NEUROSCIENCE
RESEARCH ARTICLE
C. Pei et al. / Neuroscience 502 (2022) 1–9

Auditory Dominance in Processing Chinese Semantic Abnormalities in Response to Competing Audio-visual Stimuli

Changfu Pei, [a,b†] **Xunan Huang,** [a,b,c†] **Yuqin Li,** [a,b] **Baodan Chen,** [a,b] **Bin Lu,** [a,b] **Yueheng Peng,** [a,b] **Yajing Si,** [d] **Xiabing Zhang,** [a,b] **Tao Zhang,** [e] **Dezhong Yao,** [a,b] **Fali Li** [a,b*] **and Peng Xu** [a,b*]

[a] *The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Lab for NeuroInformation, University of Electronic Science and Technology of China, Chengdu 611731, China*

[b] *School of Life Science and Technology, Center for Information in BioMedicine, University of Electronic Science and Technology of China, Chengdu 611731, China*

[c] *School of Foreign Languages, University of Electronic Science and Technology of China, Sichuan, Chengdu 611731, China*

[d] *School of Psychology, Xinxiang Medical University, Xinxiang 453003, China*

[e] *Mental Health Service and Education Center, School of Science, Xihua University, Chengdu 610039, China*

Abstract—Language is a remarkable cognitive ability that can be expressed through visual (written language) or auditory (spoken language) modalities. When visual characters and auditory speech convey conflicting information, individuals may selectively attend to either one of them. However, the dominant modality in such a competing situation and the neural mechanism underlying it are still unclear. Here, we presented participants with Chinese sentences in which the visual characters and auditory speech convey conflicting information, while behavioral and electroencephalographic (EEG) responses were recorded. Results showed a prominent auditory dominance when audio-visual competition occurred. Specifically, higher accuracy (ACC), larger N400 amplitudes and more linkages in the posterior occipital-parietal areas were demonstrated in the auditory mismatch condition compared to that in the visual mismatch condition. Our research illustrates the superiority of the auditory speech over the visual characters, extending our understanding of the neural mechanisms of audio-visual competition in Chinese.

Key words: EEG, Chinese, brain networks, auditory dominance, N400.

INTRODUCTION

Multisensory integration is generally denoted as the set of processes by which information arriving from an individual sensory modality (e.g. vision, auditory sense) interacts and influences processing in other sensory modalities, including how these sensory inputs are combined together to yield a unified perceptual experience of multisensory events (Talsma et al., 2010). However, auditory and visual modalities are likely not of equal value in audio-visual interaction (De Gelder and Bertelson, 2003). More specifically, the spatial resolution of the visual system is better than that of the auditory system, while the temporal resolution of the auditory system is better than that of the visual system (Weihing et al., 2009; Robinson and Sloutsky, 2013; Lukas et al., 2014; Huang et al., 2015). Language comprehension typically requires multisensory information, namely spoken words and visual characters to access meaning. Information processing may be facilitated when characters and speech indicate the same percept (Raij et al., 2000). However, conflicting information, i.e., incongruent phonemic and graphemic input, often results in cross-modality competition (Robinson and Sloutsky, 2010). A critical question is which modality is dominant when auditory speech and visual characters stimuli clash.

Many studies have explored the mechanisms underlying audio-visual competition (Hirst et al., 2020). Some studies have demonstrated that there is sometimes an auditory advantage (Shams et al., 2000) and, in other cases, have shown visual mechanisms are dominant (Mishra et al., 2007). For example, L. Shams et al. found

*Corresponding authors. Address: The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Lab for NeuroInformation, University of Electronic Science and Technology of China, Chengdu 611731, China.
E-mail addresses: fali.li@uestc.edu.cn (F. Li), xupeng@uestc.edu.cn (P. Xu).
† These authors contributed equally to this work.

Abbreviations: RT, response time; AI, auditory incongruent condition; VI, visual incongruent condition; AVC, audio-visual congruent condition; EEG, electroencephalographic; ACC, accuracy; ERP, event-related potential; PLV, phase-locking value; *Cpv*, denotes the Cauchy principal value; SVO, subject – verb – object; REST, Reference Electrode Standardization Technique; UESTC, University of Electronic Science and Technology of China; ANOVA, one-way analyses of variance.

https://doi.org/10.1016/j.neuroscience.2022.08.017
0306-4522/

that the number of stimulus flickers seen visually was affected by the number of short sounds sensed auditorily (Shams et al., 2000). Mishra et al. proved that when an individual hears two consecutive short notes and sees a flash, it will produce a hallucinatory phenomenon when the second short note arrives, as the person believes he or she has seen two consecutive flashes (Mishra et al., 2007). In addition, most studies of audio-visual speech-related competition evaluate the outcomes of discrepancies between visual speech signals and auditory speech stimuli (McGurk and MacDonald, 1976; Peelle and Sommers, 2015; Keitel et al., 2020) and between speech and pictures or videos (Liu et al., 2011; Manfredi et al., 2018). A prevalent example is called the McGurk Effect (McGurk and MacDonald, 1976). Ordinarily, watching a speaker's face helps us understand what is being said because integrating the sight (lip movements) and sound of speech enhances the brain activity that underlies speech perception. However, if slightly mismatched cues are paired (such as the sound for 'ba' with the lip movements for 'ga'), the resulting synthesis yields an entirely different product ('da').

The N400 effect is typically sensitive to semantic integration difficulty (Kutas and Hillyard, 1980; Kutas and Federmeier, 2000, 2011; Hu et al., 2012). A substantial body of literature describes investigations of cross-modal semantic processing by analyzing the N400 component (Kutas and Federmeier, 2011). The N400 component can be evoked by several types of multimodal information, such as cooperative actions and gestures (Proverbio et al., 2014) and speech sounds with pictures or videos (Cummings et al., 2008; Liu et al., 2011). The N400 effect elicited by videos with semantically inconsistent speech was larger and later than that elicited by videos with semantically consistent natural sound (Liu et al., 2011). However, these studies did not further explore the N400 effect induced by different types of mismatched audio-visual stimuli. In addition to event-related potential (ERP) analysis, network connectivity analysis is another efficient way to measure brain activity. It treats cognitive processing as circuits rather than any brain region considered in isolation (Bassett and Sporns, 2017; Yi et al., 2021). The cognitive process could be reflected by not only the activity of various brain regions but also the information propagation and interactions between different functional areas in the human brain (Petersen and Sporns, 2015). As a higher-level cognitive process, language has also been involved with a large-scale network spanning related brain areas (Hagoort, 2019).

Chinese speakers usually integrate multisensorial information (e.g., auditory and visual) when perceiving Chinese, while to efficiently process different modalities of information, the differential regions are involved (Wu and Thierry, 2010). Before the development of logographic and alphabetic writing systems, human language relies mainly upon spoken utterances (Houston, 2004), and the advent of written language thus provides a new, visual pathway for communication. However, since written language requires extensive training and typically follows the acquisition of spoken language, it is thought to rely more on neural pathways that originally supported spoken language (van Atteveldt et al., 2004). In fact, the interference between different modalities and the competition of brain resources occurs objectively (Raij et al., 2000; Andres et al., 2011). Although there has been plenty studies that explore audio-visual competition (Shams et al., 2000; Mishra et al., 2007; Hirst et al., 2020), for non-Indo-European Chinese, it is still not clear whether written or spoken is more dominant in language processing, and there is thus little research on the modality advantage of Chinese. In our present study, to explore this issue, first, we hypothesized that auditory is dominant when conflicting auditory speech and visual characters stimuli are presented. We then developed new Chinese character-speech materials in which the linguistic constituent structure was dissociated from prosodic or statistical cues. Thereafter, individual behavioral and electrophysiological parameters, i.e., ERP and functional networks, were investigated to identify the mechanisms underlying the character-speech competition of Chinese.

EXPERIMENTAL PROCEDURES

Participants

We recruited twenty healthy, right-handed participants (11 males, mean age: 24.82 years). The participants were all students at the University of Electronic Science and Technology of China (UESTC). The participants had never used any psychoactive medication, and none of them had any personal or family history of psychiatric or neurological illnesses. All participants were native Chinese speakers with normal hearing and normal or corrected-to-normal visual acuity. None of the participants had a history of reading difficulties/dyslexia. Before the experiment, written informed consent was obtained when the participants fully understood the procedure. The institution research ethics board approved the experimental procedure of the UESTC.

Materials

Fifty-four-syllable sentences were constructed, in which the first two of the four syllables formed a noun phrase and the last two syllables formed a verb phrase. The stimuli were adapted from a previous study (Ding et al., 2016), where the duration length of all syllables was adjusted to 250 ms. Additionally, there was no extra gap between each syllable, which prevented the speech rate of any other prosodic cue from influencing the construction of the linguistic structure building. In our experiment, the sentences were presented to the subjects either intact or out of order under the audio-visual modality. There were three experimental conditions: audio-visual congruent condition (AVC), auditory incongruent condition (AI), and visual incongruent condition (VI). In the AI condition: the visual and auditory stimuli were presented simultaneously. Under the visual modality, the syllables presented meaningful sentences. At the same time, a sequence of randomly ordered syllables could not be combined into a meaningful constituent with its neighbours under the auditory modality. In the AVC condition: four Chinese

monosyllabic words were presented in sequences in an order such that two-phrase sentences were formed under the audio-visual modality; the first two syllables constituted a noun phrase and the last two syllables constituted a verb phrase. In the VI condition: in contrast to the AI condition, meaningful sentences were presented under the auditory modality, and a sequence of randomly ordered syllables was played isochronously under the visual modality.

Procedure

This study consisted of eyes-closed resting-state EEG recordings with a 5 min duration and a 15 min long audio-visual task. In the first 2 min, participants were instructed to take deep breaths to adapt to the experimental environment. Then, 5-min eyes-closed resting-state EEG data sets were recorded before the audio-visual task. This experiment consisted of four hundred trials randomly divided into three blocks in total (AVC: 240, VI: 80, and AI: 80; 400 trials in total). Each trial lasted a period of 3 s, starting with the presentation of a fixation cross for 600 ms. After the disappearance of the fixation crosshair, the stimuli consisted of the presentation of two-phrase sentence for 1 s. Then after the stimulus, the participants were required to judge false auditory information (press the button ‘‘1”), audio-visual congruent information (press the button ‘‘2”), or false visual information (press the button ‘‘3”). The trials in which participants could not press the key within 1200 ms or pressed other keys improperly were considered invalid. After the presentation of a blank for 200 ms, the subsequent trial was initiated. Before the formal experiment, all participants were required to complete a preliminary round to ensure that they clearly understood the experimental rules. The detailed procedure is depicted in Fig. 1.

EEG data acquisition

Participants were seated individually in an electrically shielded, dimly lit room. EEG data recordings were performed with 64 Ag/AgCl electrodes (ANT Neuro, Berlin, Germany). These electrodes were positioned according to the extended 10–20 system, and the data were recorded at a sampling rate of 500 Hz. The online filter band was set between 0.01 and 100 Hz. Electrodes CPz and AFz served as the reference and ground, respectively. Electrooculograms to monitor eye movements were recorded from one additional channel located above the left eye. During the entire experimental task, the impedances of the electrodes were kept below 5 kΩ.

EEG data analysis

The EEG data analysis was divided into data preprocessing, ERP extraction, and network analysis steps, as shown in Fig. 2. The details for each process are further provided in the following sections.

Data preprocessing

In the current study, the raw data were first re-referenced to the ‘‘zero” reference by using the Reference Electrode Standardization Technique (REST) toolbox (https://www.neuro.uestc.edu.cn/rest/) (Yao, 2001; Dong et al., 2017). Thereafter, following the protocols used in previous studies (Steinhauer et al., 1999; Si et al., 2019), a narrow band of 0.1–10 Hz was applied in the offline band filtering, and [−200, 1000] ms (0 ms denotes the stimulus onset) data segmentation and artifact trial removal with a threshold of ±90 μV were further performed to extract the artifact-free trials.

ERP component

Following preprocessing, segments ranging from −200 to 1000 ms were averaged for each condition. Peak detection was performed automatically within the N400 time window. Based on the averaged ERP, the N400 amplitude was calculated within a time interval of [400, 500] ms after the onsets of the target stimulus. And across the three modalities, the same time window was used. The N400 amplitude on a single electrode was defined as the mean amplitude within a ±10 ms window with the N400 peak as the centre at eight electrodes over the frontocentral, central, and parietal-occipital areas (i.e., FC1, FCz, FC2, Cz, C1, C2, Cp1, and Cp2), where the N400 component is classically found and displays maximal sensitivity (Luck, 2014; Christoffels et al., 2016). Finally, the individual N400 amplitude was calculated by averaging over the eight electrodes.

EEG network analysis

The phase-locking value (PLV) method was used to capture the nonlinear phase synchronization between two brain areas (Sakkalis, 2011; Sun et al., 2012). In the present study, the networks were constructed at the scalp level, which may have been influenced by the volume conduction effect (Srinivasan et al., 1998; Xu et al., 2014). To alleviate the volume conduction effect, we used the 21 canonical electrodes (Fp1, Fp2, F7, F3, T7, C3, P7, P3, O1, Fpz, F8, F4, T8, C4, P8, P4, O2, Oz, Pz, Cz, and Fz) of the 10–20 international system as network nodes and applied the PLV to calculate the functional connectivity for all the EEG segments. A high value represents strong phase synchronization. To estimate the corresponding instantaneous phases, i.e., $\phi_a(t)$ and $\phi_b(t)$ of two given time series, $a(t)$ and $b(t)$, the Hilbert transform was used to form the analytical signal $S(t)$ as.

$$\begin{cases} S_a(t) = a(t) + iT_a(t) \\ S_b(t) = b(t) + iT_b(t) \end{cases} \quad (1)$$

where $T_a(t)$ and $T_b(t)$ are the Hilbert transforms of two time series, $a(t)$ and $b(t)$, which are defined as,

$$\begin{cases} T_a(t) = \frac{1}{\pi} Cpv \int_{-\infty}^{\infty} \frac{a(t')}{t-t'} dt' \\ T_b(t) = \frac{1}{\pi} Cpv \int_{-\infty}^{\infty} \frac{b(t')}{t-t'} dt' \end{cases} \quad (2)$$

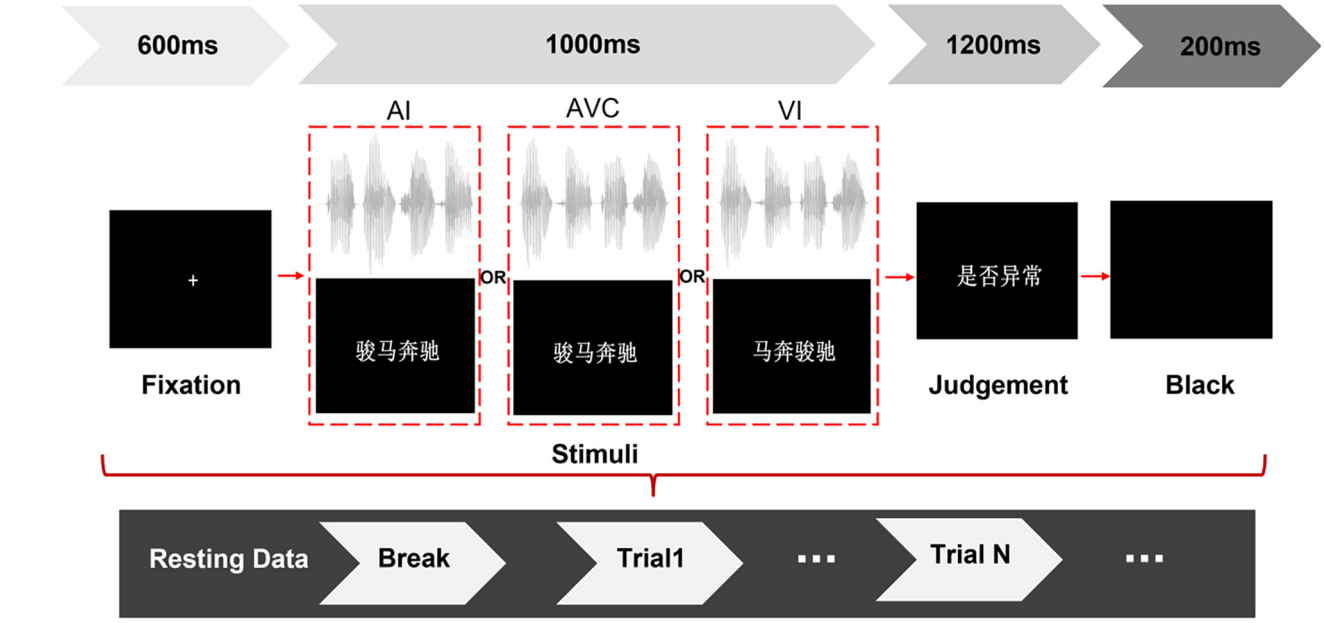

Fig. 1. Trial protocol in the experiment. A trial consisting of a 600 ms cue, 1000 ms stimulus, 1200 ms judgment period, and 200 ms break was included in each trial.

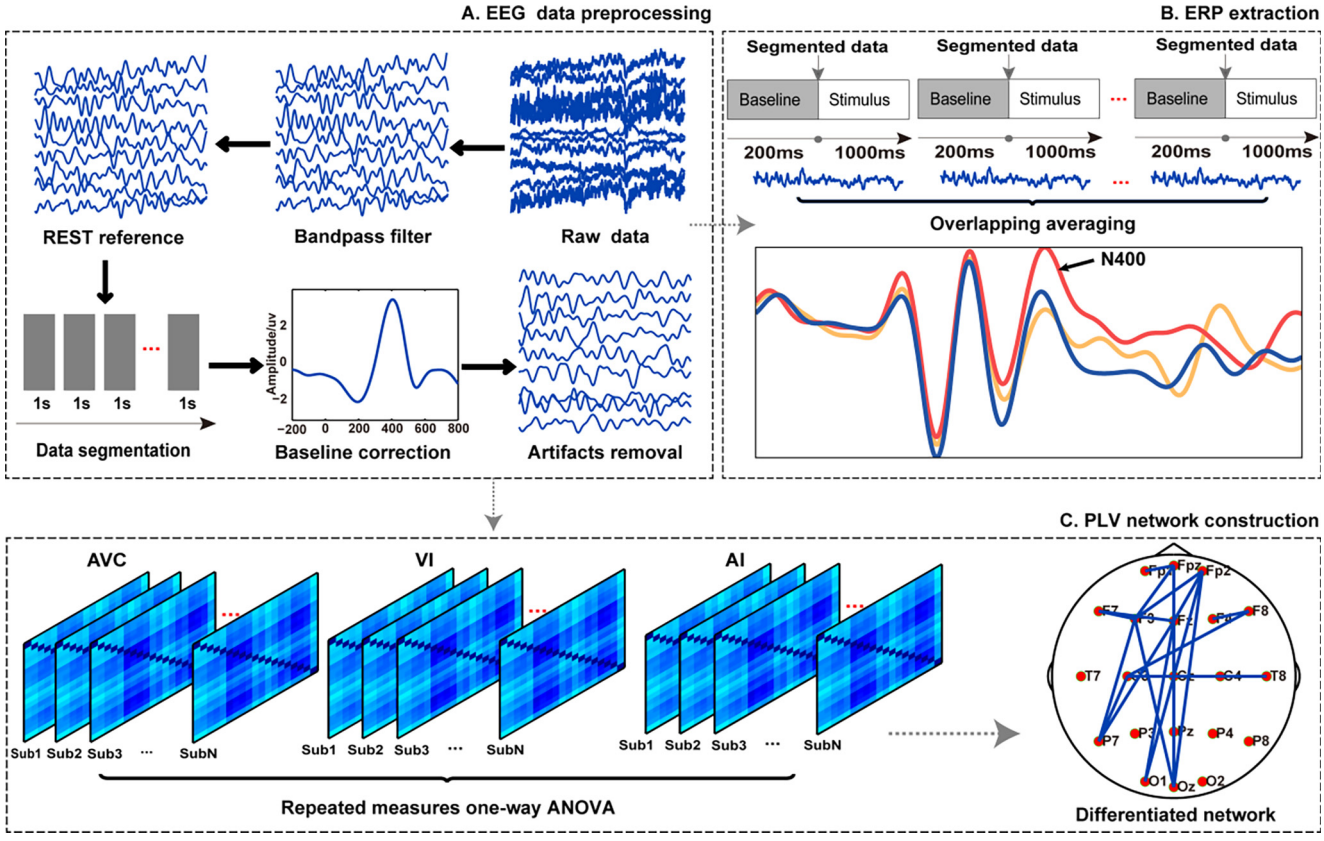

Fig. 2. The EEG data processing procedure. The data procedure consists of three phases: (**A**) EEG data preprocessing, (**B**) ERP extraction, and (**C**) PLV network pattern analysis.

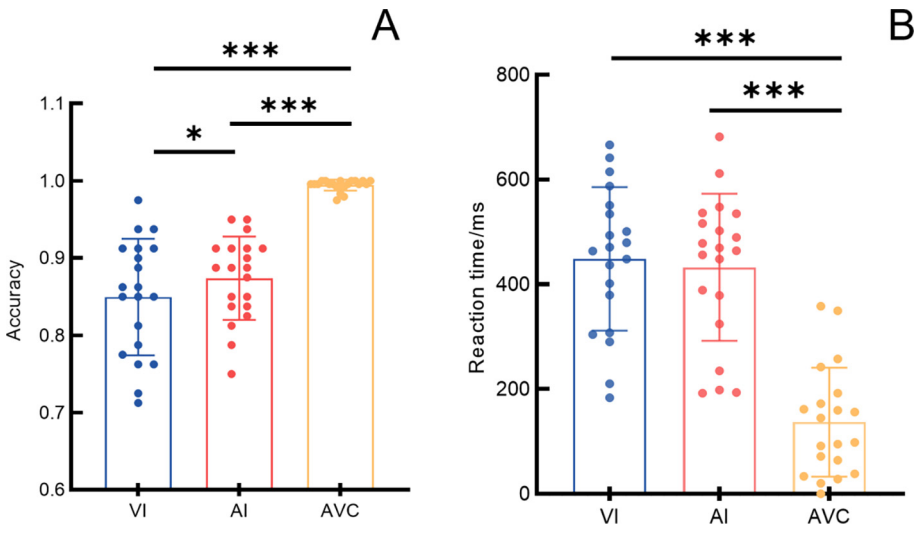

Fig. 3. Plot of mean RT and ACC in the three experimental conditions. (**A**) Accuracy for each condition. (**B**) Reaction times for each condition (*$p < 0.05$, ***$p < 0.001$, Bonferroni-corrected). Error bars represent the standard deviation of the mean.

where *Cpv* denotes the Cauchy principal value. Afterwards, the corresponding analytical signal phases $\phi_a(t)$ and $\phi_b(t)$ can be computed as.

$$\begin{cases} \phi_a(t) = \arctan\frac{H_a(t)}{a(t)} \\ \phi_b(t) = \arctan\frac{H_b(t)}{b(t)} \end{cases} \quad (3)$$

Finally, the PLV between time series *a* and *b* is formulated as.

$$Plv = \left| \frac{1}{N} \sum_{j=0}^{N-1} e^{j(\phi_a(j\Delta t) - \phi_b(j\Delta t))} \right| \quad (4)$$

where *Plv* is the connection weight estimated by PLV, Δt is the sampling period, and *N* denotes the sample number. By computing the PLV for each pair of paired electrodes, the network was constructed for the three conditions.

Statistical analysis

We used repeated measures one-way analyses of variance (ANOVA) to quantify the differences in ACC, response time (RT), and brain networks among the three conditions. Post hoc tests (pairwise comparisons, corrected by using Bonferroni-corrected) were further applied to reveal the difference between pairs of conditions. Repeated measure two-way ANOVA with conditions (3 levels) and electrodes (8 levels) as within-participants variables was performed to reveal the N400 amplitude difference among the three stimuli conditions.

RESULTS

Behavioural differences

As illustrated in Fig. 3, ACC and RT were compared in the three conditions (i.e., AI: the normal visual and abnormal auditory condition; AVC: the normal visual and normal auditory condition; VI: the normal auditory and abnormal visual condition). Repeated measures one-way ANOVA revealed significant differences in RT ($F_{(2,38)} = 102.7$, $p < 0.001$) among the three conditions, and the post hoc test further revealed that the AVC condition (136.82 ± 103.95 ms) had the shortest RT compared with the RTs of AI (432.41 ± 140.32 ms) and VI (448.50 ± 136.83 ms) conditions. In terms of ACC, the ANOVA identified significant differences among the three conditions ($F_{(2,38)} = 69.759$, $p < 0.001$). The post hoc tests revealed that the AVC (99.44 ± 0.71%) had the highest ACC compared with the ACC of the AI (87.37 ± 5.40 %) and the VI (84.94 ± 7.5%) conditions. Furthermore, the AI condition showed a higher ACC than that in the VI condition, as depicted in Fig. 3.

ERPs analysis

Statistical analyses of the N400 amplitudes revealed significant differences among experiment conditions ($F_{(2,38)} = 28.38$, $p < 0.001$; Fig. 4). Pairwise comparisons with a Bonferroni adjustment revealed that the N400 component was enhanced in the AI (−1.16 ± 0.12 μV) condition compared with that elicited in either the AVC (−0.001 ± 0.18 μV, $p < 0.001$) or the VI condition (−0.15 ± 0.18 μV, $p < 0.001$).

Patterns of the brain network

We statistically analysed the connection patterns among different conditions by performing the following steps: (1) Repeated measures one-way ANOVA was utilized to check whether there was a significant difference in the whole-brain connection strength under the three conditions. (2) The post hoc test further revealed significant linkages in three comparison conditions (i.e., AVC vs AI, AI vs VI, and AVC vs VI). Fig. 5A depicts the linkages with significant differences among the three conditions for all participants revealed by repeated measures one-way ANOVA. The post hoc test results showed that the AVC condition had a stronger connection pattern than that in the VI or AI conditions; moreover, the AI condition had a stronger network connection than that in the VI condition, which was mainly an increased long-range functional connectivity pattern (Fig. 5D). The network linkages of the AVC condition were significantly more potent than those of the AI condition, and the significant differences were mainly in the left frontal and occipital areas (Fig. 5B). Likewise, the significant differences found in the AVC condition had a stronger network connection pattern with long-range functional connectivity located in the occipital and frontal areas compared to that in the VI condition (Fig. 5C).

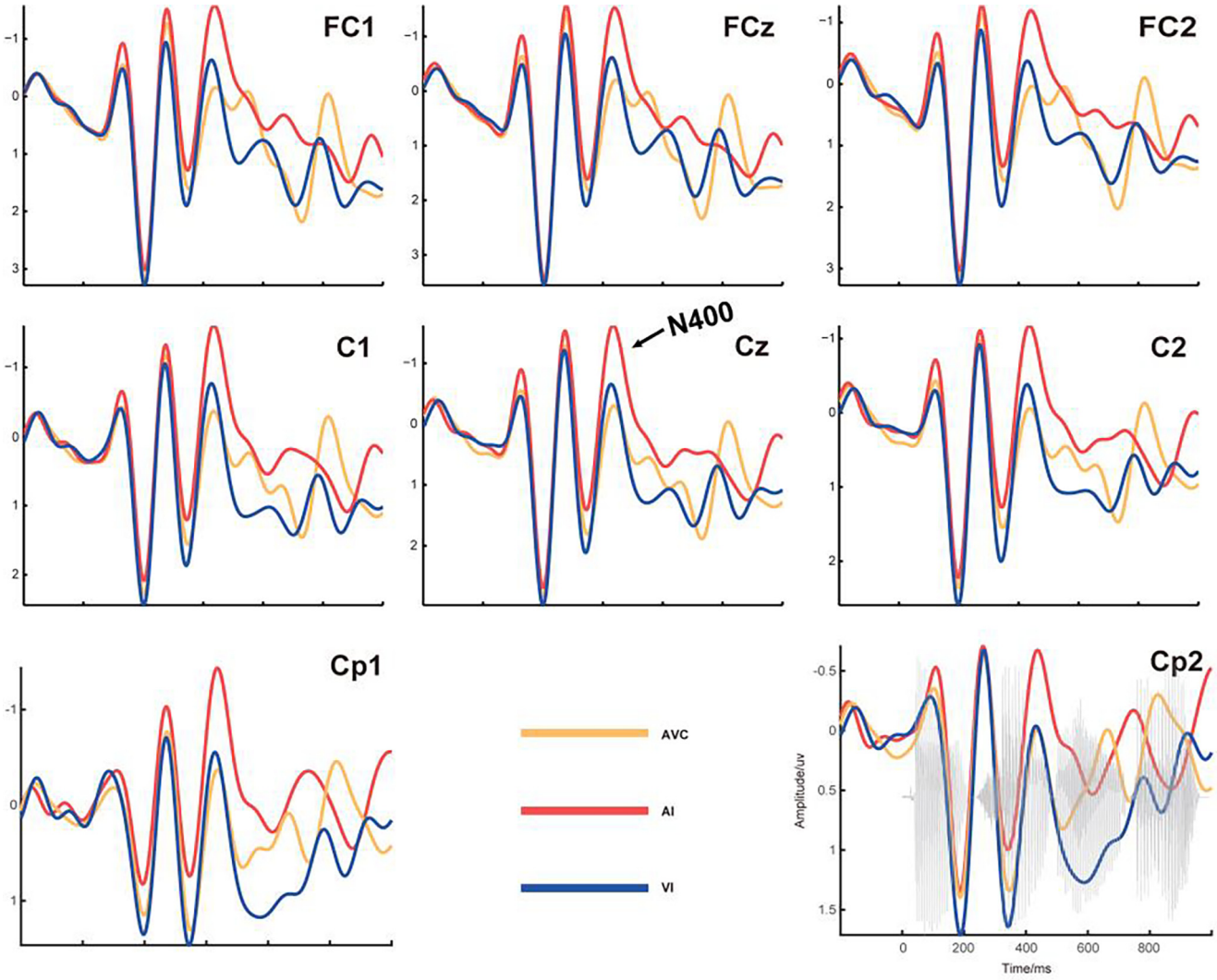

Fig. 4. ERP waveforms for the experimental conditions at eight electrodes (FC1, FCz, FC2, C1, Cz, C2, Cp1 and Cp2).

DISCUSSION

What modality dominates when the participant receives conflicted phonemic and graphemic information? In the current study, we addressed the questions in the experiment combining auditory speech and visual characters presentations in different stimuli conditions, aiming to probe how auditory speech and visual characters conflicting information is represented in the brain. To this end, we collected behavioural performance and electrophysiological evidence, including ERPs and functional brain networks to identify audio-visual competition mechanisms. We demonstrated the existence of audio-visual competition between speech and Chinese characters. Furthermore, auditory dominance prevailed when the brain processed conflicting auditory speech and visual characters information.

In terms of behavioural performance, participants responded more accurately and faster to consistent audio-visual stimuli (AVC condition) than they did in either of the unpredictable audio-visual conditions (VI and AI conditions). The results suggested that visual characters and auditory speech competition exists (Robinson and Sloutsky, 2013; Sugano et al., 2016). In addition, since audio-visual competition occurred, the brain recognized meaningless speech sounds more accurately than it detected meaningless Chinese sentences. Although there was no significant difference in RT when the participants determined that they had received an auditory or visual mismatch stimulus, they spent relatively little time identifying auditory mismatch stimuli. In general, these results suggested that the participants were better at identifying speech mismatch conditions. Moreover, we explored electrophysiological evidence of how auditory speech and visual characters stimuli compete.

Likewise, the N400 effect in both inconsistent (VI and AI) and consistent (AVC) audio-visual conditions was in agreement with those of previous studies (Kutas and Hillyard, 1980; Kutas et al., 1983; Hu et al., 2012). More specifically, the N400 amplitude was larger in the inconsistent condition than in the consistent condition. In com-

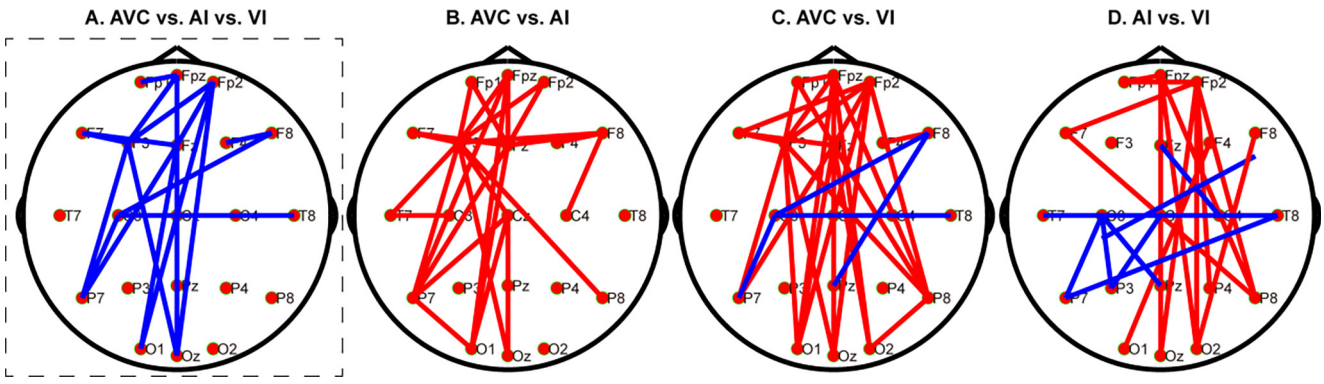

Fig. 5. Differentiated brain network among the three conditions and the post hoc tests. In the subfigures, the red and blue solid lines denote significantly enhanced and weakened functional connectivity, respectively ($p < 0.01$).

parison, the N400 amplitude was larger in the auditory mismatch condition than in the visual mismatch condition. Importantly, semantic integration becomes difficult when visual characters and auditory speech are inconsistent. In contrast to speech (auditory modality), reading is a cognitive skill that has only emerged over the last a few thousand years. Until now, it has been generally assumed that reading acquisition was evolutionarily developed, benefiting from anchored spoken systems (Karipidis et al., 2018). In previous studies, Hu and colleagues (Hu et al., 2012) instructed participants to identify whether the last character in four-character idioms is correct, on the one hand, they found tone violation elicited a more robust late positive component than vowel violation, suggesting different reanalyses of the two types of information; moreover, the vowel mismatch elicited an earlier negative component and a larger N400 effect than the tone mismatch, In fact, the N400 amplitude induced by meaningless speech sounds was larger than that of meaningless sentences as individuals are more sensitive to the auditory modality. In our present study, the N400 was accordingly considered and the corresponding results found that the auditory mismatch condition was easier to identify when the brain processes conflicting audio-visual information. The N400 amplitude induced by meaningless speech sounds was larger than that of meaningless sentences because individuals are more sensitive to auditory mismatch and more capable of identifying semantic abnormalities of auditory modality. Our results found that the auditory mismatch condition was easier to identify when the brain processes conflicting audio-visual information. In fact, several studies have investigated ERP effects of combined syntactic category and semantic anomalies in Chinese, a non-Indo-European language that does not use grammatical morphology to mark syntactic category or syntactic features (Ye et al., 2006; Zhang et al., 2010). As illustrated, syntactic processing is not a necessary prerequisite for the initiation of semantic integration in Chinese (Zhang et al., 2013; Yu et al., 2015). For example, Zhang et al. (2010) observed an N400 effect for the combined anomalies both in the SVO (subject–verb–object) and in the *ba* (把) sentences, suggesting that semantic integration proceeds even when syntactic category processing fails. Following these previous studies, in current work we mainly focused more on N400, although both semantic and syntactic abnormalities do exist in both auditory and visual incongruent conditions.

In addition, our study found a difference in network patterns among the three experimental conditions. When the visual characters and auditory speech are consistent, the information simultaneously integrates and recruits more relevant resources to handle language-related information. In particular, compared to those in the speech mismatch condition, the linkage patterns for the consistent audio-visual condition were characterized by stronger long-range connections between the left prefrontal and temporal regions, and the two brain regions are physiologically connected via fibre bundles (Catani and Thiebaut de Schotten, 2008). Moreover, this connection between the left prefrontal and left temporal areas supports the efficient processing of syntactically complex sentences (Brauer et al., 2011; Wilson et al., 2011). Namely, the left temporal areas are involved in mapping sensory or phonological representations to lexical conceptual representations, and the left frontal regions are responsible for storing and integrating linguistic information (Hickok and Poeppel, 2007). Our findings further showed strong left lateralization around both frontotemporal regions for consistent audio-visual conditions, which was congruous with previous neuroanatomical language processing models (Hickok and Poeppel, 2007; Friederici, 2011). Compared with the visual mismatch condition, the long-range connection between the frontal and occipital areas was the primary brain connection pattern for the consistent audio-visual condition. Physiologically, the occipital regions are more involved in processing visual information than other regions (Dehaene et al., 2005; Goodale, 2008), and a previous study also showed that the frontal cortex is a convergence region that might help integrate multimodal sensory information with language processing (Hagoort, 2005).

More strikingly, compared with that in the visual mismatch condition, the dominant pattern in the auditory mismatch condition showed enhanced linkages in the

posterior occipital-parietal areas. Notably, the prefrontal areas are known to resolve cross-modal conflicts (Mayer et al., 2009; Orr and Weissman, 2009) and play a critical role during decision-making (Paret et al., 2016). When the brain deals with conflicting audio-visual information, especially when a participant encounters uncertain situations, the brain allocates more resources to the decision-making process (deciding what type of stimulus it is). Under the visual modality, the pronunciation of words is activated first by visible glyphs and then there is a transition from the pronunciation of words to the meaning. Therefore, the visual modality may be less efficient than the auditory modality, which allows the language information to be processed directly from pronunciation to meaning (Coltheart et al., 2001). These results consistently prove auditory dominance in the audio-visual competition of language information.

In this paper, we have discussed the auditory dominance effect in the competition between auditory speech and visual characters. Though we have strived to make our analyses as comprehensive as possible, there may be limitations that need to be addressed in future work. The networks were constructed at the scalp level, which may have been influenced by the volume conduction effect (Schoffelen and Gross, 2009). Although we used the sparse 21 canonical electrodes of the 10–20 EEG system to lower the effect of volume conduction, the technique of EEG source connectivity which constructs networks at the cortical level is another important solution (O'Neill et al., 2017; Kabbara et al., 2018). In future work, we will construct EEG networks at the source level to probe how audio-visual conflict information is represented in the brain. The experimental design of our present study adopted a variation of the "Ding et al." work, where monosyllabic words are presented at a fixed rate (4/s) without any prosody. Ding and collegaues indicated that a hierarchy of neural processing timescales underlies grammar-based internal construction of the hierarchical linguistic structure (Ding et al., 2016). In our experimental paradigm, visual stimuli were presented to the participants along with the speech. In fact, in the study performed by Tatiana et al., a similar experimental design was used, as participants can also see visual text while receiving an auditory stimulus under one of their concerned conditions. Although this design retains more visual information and may cause a stronger visual advantage, Tatiana et al. did report seldom effects on individual N400 (Tatiana et al., 2016). While concerning this issue, new experimental protocols should be further considered in the future, for example, a next step is to simultaneously present the Chinese characters and speech in time and space to further explore the mechanism of audio-visual competition.

To conclude, the current study revealed the auditory dominance effect in the competition between auditory speech and visual characters. The higher ACC, concurrent larger N400 amplitudes and more linkages in the posterior occipital-parietal areas in the auditory mismatch condition compared to the outcomes in the visual mismatch condition suggested that auditory dominance was represented when audio-visual competition occurred. These results consistently support the supremacy of the auditory modality when subjects received conflicting auditory speech or visual characters.

ACKNOWLEDGMENTS

We wish to thank all the participants who participated in our experiment. This work was supported by the National Natural Science Foundation of China (#U19A2082, #61961160705, #62103085, #82102175), the Science and Technology Development Fund, Macau SAR (File no. 0045/2019/AFJ), and the Project of Science and Technology Department of Sichuan Province (#2021YFSY0040, #2020ZYD013).

CONFLICT OF INTEREST

All authors claim that there are no conflicts of interest.

REFERENCES

Andres AJD, Cardy JEO, Joanisse MF (2011) Congruency of auditory sounds and visual letters modulates mismatch negativity and P300 event-related potentials. Int J Psychophysiol 79:137–146.
Bassett DS, Sporns O (2017) Network neuroscience. Nat Neurosci 20:353–364.
Brauer J, Anwander A, Friederici AD (2011) Neuroanatomical prerequisites for language functions in the maturing brain. Cereb Cortex 21:459–466.
Catani M, Thiebaut de Schotten M (2008) A diffusion tensor imaging tractography atlas for virtual in vivo dissections. Cortex 44:1105–1132.
Christoffels I, Timmer K, Ganushchak L, La Heij W (2016) On the production of interlingual homophones: delayed naming and increased N400. Lang Cogn Neurosci 31:628–638.
Coltheart M, Rastle K, Perry C, Langdon R, Ziegler J (2001) DRC: A dual route cascaded model of visual word recognition and reading aloud. Psychol Rev 108:204–256.
Cummings A, Ceponiene R, Dick F, Saygin AP, Townsend J (2008) A developmental ERP study of verbal and non-verbal semantic processing. Brain Res 1208:137–149.
De Gelder B, Bertelson P (2003) Multisensory integration, perception and ecological validity. Trends Cogn Sci 7:460–467.
Dehaene S, Cohen L, Sigman M, Vinckier F (2005) The neural code for written words: a proposal. Trends Cogn Sci 9:335–341.
Ding N, Melloni L, Zhang H, Tian X, Poeppel D (2016) Cortical tracking of hierarchical linguistic structures in connected speech. Nat Neurosci 19:158–164.
Dong L, Li FL, Liu Q, Wen X, Lai YX, Xu P, Yao DZ (2017) MATLAB toolboxes for reference electrode standardization technique (REST) of scalp EEG. Front Neurosci:11.
Friederici AD (2011) The brain basis of language processing: From structure to function. Physiol Rev 91:1357–1392.
Goodale M (2008) Action without perception in human vision. Cogn Neuropsychol 25:891–919.
Hagoort P (2005) On Broca, brain, and binding: A new framework. Trends Cogn Sci 9:416–423.
Hagoort P (2019) The neurobiology of language beyond single-word processing. Science 366. 55–+.
Hickok G, Poeppel D (2007) Opinion – The cortical organization of speech processing. Nature Rev Neurosci 8:393–402.
Hirst RJ, McGovern DP, Setti A, Shams L, Newell FN (2020) What you see is what you hear: Twenty years of research using the sound-induced flash illusion. Neurosci Biobehav Rev 118:759–774.
Houston, S.D., 2004. The first writing: The first writing.
Hu JH, Gao S, Ma WY, Yao DZ (2012) Dissociation of tone and vowel processing in Mandarin idioms. Psychophysiology 49:1179–1190.
Huang S, Li Y, Zhang W, Zhang B, Liu XX, Mo L, Chen Q (2015) Multisensory competition is modulated by sensory pathway

interactions with fronto-sensorimotor and default-mode network regions. J Neurosci 35:9064–9077.
Kabbara A, Eid H, Falou WE, Khalil M, Hassan M (2018) Reduced integration and improved segregation of functional brain networks in Alzheimer's disease. J Neural Eng 15:026023.
Karipidis II, Pleisch G, Brandeis D, Roth A, Rothlisberger M, Schneebeli M, Walitza S, Brem S (2018) Simulating reading acquisition: The link between reading outcome and multimodal brain signatures of letter-speech sound learning in prereaders. Sci Rep 8.
Keitel A, Gross J, Kayser C (2020) Shared and modality-specific brain regions that mediate auditory and visual word comprehension. Elife 9.
Kutas M, Federmeier KD (2000) Electrophysiology reveals semantic memory use in language comprehension. Trends Cogn Sci 4:463–470.
Kutas M, Federmeier KD (2011) Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). Annu Rev Psychol 62:621–647.
Kutas M, Hillyard SA (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207:203–205.
Kutas M, Hillyard SA (1983) Event-related potentials to grammatical errors and semantic anomalies. Mem Cogn 11:539–550.
Liu B, Wang Z, Wu G, Meng X (2011) Cognitive integration of asynchronous natural or non-natural auditory and visual information in videos of real-world events: An event-related potential study. Neuroscience 180:181–190.
Luck SJ (2014) An introduction to the event-related potential technique. Cambridge, MA: MIT Press.
Lukas S, Philipp AM, Koch I (2014) Crossmodal attention switching: Auditory dominance in temporal discrimination tasks. Acta Psychol (Amst) 153:139–146.
Manfredi M, Cohn N, Andreoli MD, Boggio PS (2018) Listening beyond seeing: Event-related potentials to audiovisual processing in visual narrative. Brain Lang 185:1–8.
Mayer AR, Franco AR, Canive J, Harrington DL (2009) The effects of stimulus modality and frequency of stimulus presentation on cross-modal distraction. Cereb Cortex 19:993–1007.
McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264:746–748.
Mishra J, Martinez A, Sejnowski TJ, Hillyard SA (2007) Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. J Neurosci 27:4120–4131.
O'Neill GC, Tewarie PK, Colclough GL, Gascoyne LE, Hunt BAE, Morris PG, Woolrich MW, Brookes MJ (2017) Measurement of dynamic task related functional networks using MEG. Neuroimage 146:667–678.
Orr JM, Weissman DH (2009) Anterior cingulate cortex makes 2 contributions to minimizing distraction. Cereb Cortex 19:703–711.
Paret C, Ruf M, Gerchen MF, Kluetsch R, Demirakca T, Jungkunz M, Bertsch K, Schmahl C, Ende G (2016) fMRI neurofeedback of amygdala response to aversive stimuli enhances prefrontal-limbic brain connectivity. Neuroimage 125:182–188.
Peelle JE, Sommers MS (2015) Prediction and constraint in audiovisual speech perception. Cortex 68:169–181.
Petersen SE, Sporns O (2015) Brain networks and cognitive architectures. Neuron 88:207–219.
Proverbio AM, Calbi M, Manfredi M, Zani A (2014) Comprehending body language and mimics: An ERP and neuroimaging study on Italian actors and viewers. PLoS One 9.
Raij T, Uutela K, Hari R (2000) Audio-visual integration of letters in the human brain. Neuron 28:617–625.
Robinson CW, Sloutsky VM (2010) Development of cross-modal processing. Wires Cogn Sci 1:135–141.
Robinson CW, Sloutsky VM (2013) When audition dominates vision evidence from cross-modal statistical learning. Exp Psychol 60:113–121.
Sakkalis V (2011) Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. Comput Biol Med 41:1110–1117.
Schoffelen JM, Gross J (2009) Source connectivity analysis with MEG and EEG. Hum Brain Mapp 30:1857–1865.
Shams L, Kamitani Y, Shimojo S (2000) Illusions – What you see is what you hear. Nature 408:788.
Si YJ, Wu X, Li FL, Zhang LY, Duan KY, Li PY, Song LM, Jiang YL, Zhang T, Zhang YS, Chen J, Gao S, Biswal B, Yao DZ, Xu P (2019) Different decision-making responses occupy different brain networks for information processing: A study based on EEG and TMS. Cereb Cortex 29:4119–4129.
Srinivasan R, Nunez PL, Silberstein RB (1998) Spatial filtering and neocortical dynamics: Estimates of EEG coherence. IEEE Trans Biomed Eng 45:814–826.
Steinhauer K, Alter K, Friederici AD (1999) Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nat Neurosci 2:191–196.
Sugano Y, Keetels M, Vroomen J (2016) Auditory dominance in motor-sensory temporal recalibration. Exp Brain Res 234:1249–1262.
Sun JF, Li ZJ, Tong SB (2012) Inferring functional neural connectivity with phase synchronization analysis: A review of methodology. Comput Math Method Med 13.
Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG (2010) The multifaceted interplay between attention and multisensory integration. Trends Cogn Sci 14:400–410.
Tatiana GF, Esther R, Stefan S, Bleichner MG (2016) The N400 effect during speaker-switch—Towards a conversational approach of measuring neural correlates of language. Front Psychol 7.
van Atteveldt N, Formisano E, Goebel R, Blomert L (2004) Integration of letters and speech sounds in the human brain. Neuron 43:271–282.
Weihing J, Daniels S, Musiek FE (2009) The effect of visual and audiovisual competition on the auditory N1–P2 evoked potential. J Am Acad Audiol 20:569–581.
Wilson SM, Galantucci S, Tartaglia MC, Rising K, Patterson DK, Henry ML, Ogar JM, DeLeon J, Miller BL, Gorno-Tempini ML (2011) Syntactic processing depends on dorsal language tracts. Neuron 72:397–403.
Wu YJ, Thierry G (2010) Chinese-English bilinguals reading English hear Chinese. J Neurosci 30:7646–7651.
Xu P, Xiong XC, Xue Q, Li PY, Zhang R, Wang ZY, Valdes-Sosa PA, Wang YP, Yao DZ (2014) Differentiating between psychogenic nonepileptic seizures and epilepsy based on common spatial pattern of weighted EEG resting networks. IEEE Trans Biomed Eng 61:1747–1755.
Yao DZ (2001) A method to standardize a reference of scalp EEG recordings to a point at infinity. Physiol Meas 22:693–711.
Ye Z, Luo YJ, Friederici AD, Zhou XL (2006) Semantic and syntactic processing in Chinese sentence comprehension: Evidence from event-related potentials. Brain Res 1071:186–196.
Yi CL, Yao RW, Song LY, Jiang L, Si YJ, Li PY, Li FL, Yao DZ, Zhang Y, Xu P (2021) A novel method for constructing EEG large-scale cortical dynamical functional network connectivity (dFNC): WTCS. IEEE T Cybern.
Yu J, Zhang Y, Boland JE, Cai L (2015) The interplay between referential processing and local syntactic/semantic processing: ERPs to written Chinese discourse. Brain Res 1597:139–158.
Zhang Y, Yu J, Boland JE (2010) Semantics does not need a processing license from syntax in reading Chinese. J Exp Psychol Learn Mem Cogn 36:765–781.
Zhang Y, Ping L, Piao Q, Liu Y, Hua S (2013) Syntax does not necessarily precede semantics in sentence processing: ERP evidence from Chinese. Brain Lang 126:8–19.

(Received 24 May 2022, Accepted 22 August 2022)
(Available online 27 August 2022)