# Revealing differential importance of word categories in spoken sentence comprehension using phoneme-related representation

Trang-Le Thi[1], Youngmin Na[1], Inyong Choi[2], Jihwan Woo[1,3,*]

[1] *Department of Electrical, Electronic and Computer Engineering, University of Ulsan, 44610 Ulsan, Republic of Korea*

[2] *Department of Communication Sciences and Disorders, University of Iowa, Iowa City, IA 52242, USA*

[3] *Department of Biomedical Engineering, University of Ulsan, 44610 Ulsan, Republic of Korea*

*Correspondence: jhwoo@ulsan.ac.kr (Jihwan Woo)

*Background*: Verbal communication comprises the retrieval of semantic and syntactic information elicited by various kinds of words (i.e., parts of speech) in a sentence. Content words, such as nouns and verbs, convey essential information about the overall meaning (semantics) of a sentence, whereas function words, such as prepositions and pronouns, carry less meaning and support the syntax of the sentence. *Methods*: This study aimed to identify neural correlates of the differential information retrieval processes for several parts of speech (i.e., content and function words, nouns and verbs, and objects and subjects) via electroencephalography performed during English spoken-sentence comprehension in thirteen participants with normal hearing. Recently, phoneme-related information has become a potential acoustic feature to investigate human speech processing. Therefore, in this study, we examined the importance of various parts of speech over sentence processing using information about the onset time of phonemes. *Results*: The distinction in the strength of cortical responses in language-related brain regions provides the neurological evidence that content words, nouns, and objects are dominant compared to function words, verbs, and subjects in spoken sentences, respectively. *Conclusions*: The findings of this study may provide insights into the different contributions of certain types of words over others to the overall process of sentence understanding.

*Keywords*

Content words; Function words; Phoneme-related EEG; Sentence comprehension

## 1. Introduction

Sentence processing is crucial for human verbal communication. A sentence is comprised of words that can each be characterized as different parts of speech. "Parts of speech" is an abstract term for the major classes of words in a language. Words are grammatically categorized into two major classes: content words and function words (open and closed classes words, respectively). The content words class includes nouns, verbs, adjectives, and adverbs, whereas pronouns, noun adjuncts, verb adjuncts, and conjunctions are classified as function words. While parts of speech classes are grammatically distinguished, their effect on sentence processing in the human brain is still unclear. Understanding how the brain processes parts of speech may help facilitate human communication, especially in individuals with language disorders. Thus, this study aimed to provide neurophysiological evidence of the contribution of different classes of parts of speech in sentence comprehension.

Function words primarily provide grammatical continuity within a sentence. Although some function words categories can also hold meaning (e.g., pronouns and noun adjuncts), they normally act as connecting words [1]. Unlike function words, content words have lexical meanings that need to be processed and related to the sentence context, thereby affecting sentence comprehension efficiency [2]. Additionally, Cutler and Foss [3] found that the reaction time to word-initial phoneme targets was shorter for high-stress words than low-stress words and related to sentence processing difficulty. Given that high-stress words are considered easier to process than low-stress words, content words typically receive higher stress than function words [2, 3]. Moreover, the information content contained within a word is relative to the amount of stress placed upon it [3]; thus, it is reasonable to assume that content words play an important role in sentence processing by conveying irreplaceable critical meaning. Therefore, we hypothesized that content words are more dominant than function words in extracting the general meaning of the sentence. We aim to confirm this hypothesis by examining electrophysiological activation regarding each class of processing. In content words, nouns and verbs convey different kinds of information as they refer to different concepts [4]. Furthermore, objects and subjects of a verb action provide another critical aspect of sentence processing [5, 6], such that each class plays a distinct role in sentence comprehension.

Speech processing entails numerous steps that occur over short time intervals (a few milliseconds) as the speech signal unfolds [7, 8]. As electroencephalography (EEG) measures brain activity with an excellent temporal resolution, it has been widely used to investigate language and speech processing [8, 9]. Specifically, event-related potentials (ERPs) can reveal variable cognitive processes in time-locked temporal patterns during speech processing. In particular, sentence processing consists of many stages across multiple time frames, grouped into three phases occurring between 100–300 ms, 300–500 ms, and 500–1000 ms, respectively. These phases correlate with a series of peaks and troughs identified chronologically in ERP, as N1 (a negativity peak at 100 ms after stimulus onset), early left-anterior negativity, left-anterior negativity N400, and P600, reflecting brain activities responsible for tackling different sentence processing levels [10, 11]. It has been proposed that sentence processing begins by identifying acoustic-phonological events and word forms during the first phase (100–300 ms), reflected in the activation of the superior temporal gyrus and inferior frontal gyrus. The superior temporal gyrus is found to activate during speech perception of phonemes and semantic processing tasks at sentence-level processing, while the inferior frontal gyrus has been known to support syntactic structure building and verbal working memory [11]. Moreover, previous studies have shown that words with different meanings could be dissociated from the brain activity occurring 100–200 ms after the word onset in both auditory and visual paradigms with word perception tasks [12, 13]. It has been suggested that semantic processing may occur as early as within 200 ms after the stimulus word onset [12–14]. Considering this, the difference in brain activity at the N1 latency could indicate the difference in word-level processing, thereby affecting sentence comprehension. Therefore, we assumed that the distortion in scalp topography during the processing of each word class of words in sentences at the N1 latency (100–300 ms) indicates the different importance levels of that word class with respect to sentence comprehension.

However, the ERP technique involves an averaging process over several repetitions, which leads to an overlapping response to the stimuli exceeding several hundred milliseconds. Thus, conventional ERP techniques are limited to the analysis of the neural response during a continuous natural speech [15] and are typically used for isolated sensory events (e.g., isolated syllables [12, 13]). However, recent advances in EEG analysis help avoid the overlapping responses to continuous natural speech (e.g., sentences) by computing the cross-correlation between the temporal envelope of speech and corresponding EEG signals. A cross-correlation function reflects the similarity between two signals assuming that both signals have a linear relationship. Benefiting from a simple calculation, by computing sliding dot product and the assumption that the human brain act as a linear system [16], the cross-correlation approach has been widely used to investigate the phase-locked neural response to con-

tinuous speech stimuli [17, 18]. Horton *et al.* [18] investigated the neural entrainment to the envelope of attended and unattended speech using the cross-correlation between the EEG and speech envelope. Their study shows the expected range of cross-correlation values to determine if the neural responses were truly entrained to the speech stimuli, thereby confirming that they did not occur by chance; it is outside the range of between –0.0035–0.0035 [18]. According to their result, the measured EEG significantly correlated with both attended and unattended speech envelopes at some latency (e.g., 100 ms and 200 ms) after stimulus onset. Hence, cross-correlation can be used to indicate the entrainment of neural oscillation to speech.

The temporal envelope of speech has been widely used as an acoustical information cue [18, 19]. However, Di Liberto, O'Sullivan, and Lalor [20] showed that the phonemic model outperformed the envelope model in predicting neural responses to speech stimuli. Furthermore, phoneme information has been posited to reveal brain activation during speech perception tasks [21–23]. We introduce a term, phoneme-onset impulse train (PH), to indicate an impulse sequence of the onset time of phonemes in a sentence. A similar concept was used by Di Liberto *et al.* [24] to investigate the cortical encoding of melodic information in a music-related study, in which they used note-onset information. To the best of our knowledge, no previous language study has utilized such a feature to analyze speech stimuli. The combination of the PH and temporal envelope of speech was also assessed to validate the usefulness of PH information.

In this study, we used continuous speech sentences as stimuli. However, we selectively omitted typical linguistic components by using indexed PHs instead of particular instances of parts of speech as stimuli. We utilized a cross-correlation-based approach to parse early ERP components into different isolated linguistic events within the context of sentence processing. This approach ensures that listeners process a meaningful sentence rather than sole speech tokens and facilitates the examination of the importance of word categories in sentence comprehension. Passive listening was employed rather than active listening as this paradigm reduces participant fatigue during the experiment. Furthermore, Kong *et al.* [25] claimed that active and passive listening resulted in similar neural responses to the speech features in quiet conditions with book chapters as stimuli. The difference in the contribution of various classes of words in sentence comprehension was evaluated by comparing the neural response relative to each word category with that of the whole sentence stimuli.

## 2. Material and methods

### 2.1 Participants

Thirteen right-handed participants with normal hearing between 18–24 years of age (mean: 21.5 years, standard deviation: 2.2, four males and nine females) participated in the experiment. All participants were native speakers of Ameri-

can English and had normal hearing thresholds (i.e., thresholds ≤20 dB HL at all tested frequencies from 250–8000 Hz). None of the participants had neurological conditions or disorders, and none of them were taking any medications. All study procedures were approved by the Institutional Review Board of the University of Iowa.

## 2.2 Stimuli

Ten sentences from the Revised Speech Perception in Noise Test [26] (duration mean: 1.78 s, standard deviation: 0.19) spoken in English by an American male were used. Most of the sentence stimuli were simple declarative sentences in the active voice to minimize variability across participants in listening effort. Moreover, the sentences were made by the approximation of daily conversation and phonetic balance [26], making them suitable for speech comprehension investigation. A list of sentence stimuli is presented in Table 1.

**Table 1. List of sentence stimuli.**

| | Sentences | Duration (s) |
|---|---|---|
| 1 | Maple syrup is made from sap | 1.85 |
| 2 | Paul was interested in the sap | 1.60 |
| 3 | Bill heard Tom called about the coach | 1.99 |
| 4 | The team was trained by their coach | 2.01 |
| 5 | Our cat is good at catching mice | 2.00 |
| 6 | Bob should not consider the mice | 1.87 |
| 7 | Let's invite the whole gang | 1.67 |
| 8 | You were considering the gang | 1.45 |
| 9 | She wants to speak about the ant | 1.74 |
| 10 | A termite looks like an ant | 1.61 |

## 2.3 Experimental setup and procedures

The experiment was conducted in a soundproof booth. Participants were seated facing a computer monitor with a loudspeaker placed in front of them at a distance of 1 m. They were allowed to watch a silent movie played on the monitor and were asked to minimize their movement. All the sentences were presented 100 times through a loudspeaker at 65 dB SPL, resulting in 1000 trials in random order. Each trial lasted for 3 s, from –0.5 s preceding stimulus presentation to 2.5 s following onset, leading to 3 s inter-stimulus interval between trials. The passive listening task required no other response from the participants.

## 2.4 EEG recording and processing

EEG signals were recorded using a 64-channel EEG system (BioSemi Co., Netherlands) at a 2048 Hz sampling rate. The raw EEG signals were down-sampled to 256 Hz and re-referenced using the average reference. The EEG data were then filtered using the 5th order Butterworth bandpass filter (0.5–57 Hz). The extended infomax independent component analysis (ICA) algorithm, which has been proven to successfully isolate eye blinks [27, 28], was implemented in the EEGLAB toolbox and applied to separate independent noise

components mixed in EEG signals. The noise component induced by eye movement was then rejected through visual inspection based on the topography, spectral content, and time-series activity. The ICA components left were then projected back into the channel space to develop the eye movement-free EEG data. The EEG signals were epoched into a 3 s window, from 0.5 s before stimulus onset to 2.5 s after stimulus onset. Epochs with a maximum amplitude exceeding 100 $\mu$V were excluded from the analysis. Then, the epochs were concatenated and filtered using the 5th order Butterworth bandpass (1–15 Hz) filter. The bandpass filter was chosen to minimize the filtering distortion effect on EEG signals as well as to preserve the phoneme-related potential peaks in neural responses to continuous speech [25, 29].

## 2.5 Phoneme-onset impulse train

All the phonemes in the sentences were listed and extracted their onset time using Praat software (University of Amsterdam, Netherlands) [30]. The sentences were then manually categorized into noun, verb, object, subject, content word, and function word. A PH is a unit impulse sequence of the onset time of phoneme regardless of consonants or vowels. In other words, a PH is a signal vector of zeros at the sampling rate of EEG signals, marked with a value of one at all phoneme onsets with the length of 2500 ms. The PH was computed as follows (Eqn. 1):

$$PH(t) = \begin{cases} 1, t : \text{ onset time of a phoneme} \\ 0, \text{ otherwise} \end{cases}$$

PH was obtained for the whole-sentence and component-exclusion cases (content words vs. function words exclusions, nouns vs. verbs exclusions, and objects vs. subjects exclusions). In component-exclusion cases, the phoneme-onset time information related to the components was omitted from the phoneme-onset train of the whole sentence. There were no significant differences in the number of phonemes over ten stimuli sentences between nouns exclusion and verbs exclusion ($p$ = 0.9414, W = 28.5, Wilcoxon signed-rank test), and between objects exclusion and subjects exclusion ($p$ = 1, W = 11, Wilcoxon signed-rank test). Fig. 1A,B illustrate an example of the PH of the sentence, "Maple syrup is made from sap" for a whole sentence and a nouns-exclusion case, respectively. Additionally, the combination of the phoneme-onset train and the temporal envelope of speech (PHENV) was examined to validate the potential of phoneme-onset time in speech perception. The PHENV was calculated by overlaying the remaining impulse train on the temporal envelope, as observed in Fig. 1C.

## 2.6 EEG analysis

EEG signals were averaged over all epochs from each sentence at each EEG electrode. The averaged EEG signals were baseline corrected by subtracting the average amplitude between –200 and 0 ms relative to stimulus onset. EEG signals in response to the sentence "Maple syrup is made from sap",
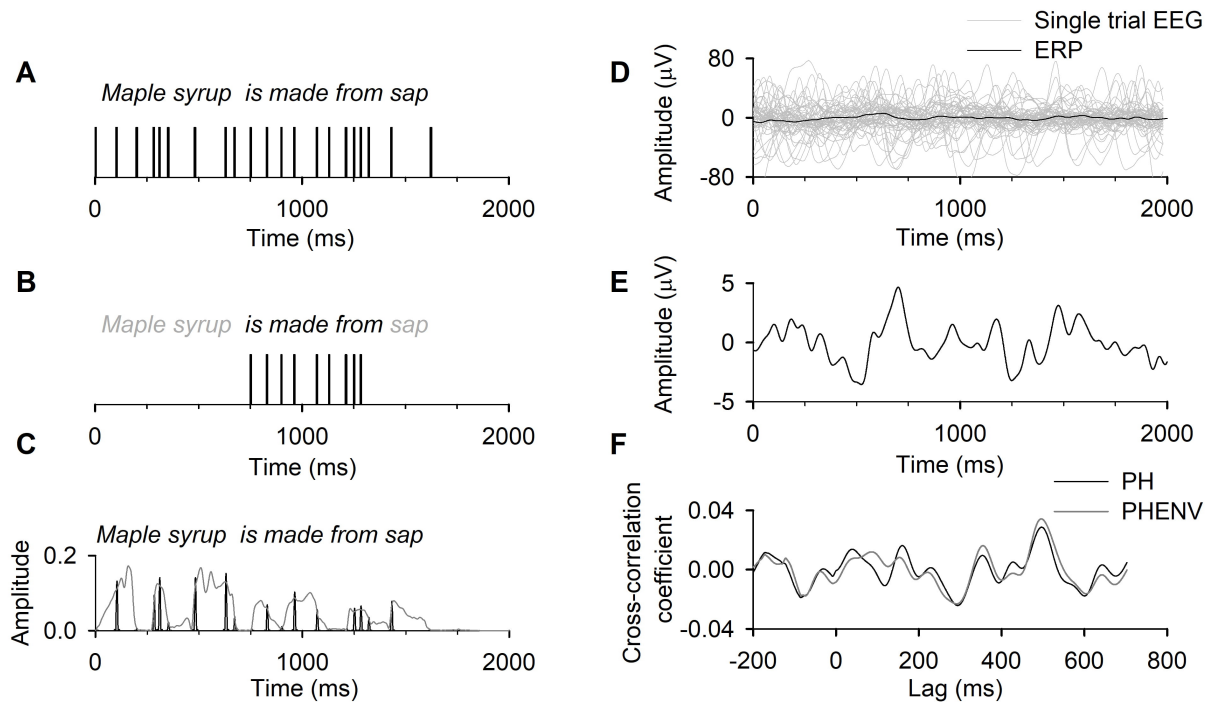
**Fig. 1. An example of a PH.** (A) for a whole sentence. (B) for a noun-exclusion case. (C) An example of a temporal envelope (coded in gray color) overlaying a PHENV (coded in black color) for a whole sentence. (D) Single trials EEG signals and the corresponding ERP. (E) The corresponding ERP in (D). (F) An example of cross-correlation coefficients between PH and ERP (coded in black color) and between PHENV and ERP (coded in gray color).

and the corresponding averaged EEG signals are shown in Fig. 1D. A prominent example of the averaged EEG signals (i.e., ERPs) in response to the continuous speech sentence of "Maple syrup is made from sap" is shown in Fig. 1E.

The averaged EEG signals were then trimmed to the range of 0 ms to 2500 ms to match the length of PHs. Cross-correlations were computed between the PHs and averaged EEG signals of each sentence and between the PHENV and averaged EEG signals of each sentence. The cross-correlations were trimmed as a function of lag and ranged from –200–700 ms. The positive lag indicates that the EEG signal lags the PH or PHENV from the onset of the PH or PHENV, respectively. The cross-correlations were averaged over all participants, denoted as the grand averaged cross-correlation. The grand averaged cross-correlations were baseline corrected by subtracting the mean and high pass (1 Hz) filtered. Examples of the grand averaged cross-correlation function corresponding to the PH and PHENV are shown in Fig. 1F.

All EEG-related potentials are shown at the anterior frontal site of AF3 electrode from subject 10 in response to a sentence of 'Maple syrup is made from sap'.

*2.7 Statistical analyses*

As the number of subjects in this study limits testing for normality of distribution [31], a non-parametric statistical test was employed. The differences in amplitude of grand averaged cross-correlations between the whole-sentence case and each component-exclusion case were evaluated using Wilcoxon signed-rank test with a two-sided hypothesis test. The exact probability distribution of W (i.e., the sum of the ranks of positive difference) was adopted to compute the $p$-value. The significance level was set at 0.05; this resulted in the critical value of W = 17 according to the Wilcoxon Signed-Ranks table with a two-sided test.

## 3. Results

Fig. 2 shows the grand average of cross-correlation coefficients between PHs and ERPs and between PHENVs and ERPs across all participants and sentences. As the grand average of cross-correlation coefficients shows a complex pattern of peaks and troughs comparable to the ERPs components in an auditory task (i.e., P1-N1-P2) [18], we focus on analyzing the shape of the cross-correlation function. We termed the first positive peak, first negative peak, and second positive peak in the cross-correlation function as P1, N1, and P2, respectively. The data are shown for the central electrode site of Cz. Panels A, B, and C in Fig. 2 illustrate data for content- vs. function words-exclusion cases, nouns- vs. verbs-exclusion cases, and objects- vs. subjects-exclusion cases in comparison with the whole sentence, respectively. The exclusion cases are represented by black or red lines in each panel, whereas the dark gray line shows the whole sentence. As shown in Fig. 2, changes with different levels in N1 amplitude were observed among component-exclusion cases compared to the whole sentence. In particular, with respect to cross-correlation coefficients between PH and ERPs, N1 amplitude in the content words-exclusion case was signifi-
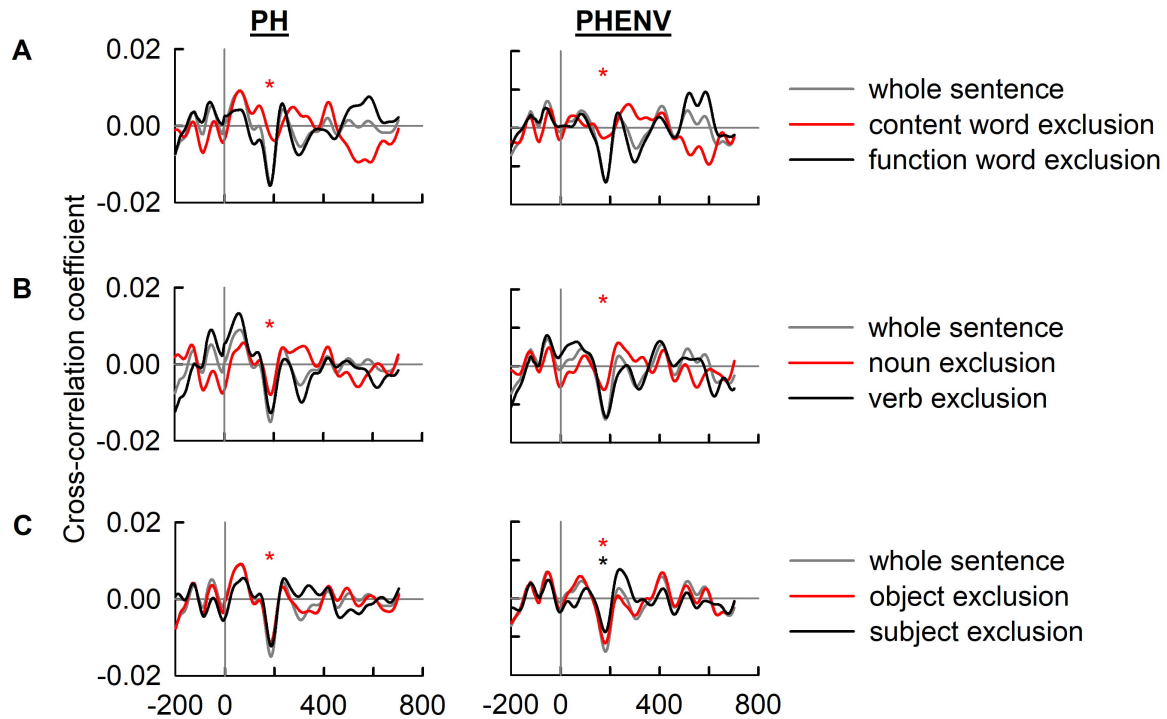
**Fig. 2. Example of cross-correlation functions between PH and ERPs and between PHENV and ERPs at the electrode Cz.** (A) Effects of content words vs. function words exclusion. (B) Effects of nouns vs. verbs exclusion. (C) Effects of objects vs. subjects exclusion. * indicates the statistical difference between cross-correlation of a component-exclusion case with the whole-sentence case at the level of $p < 0.05$, $W \leq 17$. The asterisk color corresponds to the significant difference induced by the corresponding component-exclusion case.

**Table 2. Group of electrodes at left hemisphere and central site showing significant reduction in N1 amplitude when comparing each component-exclusion case with whole-sentence case using PH (Wilcoxon signed-rank test, $p < 0.05$, $W \leq 17$).**

|  | Content word exclusion | Function word exclusion | Noun exclusion | Verb exclusion | Object exclusion | Subject exclusion |
|---|---|---|---|---|---|---|
| Number of electrodes | 4 | 0 | 12 | 0 | 7 | 6 |
| Left frontal | / | / | FC1, Fpz, AFz, FCz, Fz | / | FC1, Fz, FCz | FC1, Fpz, AFz |
| Central-parietal | C3, C5, FCz, Cz | / | C1, C3, Cpz, Cz, P1, P3, Pz | / | C1, CP1, Cz, P9 | C3, Cz, Cpz |

**Table 3. Group of electrodes at the left hemisphere and central site showing significant reduction in N1 amplitude when comparing each component-exclusion case with whole-sentence case using PHENV (Wilcoxon signed-rank test, $p < 0.05$, $W \leq 17$).**

|  | Content word exclusion | Function word exclusion | Noun exclusion | Verb exclusion | Object exclusion | Subject exclusion |
|---|---|---|---|---|---|---|
| Number of electrodes | 5 | 1 | 8 | 0 | 19 | 4 |
| Left frontal | / | / | Fpz, AFz, Fz | / | AF3, F1, F3, F5, FC3, FC1, AFz, Fz, FCz | Fpz, AFz |
| Central-parietal | C3, C5, Cz, P7 | P9 | C3, C5, CPz, Cz, P7 | / | C1, C3, CPz, Cz, CP3, CP1, P9, PO3 | P7 |
| Temporal | T7 | / | / | / | / | T7 |
| Occipital | / | / | / | / | O1, Oz | / |

cantly reduced at the Cz channel ($p = 0.046$, $W = 17$), whereas that in the function words-exclusion case revealed a subtle reduction (see Fig. 2A). N1 amplitude in the nouns-exclusion case decreased significantly at the Cz channel ($p = 0.033$, $W = 15$), while N1 amplitude in the verbs-exclusion case showed a non-significant reduction (as seen in Fig. 2B). Additionally, Fig. 2C showed a significant reduction in the object-exclusion case ($p = 0.003$, $W = 3$) but showed no significant reduction in

the case of subjects exclusion. Consistent findings were observed in the grand average of cross-correlation coefficients between PHENV and ERP (see Fig. 2, PHENV panel). Table 2 summarizes the electrode sites located in the left hemisphere and the central site that showed significant differences in N1 amplitude ($p < 0.05$, $W \leq 17$) between each exclusion case and the whole-sentence case when using PH. The object-exclusion case significantly reduced N1 activity at the greatest

number of electrode sites (i.e., 19 electrodes out of 37). However, no electrode site showed such a significant difference in the verb-exclusion case. The comparable details for that type of information using PHENV are shown in Table 3.

The cross-correlation scalp distribution of P1, N1, and P2 amplitudes using PH and PHENV are shown as topographies in panels A and C of Figs. 3,4,5. The time labels in these figures correspond to the latencies of P1, N1, and P2 peaks as 152, 199, and 285 ms for the PH, and 156, 199, and 293 ms for the PHENV case in the cross-correlation functions. Wilcoxon signed-rank tests were performed with respect to each electrode to test the effect of component exclusion (i.e., function words vs. content words exclusions, verbs vs. nouns exclusions, and subjects vs. objects exclusions). Figs. 3B,4B,5B illustrate brain regions that showed significant differences at P1, N1, and P2 amplitudes between the whole sentence and each exclusion case when using PH. The red color indicates that the whole-sentence case showed greater peak amplitude ($p < 0.05$, $W \leq 17$). In contrast, the blue color indicates that a component-exclusion case elicited greater P1, N1, or P2 peak amplitude, respectively. The corresponding results using PHENV are shown in Figs. 3D,4D,5D, respectively.

Fig. 3 illustrates the results when comparing the cross-correlation of the content words- and function words-exclusion cases with that of the whole sentence. Significantly reduced activity was observed in left-temporal and central sites in the content words-exclusion case ($p < 0.05$, $W \leq 17$; Fig. 3B–D, top mid-panel). This result may reflect a weaker event-related response in regions supporting speech perception located in the left temporal (i.e., left superior temporal gyrus) and central sites. However, the function words-exclusion case did not exhibit such a difference at the N1 peak.

Compared with the whole sentence, the nouns-exclusion case elicited significantly reduced N1 amplitudes in the left temporal and central sites (top mid-panel in Fig. 4B–D; $p < 0.05$, $W \leq 17$). In contrast, the verbs-exclusion case did not exhibit such a difference at the N1 peak.

Fig. 5 shows the results when comparing the cross-correlation of the objects- and subjects-exclusion cases with the whole sentence. While the objects-exclusion case exhibited considerably reduced N1 amplitude at broad fronto-central and temporal sites, the subjects-exclusion case did not show such reduction. Results observed when using PH and PHENV are comparable.

## 4. Discussion

In this study, we examined the cortical tracking of the onset time of phonemes in spoken sentences in different cases: when phoneme information from the whole sentence was used and when omission of critical parts of the sentence was used (i.e., component exclusions). In the component-exclusion cases, three pairs of reciprocal components were investigated: content words vs. function words exclusions, nouns vs. verbs exclusions, and objects vs. subjects exclu-

sions. Our findings reveal that the cross-correlation between the phoneme information and neural responses to various component-exclusion cases and the whole sentence vary significantly. The results show a significant decrease in phase-locking of the N1-evoked amplitude in the cross-correlation coefficients of content words-, nouns-, and objects-exclusion cases compared with that of the whole-sentence case. Such significant differences were observed at language-related brain regions (i.e., left temporal gyrus, left inferior frontal gyrus, and central regions), consistent with the typical auditory N1 scalp peak [11]. Such significant differences were not observed when comparing the cross-correlation of function words-, verbs-, or subjects-exclusion cases with that of the whole-sentence case. The findings indicate the dominance of content words over function words in sentence comprehension and are supported by previous studies [1, 2]. One possible reason for the dominance of nouns over verbs is that nouns are conceptually simpler than verbs [4], which may allow them to be processed more easily than relatively complex concepts referred to by verbs during passive listening. Furthermore, most of the sentences used as stimuli were in active voice, which may favor the importance of objects to subjects.

Given that N1 reflects discrimination of auditory information [10, 11], the main result of this study indicates that the importance of linguistic components may be encoded early as the N1 latency. This assumption is partly in line with a previous study by Moseley *et al.* [13], which found that the brain retrieved the semantic information provided by words and contexts relatively early, at 100–200 ms after word onset. Additionally, the left hemisphere has been considered dominant for processing acoustic information [10, 11]. Our results are consistent with the finding of a left-hemisphere bias of activities observed from topographies in Figs. 3A,4A,5A.

To validate the usefulness of PH, we used PHENV, the combination of PH and the temporal envelope of speech. The temporal envelope has been widely investigated in speech comprehension as it reflects the acoustic changes in a sentence [17, 18, 25]. Thus, the PHENV captures the onset time of phonemes and acoustic information of the speech. Then, we computed the cross-correlation between the PHENV and the neural response corresponding to each sentence stimulus. The grand average of the cross-correlation coefficient shows a significant difference in the central region (Cz electrode) for content words-, nouns-, and objects-exclusion cases compared with the whole-sentence case. These results are in line with those obtained when using the PH. However, the subjects-exclusion case showed a significant difference in the central, while such a difference was not observed when using the PH (Fig. 2C). As expected, in each comparison case (i.e., function words vs. content words exclusions, verbs vs. nouns exclusions, and subjects vs. objects exclusions), brain activation patterns were somewhat akin to the mentioned results. These findings indicate that the phoneme-onset time is a practical aspect to consider while investigating speech comprehension.
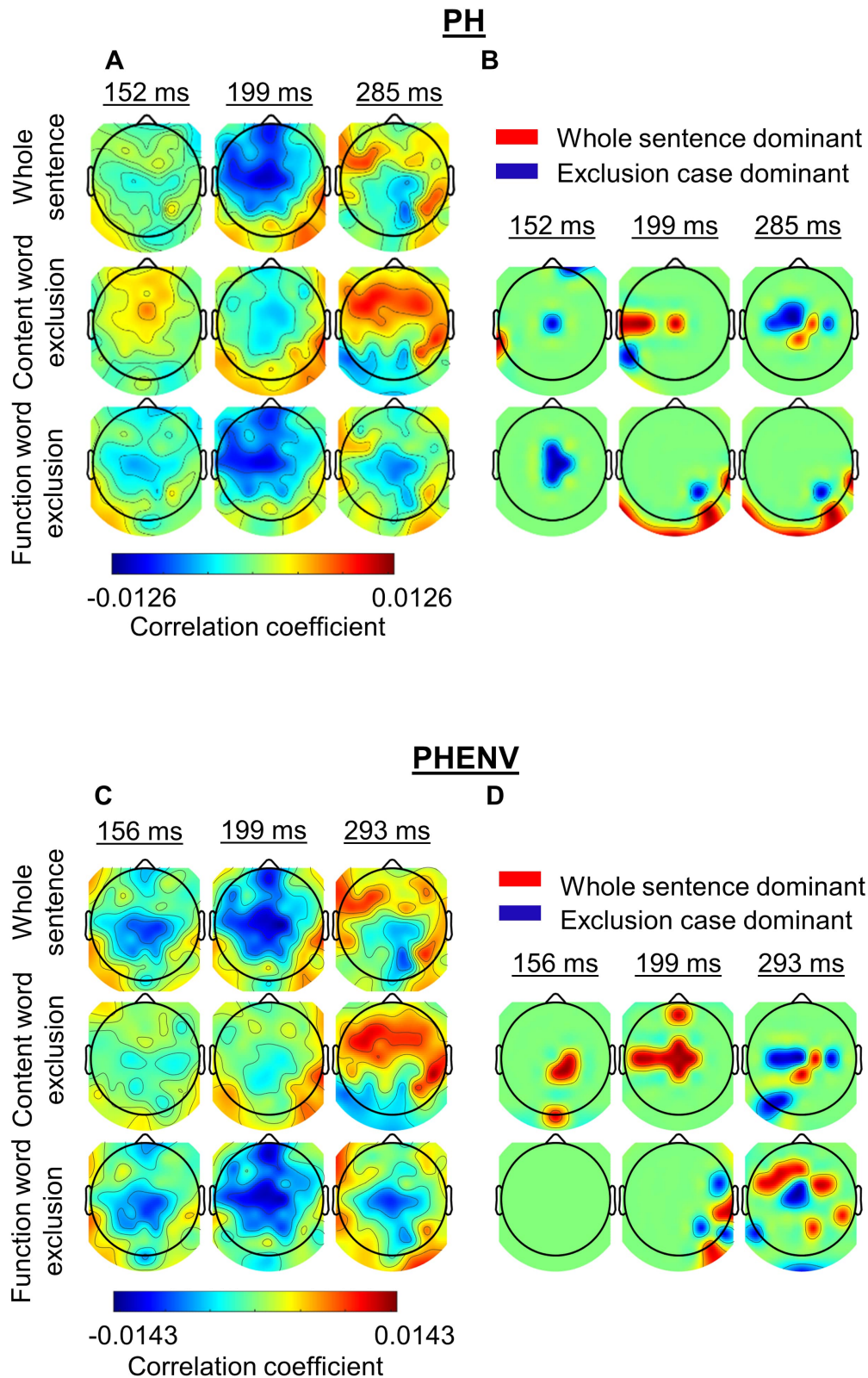
# PH

**A**

152 ms     199 ms     285 ms

Whole sentence

Content word exclusion

Function word exclusion

-0.0126            0.0126
Correlation coefficient

**B**

■ Whole sentence dominant
■ Exclusion case dominant

152 ms     199 ms     285 ms

# PHENV

**C**

156 ms     199 ms     293 ms

Whole sentence

Content word exclusion

Function word exclusion

-0.0143            0.0143
Correlation coefficient

**D**

■ Whole sentence dominant
■ Exclusion case dominant

156 ms     199 ms     293 ms

**Fig. 3. The scalp map of cross-correlation coefficients and dominance for the whole sentence, content words exclusion, and function words exclusion at P1-N1-P2 peak time.** (A) Scalp map of cross-correlation coefficient between the PH of the whole sentence, content words exclusion, and function words exclusion and ERPs at the time lags of 152, 199, and 285 ms. (B) Dominance map showing the whole-sentence vs. content words-exclusion and vs. function words-exclusion cases (Wilcoxon signed-rank test, $p < 0.05$, $W \leq 17$). (C) Scalp map of cross-correlation coefficient between the PHENV of the whole sentence, content words exclusion, and function words exclusion and ERPs at the time lags of 156, 199, and 293 ms. (D) Dominance map showing the whole-sentence vs. content words-exclusion and vs. function words-exclusion cases (Wilcoxon signed-rank test, $p < 0.05$, $W \leq 17$).
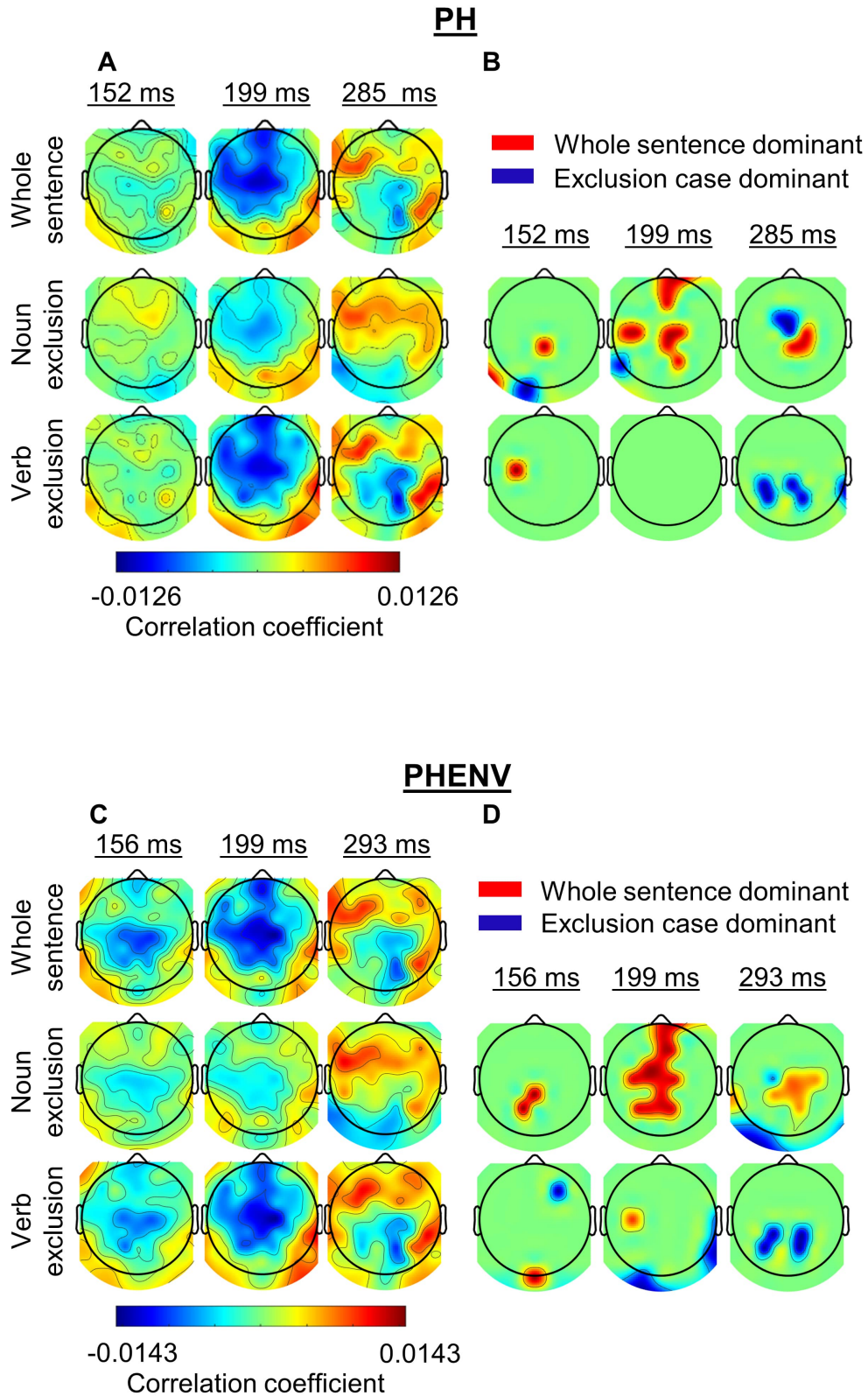
**Fig. 4. The scalp map of cross-correlation coefficients and dominance for the whole sentence, nouns exclusion, and verbs exclusion at P1-N1-P2 peak time.** (A) Scalp map of the cross-correlation coefficient between the PH of the whole sentence, nouns exclusion, and verbs exclusion and ERPs at the time lags of 152, 199, and 285 ms. (B) Dominance map showing whole-sentence vs. nouns-exclusion vs. verbs-exclusion cases (Wilcoxon signed-rank test, $p < 0.05$, $W \leq 17$). (C) Scalp map of the cross-correlation coefficient between the PHENV of the whole sentence, nouns exclusion, and verbs exclusion and ERPs at the time lags of 156, 199, and 293 ms. (D) Dominance map showing whole-sentence vs. nouns-exclusion vs. verbs-exclusion cases (Wilcoxon signed-rank test, $p < 0.05$, $W \leq 17$).
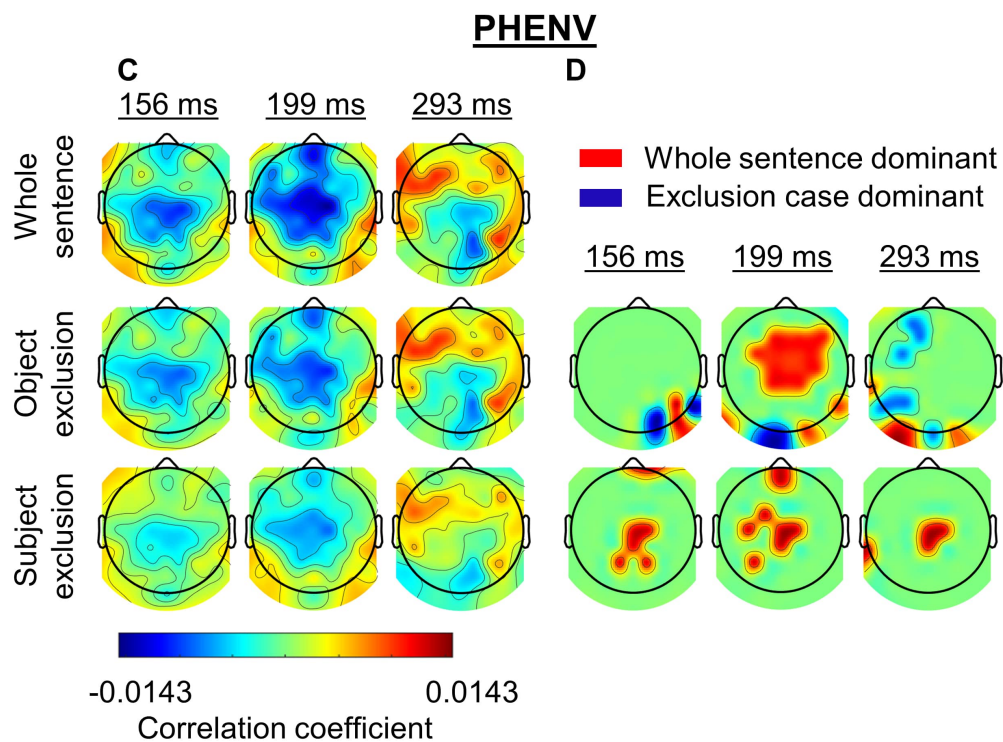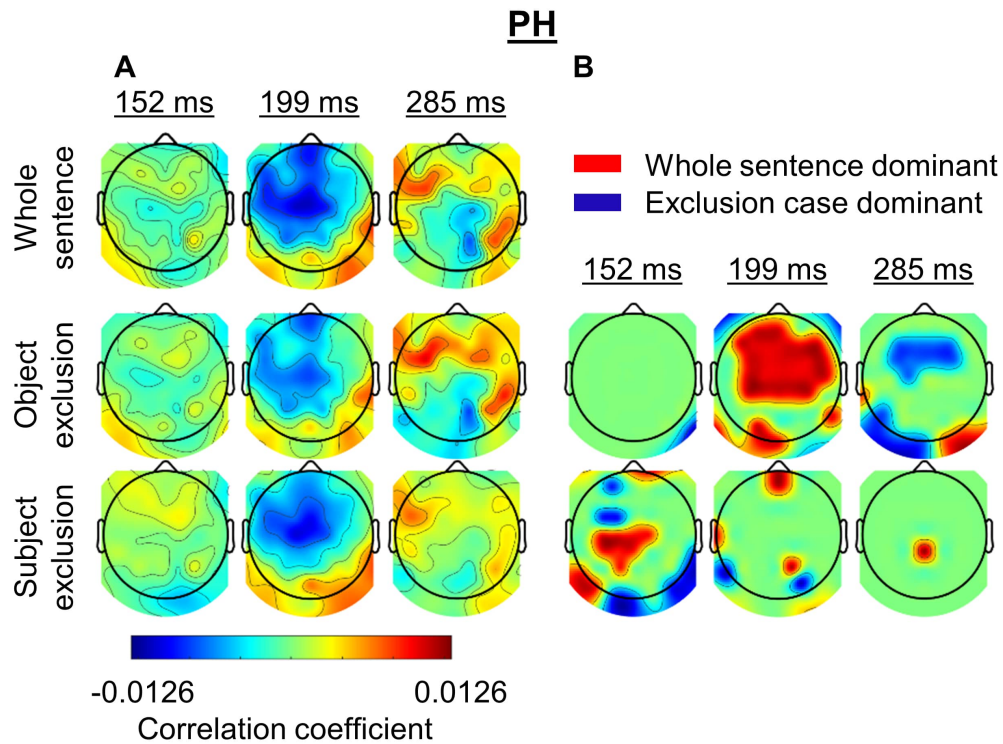
**Fig. 5. The scalp map of cross-correlation coefficients and dominance for the whole sentence, objects-exclusion, and subjects-exclusion at P1-N1-P2 peak time.** (A) Scalp map of the cross-correlation coefficient between the PH of the whole sentence, objects exclusion, and subjects exclusion and ERPs at the time lags of 152, 199, and 285 ms. (B) Dominance map showing the whole-sentence vs. objects-exclusion vs. subjects-exclusion (Wilcoxon signed-rank test, $p < 0.05$, $W \leq 17$). (C) Scalp map of the cross-correlation coefficient between the PHENV of the whole sentence, objects exclusion, and subjects exclusion and ERPs at the time lags of 156, 199, and 293 ms. (D) Dominance map showing the whole-sentence vs. objects-exclusion vs. subjects-exclusion (Wilcoxon signed-rank test, $p < 0.05$, $W \leq 17$).

There are several limitations in the current study as follows. First, all sentences were simple declarative sentences in the active voice. Function words in sentence stimuli act as linking words without holding much meaning, leading to the trivial role of function words in sentence comprehension. Second, our hypothesis was validated based on the brain activations on sensor-space, which can include not only the local active source but also the concurrent electrical sources in the brain [32, 33]. Third, the analysis mainly focused on the early ERP component (N1), which does not reflect other complex processes, such as the integration of semantics and the process of reanalysis. Future studies should employ various types of sentences as stimuli and recruit additional linguistic components of sentence comprehension (e.g., stress and intonation [2, 3]). Analysis of late components in speech-evoked potentials during sentence comprehension and techniques to estimate cortical source activity should also be considered.

## 5. Conclusions

In summary, the phoneme-based ERP analyses reveal the differential importance of linguistic components for sentence comprehension. Such information is encoded early in sentence processing, even while listening to sentences passively. Our findings suggest that content words, nouns, and objects are dominant components in sentence comprehension compared to function words, verbs, and subjects, respectively.

## Abbreviations

ERP, event-related potentials; PH, phoneme onset impulse train; PHENV, a combination of phoneme onset impulse train and temporal envelope.

## Author contributions

IC and JW conceived and designed the experiments; YN performed the experiments; TLT analyzed the data and wrote the paper; IC and JW revised the paper.

## Ethics approval and consent to participate

IRB documents were obtained with the informed consent of all participants. The institutional review board of the University of Iowa approved all the study procedures, code 201609847.

## Acknowledgment

We thank two anonymous reviewers for their comments and suggestions which helped improve the manuscript.

## Funding

## Conflict of interest

The authors declare no conflict of interest.

## References

[1] Shopen T. Language Typology and Syntactic Description. 2nd ed. Cambridge University Press: UK. 2007.

[2] Cutler A. Phonological cues to open- and closed-class words in the processing of spoken sentences. Journal of Psycholinguistic Research. 1993; 22: 109–131.

[3] Cutler A, Foss DJ. On the role of sentence stress in sentence processing. Language and Speech. 1978; 20: 1–10.

[4] Gentner D. Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. Language, Thought and Culture. 1982; 2: 301–334.

[5] Wang J, Cherkassky VL, Yang Y, Chang KK, Vargas R, Diana N, et al. Identifying thematic roles from neural representations measured by functional magnetic resonance imaging. Cognitive Neuropsychology. 2017; 33: 257–264.

[6] Thothathiri M, Asaro CT, Hsu NS, Novick JM. Who did what? A causal role for cognitive control in thematic role assignment during sentence comprehension. Cognition. 2019; 178: 162–177.

[7] Kaan E. Event-Related Potentials and Language Processing: a Brief Overview. Language and Linguistics Compass. 2007; 1: 571–591.

[8] Beres AM. Time is of the Essence: a Review of Electroencephalography (EEG) and Event-Related Brain Potentials (ERPs) in Language Research. Applied Psychophysiology and Biofeedback. 2018; 42: 247–255.

[9] Alday PM. M/EEG analysis of naturalistic stories: a review from speech to language processing. Language, Cognition and Neuroscience. 2019; 34: 457–473.

[10] Friederici AD. Towards a neural basis of auditory sentence processing. Trends in Cognitive Sciences. 2005; 6: 78–84.

[11] Friederici AD. The brain basis of language processing: from structure to function. Physiological Reviews. 2011; 91: 1357–1392.

[12] Shtyrov Y, Hauk O, Pulvermüller F. Distributed neuronal networks for encoding category-specific semantic information: the mismatch negativity to action words. the European Journal of Neuroscience. 2004; 19: 1083–1092.

[13] Moseley RL, Pulvermüller F, Shtyrov Y. Sensorimotor semantics on the spot: brain activity dissociates between conceptual categories within 150 ms. Scientific Reports. 2013; 3: 1928.

[14] Hoenig K, Sim E, Bochev V, Herrnberger B, Kiefer M. Conceptual flexibility in the human brain: dynamic recruitment of semantic maps from visual, motor, and motion-related areas. Journal of Cognitive Neuroscience. 2008; 20: 1799–1814.

[15] Lalor EC, Foxe JJ. Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. European Journal of Neuroscience. 2010; 31: 189–193.

[16] Boynton GM, Engel SA, Glover GH, Heeger DJ. Linear systems analysis of functional magnetic resonance imaging in human V1. Journal of Neuroscience. 1996; 16: 4207–4221.

[17] Aiken SJ, Picton TW. Human cortical responses to the speech envelope. Ear and Hearing. 2008; 29: 139–157.

[18] Horton C, D'Zmura M, Srinivasan R. Suppression of competing speech through entrainment of cortical oscillations. Journal of Neurophysiology. 2014; 109: 3082–3093.

[19] Bidelman GM, Howell M. Functional changes in inter- and intrahemispheric cortical processing underlying degraded speech perception. NeuroImage. 2016; 124: 581–590.

[20] Di Liberto GM, O'Sullivan JA, Lalor EC. Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. Current Biology. 2016; 25: 2457–2465.

[21] Mesgarani N, Cheung C, Johnson K, Chang EF. Phonetic feature encoding in human superior temporal gyrus. Science. 2014; 343: 1006–1010.

[22] Khalighinejad B, Cruzatto da Silva G, Mesgarani N. Dynamic Encoding of Acoustic Features in Neural Responses to Continuous Speech. Journal of Neuroscience. 2017; 37: 2176–2185.

[23] McCloy DR, Lee AKC. Investigating the fit between phonological feature systems and brain responses to speech using EEG. Language, Cognition and Neuroscience. 2019; 34: 662–676.

[24] Di Liberto GM, Pelofi C, Bianco R, Patel P, Mehta AD, Herrero JL, *et al*. Cortical encoding of melodic expectations in human temporal cortex. Elife. 2020; 9: 1–26.

[25] Kong Y, Mullangi A, Ding N. Differential modulation of auditory responses to attended and unattended speech in different listening conditions. Hearing Research. 2015; 316: 73–81.

[26] Bilger RC, Nuetzel JM, Rabinowitz WM, Rzeczkowski C. Standardization of a test of speech perception in noise. Journal of Speech and Hearing Research. 1984; 27: 32–48.

[27] Xue Z, Li J, Li S, Wan B. Using ICA to Remove Eye Blink and Power Line Artifacts in EEG. IEEE. 2006; 107–110.

[28] Delorme A, Palmer J, Onton J, Oostenveld R, Makeig S. Independent EEG sources are dipolar. PLoS ONE. 2012; 7: e30135.

[29] Widmann A, Schröger E, Maess B. Digital filter design for electro-physiological data–a practical approach. Journal of Neuroscience Methods. 2016; 250: 34–46.

[30] Paul B. Praat, a system for doing phonetics by computer. Glot International. 2001; 5: 341–345.

[31] O'Brien CM. Statistics for Bioengineering Sciences: with MATLAB and WinBUGS Support by Brani Vidakovic. International Statistical Review. 2013; 81: 471–472.

[32] Brunner C, Billinger M, Seeber M, Mullen TR, Makeig S. Volume Conduction Influences Scalp-Based Connectivity Estimates. Frontiers in Computational Neuroscience. 2016; 10: 121.

[33] van den Broek SP, Reinders F, Donderwinkel M, Peters MJ. Volume conduction effects in EEG and MEG. Electroencephalography and Clinical Neurophysiology. 1998; 106: 522–534.