure is realized not just by surface acoustic periodicity, but instead directly by regularity on the number of syllables per word. Consequently, the speech rhythm of the present study (compared to previous studies) is more directly related to word segmentation and lexical processing, which might be the reason why, in the present study, the facilitating effect of rhythmic context on speech processing occurs at the early acoustic-perceptual stage of processing. That is, the specific speech rhythm properties might affect its effect on speech processing and the underlying cognitive mechanisms, which need to be examined further in future studies.

The dynamic attending theory proposes that the temporal rhythm structure of auditory/visual sequences can directly affect neural excitability and stimulus processing by modulating internal neural oscillatory activities (e.g., Jones & Boltz, 1989; Large & Jones, 1999). In line with this theory, some auditory/visual perception studies have demonstrated that external rhythmic stimuli not only modulate behavioral responses, but also directly modulate very early perceptual coding (e.g., Jepma et al., 2012; Rohenkohl et al., 2012) and cortical excitability in the visual/auditory sensory brain areas (e.g., Arnal et al., 2015; Besle et al., 2011; Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008; Large & Jones, 1999; Wöstmann et al., 2016). The present study provides further experimental evidence for the dynamic attending theory by showing that, even during complex spoken sentence comprehension, both the later semantic processing and very early acoustic-phonological processing are facilitated by the temporal rhythm structure of speech signals. Moreover, the present results also enable us to gain further understanding of the neural oscillatory mechanisms by which rhythm regularity affects speech comprehension. That is, the facilitating effect of rhythm regularity on speech comprehension in part relies on alpha-beta band neural oscillations, especially on beta activity, with beta oscillations directly predicting the ease of subsequent semantic processing.

5. Conclusion

The present EEG study examined whether and how the rhythm regularity of sentence context affects Mandarin Chinese speech comprehension. Our results showed that a rhythmical context realized by the regular distribution of phonological foot types speeds up the processing of Mandarin Chinese sentences. Rhythm regularity facilitates speech comprehension both by enhancing neural excitability associated with early acoustic-phonological processing, and by reducing cognitive costs associated with later semantic processing. Moreover, the facilitating effect of rhythm regularity on speech comprehension relies at least in part on alpha-beta band neural oscillations, especially on beta activity, with beta oscillations directly predicting the ease of semantic resolution of upcoming words in speech.

Author contributions

Xiaoqing Li designed the experiment, analyzed the data, and wrote the paper. Xiaoying Xu contributed a lot to prepare the experimental sentences. Jinyan Xia and Ximing Shao collected the data and co-analyzed the data with Xiaoqing Li.

Conflicts of interest

The author has declared that no conflict of interest exists.

Acknowledgments

This work was supported by the State Key Program of the National Natural Science Foundation of China [61433018]; and Beijing Advanced Innovation Center for Imaging Technology [BAICIT-2016018].

References

- Arnal, L. H., Doelling, K. B., & Poeppel, D. (2015). Delta–beta coupled oscillations underlie temporal prediction accuracy. *Cerebral Cortex*, 25(9), 3077–3085.
- Bailly, G. (1989). Integration of rhythmic and syntactic constraints in a model of generation of French prosody. *Speech Communication*, 8(2), 137–146.
- Besle, J., Schevon, C. A., Mehta, A. D., Lakatos, P., Goodman, R. R., McKhann, G. M., ... Schroeder, C. E. (2011). Tuning of the human neocortex to the temporal dynamics of attended events. *Journal of Neuroscience*, 31(9), 3176–3185.
- Bornkessel, I., McElree, B., Schlesewsky, M., & Friederici, A. D. (2004). Multi-dimensional contributions to garden path strength: Dissociating phrase structure from case marking. *Journal of Memory and Language*, 51(4), 495–522.
- Cason, N., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia*, 50(11), 2652–2658.
- Dalla Bella, S., Białuńska, A., & Sowiński, J. (2013). Why movement is captured by music, but less by speech: Role of temporal regularity. *PLoS One*, 8(8) e71945.
- Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, 19(1), 158–164.
- Donner, T. H., Siegel, M., Fries, P., & Engel, A. K. (2009). Buildup of choice-predictive activity in human motor cortex during perceptual decision making. *Current Biology*, 19(18), 1581–1585.
- Endress, A. D., & Hauser, M. D. (2010). Word segmentation with universal prosodic cues. *Cognitive Psychology*, 61(2), 177–199.
- Feng, S. (1998). Prosodic structure and compound words in Classical Chinese. *New approaches to Chinese word formation: Morphology, phonology and the lexicon in modern and ancient Chinese* (pp. 197–260).
- Fowler, C. A. (1979). “Perceptual centers” in speech production and perception. *Attention, Perception, & Psychophysics*, 25(5), 375–388.
- Fujioka, T., Trainor, L. J., Large, E. W., & Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. *Journal of Neuroscience*, 32(5), 1791–1802.
- Goupillaud, P., Grossmann, A., & Morlet, J. (1984). Cycle-octave and related transforms in seismic signal analysis. *Geoexploration*, 23(1), 85–102.
- Gulberti, A., Moll, C. K. E., Hamel, W., Buhmann, C., Koeppen, J. A., Boelmans, K., ... Engel, A. K. (2015). Predictive timing functions of cortical beta oscillations are impaired in Parkinson's disease and influenced by L-DOPA and deep brain stimulation of the subthalamic nucleus. *NeuroImage: Clinical*, 9, 436–449.
- Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech: Semantic ERP effects. *Neuropsychologia*, 38(11), 1518–1530.
- Hoequist, C. E., Jr. (1983). The perceptual center and rhythm categories. *Language and Speech*, 26(4), 367–376.
- Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience*, 4.
- Jepma, M., Wagenmakers, E. J., & Nieuwenhuis, S. (2012). Temporal expectation and information processing: A model-based analysis. *Cognition*, 122(3), 426–441.
- Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, 96(3), 459.
- Juszyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences*, 3(9), 323–328.
- Kilavik, B. E., Zaepffel, M., Brovelli, A., MacKay, W. A., & Riehle, A. (2013). The ups and downs of beta oscillations in sensorimotor cortex. *Experimental Neurology*, 245, 15–26.
- Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, 16(12), 606–617.
- Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified. *Handbook of psycholinguistics* (pp. 83–143).
- Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, 320(5872), 110–113.
- Lakatos, P., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., & Schroeder, C. E. (2009). The leading sense: Supramodal control of neurophysiological context by attention. *Neuron*, 64(3), 419–430.
- de Lange, F. P., Rahnev, D. A., Donner, T. H., & Lau, H. (2013). Prestimulus oscillatory activity over motor cortex reflects perceptual expectations. *Journal of Neuroscience*, 33(4), 1400–1410.
- Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106(1), 119.
- Lee, C. S., & Todd, N. P. M. (2004). Towards an auditory account of speech rhythm: Application of a model of the auditory ‘primal sketch’ to two multi-language corpora. *Cognition*, 93(3), 225–254.
- Lewis, A. G., Schoffelen, J. M., Schriefers, H., & Bastiaansen, M. (2016). A predictive coding perspective on beta oscillations during sentence-level language comprehension. *Frontiers in Human Neuroscience*, 10.
- Lewis, A. G., Wang, L., & Bastiaansen, M. (2015). Fast oscillatory dynamics during language comprehension: Unification versus maintenance and prediction? *Brain and Language*, 148, 51–63.
- Luo, Y., Duan, Y., & Zhou, X. (2015). Processing rhythmic pattern during Chinese sentence reading: An eye movement study. *Frontiers in Psychology*, 6.
- Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, 54(6), 1001–1010.
- Luo, Y., & Zhou, X. (2010). ERP evidence for the online processing of rhythmic pattern during Chinese sentence reading. *NeuroImage*, 49(3), 2836–2849.
- Magne, C., Astésano, C., Aramaki, M., Ystad, S., Kronland-Martinet, R., & Besson, M. (2007). Influence of syllabic lengthening on semantic processing in spoken French: Behavioral and electrophysiological evidence. *Cerebral Cortex*, 17(11), 2659–2668.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190.
- Mattys, S. L., & Samuel, A. G. (1997). How lexical stress affects speech segmentation and interactivity: Evidence from the migration paradigm. *Journal of Memory and Language*, 36(1), 87–116.
- Merchant, H., Grahn, J., Trainor, L., Rohrmeier, M., & Fitch, W. T. (2015). Finding the beat: A neural perspective across humans and non-human primates. *Philosophical Transactions of the Royal Society B*, 370(1664), 20140093.
- Morton, J., Marcus, S., & Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review*, 83(5), 405.
- Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, 24(4), 375–425.
- Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication*, 41(1), 233–243.
- Oblaser, J., Elbert, T., Lahiri, A., & Eulitz, C. (2003). Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. *Cognitive Brain Research*, 15(3), 207–213.
- Oblaser, J., Lahiri, A., & Eulitz, C. (2004). Magnetic brain response mirrors extraction of phonological features from spoken vowels. *Journal of Cognitive Neuroscience*, 16(1), 31–39.
- Parviainen, T., Helenius, P., & Salmelin, R. (2004). Cortical differentiation of speech and nonspeech sounds at 100 ms: Implications for dyslexia. *Cerebral Cortex*, 15(7), 1054–1063.
- Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: The action simulation for auditory prediction (ASAP) hypothesis.

- Frontiers in Systems Neuroscience*, 8 <https://doi.org/10.3389/fnsys.2014.00057>.
- Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, 23(6), 1378–1387.
- Roach, P. (1982). On the distinction between ‘stress-timed’ and ‘syllable-timed’ languages. *Linguistic Controversies*, 73–79.
- Rohenkohl, G., Cravo, A. M., Wyart, V., & Nobre, A. C. (2012). Temporal expectation improves the quality of sensory information. *Journal of Neuroscience*, 32(24), 8424–8428.
- Roncaglia-Denissen, M. P., Schmidt-Kassow, M., & Kotz, S. A. (2013). Speech rhythm facilitates syntactic ambiguity resolution: ERP evidence. *PLoS One*, 8(2) e56000.
- Rothermich, K., & Kotz, S. A. (2013). Predictions in speech comprehension: fMRI evidence on the meter–semantic interface. *Neuroimage*, 70(4), 89–100.
- Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm’s gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia*, 50(2), 232–244.
- Sadaghiani, S., Scheeringa, R., Lehongre, K., Morillon, B., Giraud, A. L., & Kleinschmidt, A. (2010). Intrinsic connectivity networks, alpha oscillations, and tonic alertness: A simultaneous electroencephalography/functional magnetic resonance imaging study. *Journal of Neuroscience*, 30(30), 10243–10250.
- Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing: I. Segmentation, semantics, and syntax in native speakers. *Cognitive Brain Research*, 15(3), 228–240.
- Schlesewsky, M., & Bornkessel, I. (2006). Context-sensitive neural responses to conflict resolution: Electrophysiological evidence from subject–object ambiguities in language comprehension. *Brain Research*, 1098(1), 139–152.
- Schmidt-Kassow, M., & Kotz, S. A. (2009a). Attention and perceptual regularity in speech. *NeuroReport*, 20(18), 1643–1647.
- Schmidt-Kassow, M., & Kotz, S. A. (2009b). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *Journal of Cognitive Neuroscience*, 21(9), 1693–1708. <https://doi.org/10.1162/jocn.2008.21153>.
- Teki, S. (2014). Beta drives brain beats. *Frontiers in Systems Neuroscience*, 8 155–155.
- Thiessen, E. D., & Saffran, J. R. (2007). Learning to learn: Infants’ acquisition of stress-based strategies for word segmentation. *Language Learning and Development*, 3(1), 73–100.
- Wang, L., Schlesewsky, M., Bickel, B., & Bornkessel-Schlesewsky, I. (2009). Exploring the nature of the ‘subject’-preference: Evidence from the online comprehension of simple sentences in Mandarin Chinese. *Language & Cognitive Processes*, 24(7–8), 1180–1226.
- Wenk, B. J., & Wioland, F. (1982). Is French really syllable-timed? *Journal of Phonetics*, 10(2), 193–216.
- Wolff, S., Schlesewsky, M., Hirotsu, M., & Bornkessel-Schlesewsky, I. (2008). The neural mechanisms of word order processing revisited: Electrophysiological evidence from Japanese. *Brain and Language*, 107(2), 133–157.
- Wöstmann, M., Herrmann, B., Maess, B., & Obleser, J. (2016). Spatiotemporal dynamics of auditory attention synchronize with speech. *Proceedings of the National Academy of Sciences*, 113(14), 3873–3878.
- Wu, W. S. (2005). Tonal patterns in classical Chinese poetry, stress and confirmation of enclitic weakened stress in a rhythm unit in the construction of metre in Chinese. *Studies in Language and Linguistics*, 25(3), 90–94.