Cerebral Cortex, 2022, 1–11

https://doi.org/10.1093/cercor/bhac354

Original Article

Delta-band neural tracking primarily reflects rule-based chunking instead of semantic relatedness between words

Yuhan Lu[1,2,3], Peiqing Jin[3], Nai Ding[3,4,*], Xing Tian[1,2,5,*]

[1]Shanghai Key Laboratory of Brain Functional Genomics (Ministry of Education), School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062, China,

[2]NYU-ECNU Institute of Brain and Cognitive Science at NYU Shanghai, Shanghai 200062, China,

[3]Key Laboratory for Biomedical Engineering of Ministry of Education, College of Biomedical Engineering and Instrument Sciences, Zhejiang University, Hangzhou 310027, China,

[4]Research Center for Applied Mathematics and Machine Intelligence, Research Institute of Basic Theories, Zhejiang Lab, Hangzhou 311121, China,

[5]Division of Arts and Sciences, New York University Shanghai

*Corresponding authors: Xing Tian, New York University Shanghai, Shanghai 200122, China. Email: xing.tian@nyu.edu; Nai Ding, Key Laboratory for Biomedical Engineering of Ministry of Education, College of Biomedical Engineering and Instrument Sciences, Zhejiang University 310027, China. Email: ding_nai@zju.edu.cn

It is debated whether cortical responses matching the time scales of phrases and sentences mediate the mental construction of the syntactic chunks or are simply caused by the semantic properties of words. Here, we investigate to what extent delta-band neural responses to speech can be explained by semantic relatedness between words. To dissociate the contribution of semantic relatedness from sentential structures, participants listened to sentence sequences and paired-word sequences in which semantically related words repeated at 1 Hz. Semantic relatedness in the 2 types of sequences was quantified using a word2vec model that captured the semantic relation between words without considering sentential structure. The word2vec model predicted comparable 1-Hz responses with paired-word sequences and sentence sequences. However, empirical neural activity, recorded using magnetoencephalography, showed a weaker 1-Hz response to paired-word sequences than sentence sequences in a word-level task that did not require sentential processing. Furthermore, when listeners applied a task-related rule to parse paired-word sequences into multi-word chunks, 1-Hz response was stronger than that in word-level task on the same sequences. Our results suggest that cortical activity tracks multi-word chunks constructed by either syntactic rules or task-related rules, whereas the semantic relatedness between words contributes only in a minor way.

Key words: speech; chunking; semantics; entrainment; MEG.

Introduction

A sensory sequence, such as sounds, words, and images, can be represented at different levels of hierarchical structures, from individual items to their associated structures. Chunking has been hypothesized to be one of the sequential representation mechanisms—several contiguous items in a sequence can be grouped into a superordinate chunk that is stored and manipulated as a single unit (Dehaene et al. 2015). In the domain of speech, it is heavily debated how the brain integrates information across words to form multi-word chunks, such as phrases and sentences, and to derive the meaning of such chunks. One viewpoint is that the brain applies tacit syntactic rules to combine words into larger multi-word chunks, forming a hierarchical linguistic structure (Martin and Doumas 2017; Ding et al. 2017a; Meyer and Gumbert 2018); another viewpoint states that the brain does not have recourse to hierarchical syntax at all, but instead, simply binds information across words based on semantic and statistical analysis (Elman 1990; Frank et al. 2012; Christiansen and Chater 2016).

Recent neurophysiological studies show that the neural oscillations at the delta-band frequency (< 4 Hz) are in synchrony with multi-word chunks, e.g. phrases and sentences (Ding et al. 2016; Meyer et al. 2016; Makov et al. 2017; Keitel et al. 2018; Henke and Meyer 2021). Importantly, cortical activity correlates with the time scales of phrases and sentences are independent of prosodic and statistical word cues, which has been taken as strong evidence for the hypothesis that the brain groups words into chunks based on syntactic rules (Fig. 1A and B, solid curve; Ding et al. 2016, 2018). Challenging this position, it is possible that neural tracking of phrases and sentences can be explained by semantic relatedness between words (Jin et al. 2020). In the following, we illustrate this idea using the neural responses to sentences as an example; the same principle also applies to neural responses to phrases.

The simplest form of word sequence processing is to analyze the relations between words. It is well-established that if a word is semantically related to its preceding context, it can be easier to access and its N400 component of event-related potentials (ERP) is reduced (Halgren et al. 2002; Lau et al. 2008; Kutas and Federmeier 2011). In most instances, the words within a sentence are semantically more related to their neighboring words, compared with words in other sentences. The context-dependent neural response, therefore, is generally expected to be high at the beginning of a sentence and to be low at the end of a sentence, forming an apparent neural tracking of sentences (Fig. 1B,

Received: May 13, 2022. **Revised:** August 12, 2022. **Accepted:** August 13, 2022

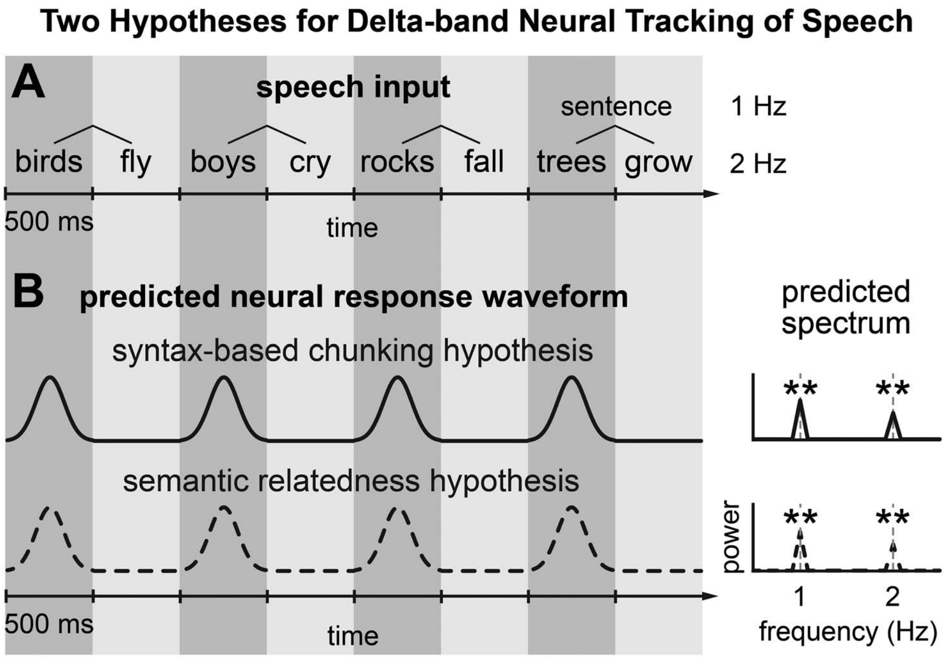

Fig. 1. Two hypotheses for neural tracking of sentences. A) Two-word sentences (noun + verb) are used for illustration purposes. Each word lasts for 500 ms and hence words and sentences are isochronously presented at 2 and 1 Hz, respectively. B) The syntax-based chunking hypothesis assumes a consistent change of neural activity within a mentally constructed chunk. The semantic-relatedness hypothesis assumes that the neural response to a word is attenuated if the word is preceded by a semantically related word. In addition, it is assumed that words within a sentence are more closely related than words across a sentence boundary. Both hypotheses can explain the neural tracking of sentences as illustrated in the response spectrum.

dotted curve). This possibility provides an alternative explanation for neural tracking of sentences without assuming chunk-level representation. It is crucial to investigate to what extent neural tracking of sentences can be explained by semantic relatedness between words so that different hypotheses about neural tracking can be tested.

Relative to automatic syntax-based chunking processes, sequence chunking can be implemented by more controlled processes (Jeon and Friederici 2015). A recent study shows that when listeners are asked to explicitly parse word lists following given chunking rules, the delta-band cortical activity reliably tracks multi-word chunks defined by the rules rather than semantic relatedness between words (Jin et al. 2020). It remains unclear, however, whether semantic relatedness could predominantly drive delta-band neural activity when listeners have access to the semantic properties of individual words in the sentential context.

To investigate these questions, we examined how sequential structures and tasks separately modulated neural response to speech using magnetoencephalography (MEG). First, to investigate the influence of sentential structures, we used a word-level task that directed participants' attention to the semantic properties of individual words. Under this task, we presented to listeners sentence sequences, as well as paired-word sequences designed to frequency tag the neural responses tracking semantic relatedness between words. We quantified the semantic relatedness of the sequences using the word2vec model—a connectionism model that only considers the semantic relation between words (Mikolov et al. 2013). If measured neural responses were inconsistent with the word2vec model predictions, it would suggest that semantic relatedness between words was not enough to explain the neural responses tracking sentences. Next, to assess to what extent neural response to paired-word sequences can be modulated by tasks, we compared neural responses in the word-level task with those in a chunk-level task. If neural tracking responses can be enhanced during the chunk-level task compared with the word-level task, it would suggest again that delta-band cortical activity was primarily related to chunk-level representations and chucking operations rather than semantic relatedness between words.

Materials and methods

Participants

Sixteen native speakers of Mandarin Chinese (19–25 years old, mean 21 years old, 8 males) participated in this study. All listeners were right-handed, with no self-reported hearing loss or neurological disorders. The sample size was predetermined by a previous MEG study on neural tracking of speech in which the post-hoc effect size calculation indicated a powerful study given a cohort of 16 participants (Jin et al. 2020). The experimental procedures were approved by the Research Ethics Committee of Peking University, Zhejiang University, and New York University Shanghai, and were in accordance with the declaration of Helsinki. The participants provided written consent and received financial compensation.

Word sequences

The study presented 3 types of word sequences, i.e. paired-word sequences, random-word sequences, and sentence sequences (Fig. 2A). Each sequence was constructed by 24 disyllabic words: These words were independently synthesized using iFLYTEK synthesizer (http://peiyin.xunfei.cn/; Mandarin Chinese; female voice) and were concatenated without any silence gap between words. The synthesized words were 500 ms in duration and were adjusted to the same intensity, following the procedure in Ding et al. (2016). Individual words synthesized as a whole sounded more natural compared with the speech materials in which each syllable was independently synthesized. Within a word, no further

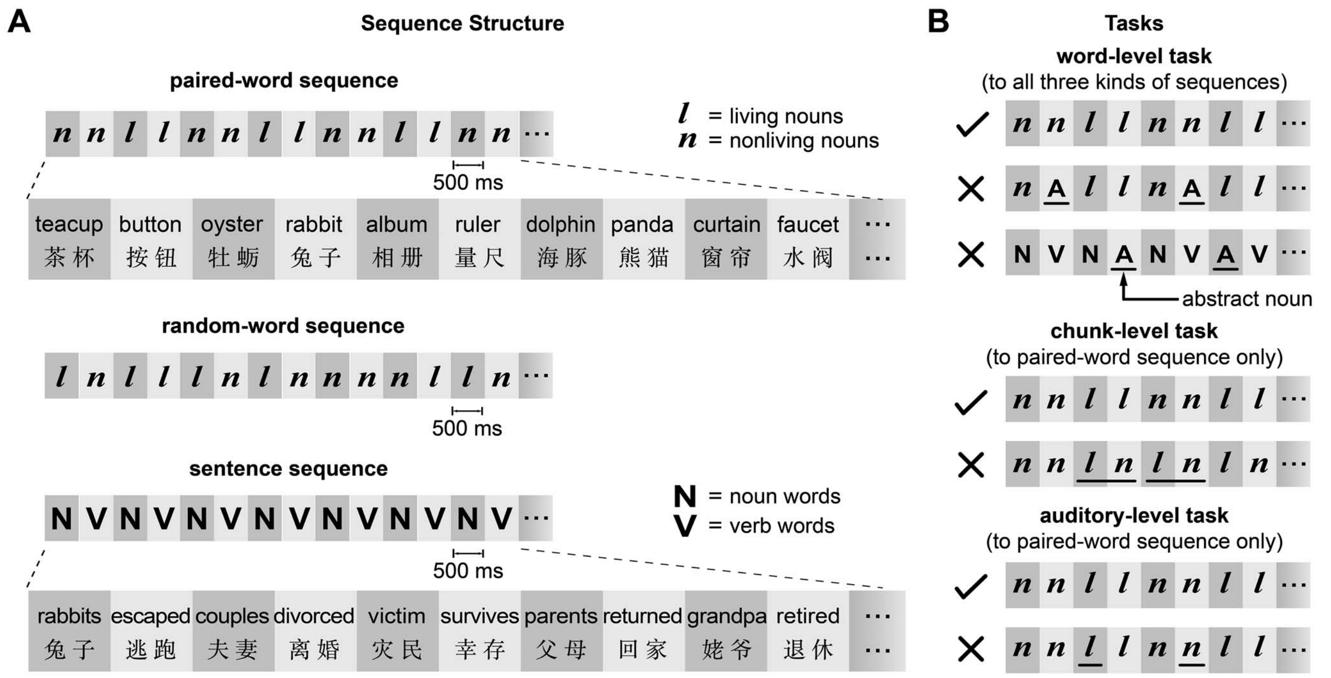

Fig. 2. Stimuli and tasks. A) Sequences of auditory words that were organized according to different rules were used as stimuli. The paired- and random-word sequences consisted of isochronously-presented disyllabic nouns from 2 semantic categories, i.e. living (*l*) and nonliving (*n*) things. In the paired-word sequences, the semantic category of words formed cyclic patterns repeating every 4 words. In the random-word sequences, the 2 semantic categories of words were randomly presented. In the sentence sequences, each sentence was constructed by presenting a disyllabic noun (N) followed by a disyllabic verb (V). B) Multiple tasks were designed by inserting different outliers in the sequences to manipulate and control the level of speech processing. During the word-level task, outlier trials were built by replacing 2 randomly selected words with 2 abstract nouns ("A" in the illustration, e.g. honor and spirit), and listeners were asked to detect these abstract nouns while listening to paired-word, random-word, and sentence sequences. During the chunk-level and auditory-level tasks, participants listened to paired-word sequences only. In the chunk-level task, outlier trials were built by switching the position of a *l*-noun with its neighboring *n*-noun, therefore 2 nouns in a chunk fell into the different semantic categories forming an illegal chunk. Listeners were asked to detect these invalid chunks. During the auditory-level task, outlier trials were built by changing the gender of the voice of 2 randomly selected words, and listeners were asked to detect these voice changes.

control was applied to the intensity and the duration of individual syllables. Coarticulation could exist between these syllables. Each disyllabic word was an acoustically independent unit and was isochronously presented at 2 Hz.

The first 2 kinds of word sequences, i.e. the paired-word sequences and random-word sequences, were constructed by disyllabic nouns ($N = 240$) of living (*l*) and nonliving (*n*) things (see Table S1 for speech materials). In each sequence, the living things were animals ($N = 60$; e.g. monkey and panda) and plants ($N = 60$; e.g. tulip and strawberry); nonliving things were small manipulatable objects ($N = 60$; e.g. teacup and toothbrush), and large non-manipulatable objects ($N = 60$; e.g. playground and hotel). The paired-word sequences and the random-word sequences had no syntactic structure. The sequences were constructed using the following procedure.

In each paired-word sequence, the semantic category of words changed periodically with a cycle of 4 words, i.e. "nnll," with 2 nouns of nonliving things followed by 2 nouns of living things (Fig. 2A). In "nn" or "ll," 2 nouns were randomly selected from the same or different narrow categories. For instance, 2 nouns of living things, i.e. "ll," can be 2 animals or 2 plants together, or a mixture of 1 animal and 1 plant. Based on the position of each noun, semantic relatedness between 2 living or nonliving nouns was high, whereas semantic relatedness between 1 living noun and 1 nonliving noun was low. Therefore, semantic relatedness between words alternated periodically at 1 Hz. Consequently, neural activity tracking semantic relatedness was expected to show periodicity at 1 Hz, which was further illustrated using a word2vec-based semantic relatedness model (see section "Results"). In the random-word sequences, each word was randomly selected from all disyllabic nouns.

In the sentence sequences, 80 four-syllable sentences were constructed, in which the first 2 syllables formed a noun (or a common noun phrase) and the last 2 syllables formed a verb (or a common verb phrase; see Table S1 for speech materials). Since the distinction between words and short phrases was often ambiguous in Chinese, to simplify the discussion, we always referred to the first 2 syllables in a sentence as a disyllabic noun (N) and the last 2 syllables as a disyllabic verb (V). Each sentence was isochronously presented at 1 Hz; Therefore, neural tracking of sentences was expected to show periodicity at 1 Hz.

In the 3 types of sequences, no word was repeated within a sequence. In some conditions, outlier sequences were constructed by replacing words with abstract nouns (e.g. honor and spirit; see Table S1 for speech materials). The abstract noun was chosen from a pool of 30 abstract nouns, and each abstract noun only appeared once in the experiment. The participants were familiarized with these abstract nouns before the experiment.

Procedures and tasks

At the beginning of the experiment, participants were required to familiarize all synthesized words used in this experiment in case they were not used to the synthesized sound. In this session, all words were presented to participants in random order. When participants heard a word, they pressed a key to see the written word on the computer screen. Then, they could choose to listen to the word again or proceed to the next word by pressing different keys.

We first preview the experimental procedures. In the experiment, neural responses were recorded by MEG. The experiment consisted of 5 conditions that were presented in 5 separate blocks. Three blocks separately presented paired-word sequences, random-word sequences, and sentence sequences, and participants were asked to perform the same word-level task. In the other 2 blocks, we only presented paired-word sequences while participants were asked to separately perform a chunk-level task and an auditory-level task. In each block, 20 normal sequences and 5 outlier sequences were mixed and presented in a random order. After listening to a sequence, participants pressed key 0 or key 1 to indicate normal and outlier trials, respectively. After the keypress, the next sequence was presented after a silent interval randomized between 1 and 2 s (uniform distribution). All 5 blocks of conditions were presented in a random order, with the constraint that the 3 blocks using paired-word sequences were following each other and 2 blocks using the random-word sequences and sentence sequences were also next to each other. We applied such constrain so that we could directly compare the task effect and sequential structure effect on neural response without interference. Participants were informed of the task before each condition block and had a 2-min rest between blocks.

The detailed procedures are as follows. In the first 3 blocks, paired-word sequences, random-word sequences, and sentence sequences were presented, respectively, and participants were asked to perform the word-level tasks. In this task, participants were instructed to detect occasionally presented abstract nouns (Fig. 2B, upper panel). An outlier trial had the same design as a normal sequence did, except that 2 randomly selected words in non-adjacent sites were replaced with 2 abstract nouns.

In the other 2 blocks, we only presented paired-word sequences while participants were asked to perform a chunk-level task and an auditory-level task, respectively. In the chunk-level task, participants were instructed that 2 neighboring words in the paired-word sequence could construct a chunk and those 2 words fell into the same semantic category, i.e. 2 words were all living nouns or all nonliving nouns. They had to detect if any chunk contained words from different semantic categories (Fig. 2B, middle panel). To build such invalid chunks in an outlier trial, a living noun was switched with its neighboring nonliving noun, and therefore 2 nouns in a chunk fell into different semantic categories, violating the chunk construction rule. In addition, In the auditory-level task, participants were instructed to detect occasionally presented voice changes (Fig. 2B, lower panel). An outlier trial was constructed by changing the voice of 2 randomly selected words in non-adjacent sites using the change-gender function in Praat (Boersma 2006).

Before MEG recording, instructions were given about all 3 tasks. Therefore, the sequence structure of word sequences was disclosed to participants. Following the instruction of each task, participants were also familiarized with the task by listening to 1 normal sequence and 2 outlier sequences, which were presented randomly. After each sequence, they had to verbally report to the experimenter whether it was a normal or outlier sequence and which words were the outlier, e.g. words with a changed voice, abstract nouns, or 2 words in a chunk falling into different semantic categories. The sequences could be replayed until the participants understood the experimental tasks and correctly reported the outliers.

Data acquisition

Neuromagnetic responses were recorded using a 306-sensor whole-head MEG system (Elekta-Neuromag, Helsinki, Finland) at Peking University, sampled at 1 kHz. The system had 102 magnetometers and 204 planar gradiometers. Four MEG-compatible electrodes were used to record electrooculography (EOG). For MEG source localization purposes, structural magnetic resonance imaging (MRI) data were collected from all participants using a Siemens Magnetom Prisma 3-T MRI system (Siemens Medical Solutions, Erlangen, Germany) at Peking University. A 3-D magnetization-prepared rapid gradient-echo T1-weighted sequence was used to obtain $1 \times 1 \times 1$ mm^3 resolution anatomical images.

Data preprocessing

Only the neural responses to normal sequences were analyzed. The temporal signal-space separation (tSSS) was used to remove the external interference from MEG signals (Taulu and Hari 2009). To remove ocular artifacts both in MEG signals, the horizontal and vertical EOG were regressed out using the least-squares method (Ding et al. 2017b). Since the current study only focused on responses at 1 and 2 Hz, the MEG signals were bandpass filtered between 0.3 and 2.7 Hz using a linear-phase finite impulse response (FIR) filter (−6 dB attenuation at the cut-off frequencies, 10-s Hamming window), and down-sampled at 20 Hz. The response during each trial was extracted.

The 306-channel MEG signals were further denoised using a semi-blind source separation technique, the denoising source separation (DSS). The DSS was a linear transform that decomposed multi-sensor MEG signals into components (de Cheveigné and Simon 2008). The bias function of the DSS was chosen as the response averaged over trials within each condition. A common DSS for all conditions was derived based on the response covariance matrices averaged over conditions. The first 6 DSS components were retained and transformed back to the sensor space for further analysis. This DSS procedure was commonly used to extract cortical responses entrained to speech (Zhang and Ding 2017; Jin et al. 2020).

Frequency-domain analysis

In the frequency-domain analysis, to avoid the response to the sound onset, the responses during the first 2 s of each trial were removed. Consequently, the neural response was 10 s in duration for each trial. The average of all trials was transformed into the frequency domain using the discrete fourier transform (DFT) without any additional smoothing window. The frequency resolution of the DFT analysis was 1/10 Hz. The complex-valued DFT coefficient at frequency f was denoted as $X(f)$, and the response power was $|X(f)|^2$. The DFT was separately applied to each MEG channel. For the MEG response power analysis, responses from the 2 collocated gradiometers were averaged. When the left and right hemispheres were separately analyzed, each hemisphere included 96 gradiometers.

Source localization

The MEG responses averaged over trials were mapped into source space using cortex constrained minimum norm estimate (MNE; Hämäläinen and Ilmoniemi 1994), implemented in the Brainstorm software (Tadel et al. 2011). The T1-weighted MRI images were used to extract the brain volume, cortex surface, and innermost skull surface using Freesurfer (http://surfer.nmr.mgh.harvard.edu/; Dale et al. 1999). Three anatomical landmarks in the MRI images, i.e. nasion, left, and right preauricular points were marked manually. The 3 anatomical landmarks and digitized head points were used to align the MRI images with the MEG sensor array. The forward MEG model was derived based on the overlapping-sphere model (Huang et al. 1999). The identity matrix was used as noise covariance. Source-space activation was measured by the dynamic statistical parametric map (dSPM; Dale et al. 2000). Individual source-space responses, consisting of 15,002 elementary dipoles over the cortex, were rescaled to the ICBM 152 brain template (Fonov et al. 2011) for source-space analyses.

Modeling

Semantic relatedness model

The study was designed to test to what extent neural responses to sentences can be explained by semantic relatedness between words. To quantify the semantic relatedness in word sequences, we used a word2vec model to convert lexical semantics into high-dimensional vectors. The vector representations of semantics have been widely used to account for psycholinguistic and neuroscientific findings (Mitchell et al. 2008; Frank and Yang 2018). Previous studies show that the semantic distance or correlation coefficient between word vectors can predict neural activations during speech comprehension (Frank and Willems 2017; Broderick et al. 2018). Here, we constructed a word2vec-based semantic relatedness model using similar procedures (Fig. S1, see online supplementary material for a color version of this figure) and assumed that simulated neural activity only encoded semantic relatedness in word sequences.

Pulse response

We constructed a semantic relatedness model to simulate how the semantic relatedness contributed to each word sequence. In the model, the smallest unit was a word. Each word was converted to 300-dimensional vectors using the word2vec algorithm learned from a large corpus (the "combination" corpora in Li et al. 2018). Each of the 300-feature dimensions was coded by a real number. The semantic-relatedness feature of a word sequence was first simulated using a pulse sequence, in which a pulse was placed at the onset of each word. The pulse amplitude of each word was defined by comparing (via Pearson's linear correlation coefficient) the current word vector representations with the preceding ones (Frank and Willems 2017; Broderick et al. 2018). The correlation coefficient was a scalar. In addition, because the neural responses to a stimulus were usually weaker if the stimulus was preceded by a similar stimulus, we used one minus the correlation coefficient to modulate the pulse sequences. This measure has proved effective to index the neural semantic processing of speech (Broderick et al. 2018).

Simulation of neural waveform responses

The neural responses in noninvasive electrophysiological recordings were smooth waveforms rather than sharp pulses. Therefore, neural response waveforms were further simulated by convolving the pulse sequences with a response function, which was a 500-ms duration Gaussian window (Jin et al. 2020; Lu et al. 2022). Note that the model assumed a phase-locked response to the onset of each word and the consistent response waveforms were repetitive across all words. The power spectrum of tracking responses cannot be affected by the shape of response waveforms. In other words, using the Gaussian window or other response functions would not change the model predictions of neural tracking results. We chose the Gaussian function to perform the simulation because it was a simple and widely-used approach to simulate neuronal response (Meyer et al. 2017; Lebedev et al. 2019). Lastly, the neural responses to word sequences were transformed into the frequency domain.

Word frequency model

Although the current study aimed to test to what extent semantic relatedness between words can explain the neural responses to speech, we also simulated the properties of neural responses tracking word frequency (Fig. S2, see online supplementary material for a color version of this figure). We retrieved the word frequency in our materials based on a large Chinese corpus (Zhan et al. 2019). Carrying out the same procedure as building the semantic relatedness model, the amplitude of the pulse to each word was set to its value of word frequency and then convolved a response function to derive simulated neural activity responding to word frequency.

Statistical tests

All tests were based on bias-corrected and accelerated bootstrap (Efron and Tibshirani 1994). In the bootstrap procedure, all participants were resampled 10,000 times with replacement. All comparisons in this study were paired comparisons. For a 1-sided comparison of response power, if the data population in 1 condition was greater than A% of the data population in the other condition, the significance level was (100A + 1)/10,001. For 2-sided comparisons, the significance level was (200A + 1)/10,001. A false discovery rate (FDR) correction was applied.

Spectral peak

The statistical significance of a spectral peak at frequency f was tested by comparing the response power at f with the response power in the frequency bin just below f using bootstrap (Jin et al. 2018). The comparison was 1-sided. The comparison only considered the frequency bin below f since spontaneous neural activity had a $1/f$ spectrum and the response power in the frequency bin above f tended to be weaker than the power at f even without stimulus-evoked activity. This significance test was only applied to the response power at 1 and 2 Hz.

Power difference

A 2-sided test was used to compare the normalized power between conditions. The normalized response power was calculated as the power at f minus the power at the frequency bin just below f. This significance test was only applied to the response power at 1 and 2 Hz.

Results

Word2vec model quantified semantic relatedness and predicted the neural tracking of sequences

We used paired-word sequences, sentence sequences, and random-word sequences to investigate to what extent semantic relatedness was involved in the neural tracking of sentences. We used a word2vec-based model to predict the neural responses to semantic relatedness based on the word-level semantic features captured by word2vec. The word2vec algorithm statistically learned from a large corpus constructed 300-dimensional vectorial representations for each word that characterized the semantic features of the word. Words with similar meanings had a higher correlation between their word2vec-based representations. For the paired-word sequences, since every 2 words, i.e. "nn" or "ll," fell into the same semantic category, the semantic similarity was high within the word pairs (Fig. 3A, left 2 plots); and for the sentence sequences, a noun and the following verb constructing a sentence established a predictable semantic and syntactic relations, the semantic correlation was also high within the sentences (Fig. 3A, most right plot, see the bright diagonal line between nouns and verbs).

Based on the word2vec representation, the semantic relatedness model predicted a 1-Hz response to paired-word sequences and sentence sequences ($Ps = 0.0001$ for both sequences, paired 1-sided bootstrap; Fig. 3B). The predicted 1-Hz responses to the sentence sequences and paired-word sequences were not significantly different ($P = 0.251$, paired 2-sided bootstrap; Fig. 3C), suggesting that semantic-relatedness features in the 2 sequences were similar. Furthermore, the semantic relatedness model predicted 2-Hz responses to all 3 sequences, i.e. paired-word sequences, sentence sequences, and random-word sequences ($Ps = 0.0001$ for all sequences, paired 1-sided bootstrap; Fig. 3C); The predicted 2-Hz power was comparable among sequences.

Although the current study aimed to test to what extent the semantic relatedness between words can explain the neural response to speech, we also considered a basic lexical property—word frequency—and analyzed how the word frequency might differ between word sequences. The word frequency model only predicts a 1-Hz response to sentence sequences ($P = 0.0001$; paired 2-sided bootstrap; Fig. S2, see online supplementary material for a color version of this figure), but not to paired-word sequences and random-word sequences.

Next, we tested whether the measured neural responses were consistent with the predictions of the semantic relatedness model and word frequency model using a MEG experiment. Specifically, the experiment tested whether the 1-Hz response to paired-word sequences was similar to the 1-Hz response to sentence sequences, as predicted by the word2vec-based semantic relatedness model. If the measured 1-Hz response to the 2 sequences were statistically similar, it suggests that delta-band neural tracking of sentences can be sufficiently explained by semantic relatedness between words; Otherwise, the neural tracking response cannot be explained by the semantic relatedness.

Neural tracking of semantic relatedness in sequences

In 3 separate blocks, participants listened to the paired-word sequences, sentence sequences, and random-word sequences, whereas they were asked to perform the word-level task—monitoring occasionally presented abstract nouns. The behavioral accuracy was 91 ± 2%, 88 ± 2%, and 95 ± 1%, for the paired-word sequences, sentence sequences, and random-word sequences, respectively (Fig. S3, see online supplementary material for a color version of this figure). In the MEG spectrum shown in Fig. 3D, a significant 1-Hz peak was observed, bilaterally distributed (Fig. 3EF), for the paired-word sequences and the sentence sequences ($Ps = 0.0006$ and 0.0001 for paired-word and sentence sequences, respectively, paired 1-sided bootstrap), but not for the random-word sequences ($P = 0.468$). The significant 1-Hz response to the paired-word sequence was consistent with the prediction of the word frequency model. However, it was significantly weaker than the 1-Hz response to sentence sequences in both left and right hemispheres ($Ps = 0.0003$ for both hemispheres, paired 2-sided bootstrap; Fig. 3G), including left inferior frontal gyrus (IFG), left middle temporal gyrus (MTG), and bilateral superior temporal sulcus (STS; Fig. S4A, see online supplementary material for a color version of this figure). The weaker 1-Hz response to paired-word sequences relative to sentence sequences was inconsistent with the semantic relatedness model, which predicted a comparable 1-Hz response power between paired-word sequences and sentence sequences (Fig. 3C, upper panel). Furthermore, significant bilaterally distributed 2-Hz responses were observed in all sequences ($Ps < 0.0003$, paired 1-sided bootstrap; Fig. 3DEF) and were not significantly different among sequences (Fig. 3H), which were consistent with the model prediction showing that neural tracking of individual words was equally strong regardless of sequential structures. These results suggested that neural tracking of semantic relatedness between words was rather weak and cannot sufficiently explain neural tracking of sentences.

Task-dependent neural tracking of semantic relatedness

Next, we further tested whether neural tracking could be enhanced during explicit sequence chunking in the chunk-level task compared with the responses in the word-level task. In addition to the word-level task, listeners were asked to perform another 2 tasks, i.e. chunk-level task and auditory task, in separate blocks while listening to the paired-word sequences. The behavioral accuracy was 91 ± 2%, 80 ± 5%, and 96 ± 2% (mean ± SEM across subjects) for the word-level, chunking-level, and auditory-level tasks, respectively (Fig. S3, see online supplementary material for a color version of this figure). In the MEG spectrum shown in Fig. 4A, 1-Hz responses to paired-word sequence were observed during chunk-level tasks ($P = 0.0001$, paired 2-sided bootstrap) but not during the auditory task ($P = 0.456$). To compare neural responses between tasks, we replicated the results of the word-level task from Fig. 3. The response topography of significant 1-Hz responses showed a bilateral activation for the chunk-level task, and the neural sources were mainly located in bilateral temporal and frontal lobes for chunk-level tasks (Fig. 4BC). Importantly, the 1-Hz response to the paired-word sequences was significantly stronger during the chunk-level task than the word-level task ($Ps = 0.0003$ for both hemispheres, paired 1-sided bootstrap; Fig. 4D), including left IFG, left MTG, and bilateral STS (Fig. S4B, see online supplementary material for a color version of this figure). Furthermore, significant bilaterally distributed 2-Hz responses were observed in all tasks ($Ps < 0.0003$, paired 1-sided bootstrap; Fig. 4ABC) and were not significantly different (Fig. 4E). These results suggest that the chunk-level task, i.e. explicit sequence chunking, can drive cortical responses more effectively than semantic relations between words, even when semantic relatedness was identical in the same paired-word sequences.

Discussion

How the brain groups words into multi-word chunks, e.g. phrases and sentences, has been heavily debated. Numerous studies support that the neural tracking of phrases and sentences more likely reflects mental representations of multi-word chunks (Ding et al. 2016, 2017b; Jin et al. 2020) and cannot be fully explained by lexical properties of individual words (Jin et al. 2020; Burroughs et al. 2021; Lu et al. 2022). However, it remains unclear whether the neural tracking of speech can be explained by semantic relatedness between words. The study compared neural activity tracking semantic relatedness with neural activity tracking 2 types of multi-word chunks, i.e. sentences defined by tactic syntactic knowledge and chunks defined by artificial chunking rule. Our findings showed that delta-band neural activity was predominately driven by either syntax-defined sentences (Fig. 3G) or artificial rule-defined chunks (Fig. 4D), rather than semantic relatedness between words.

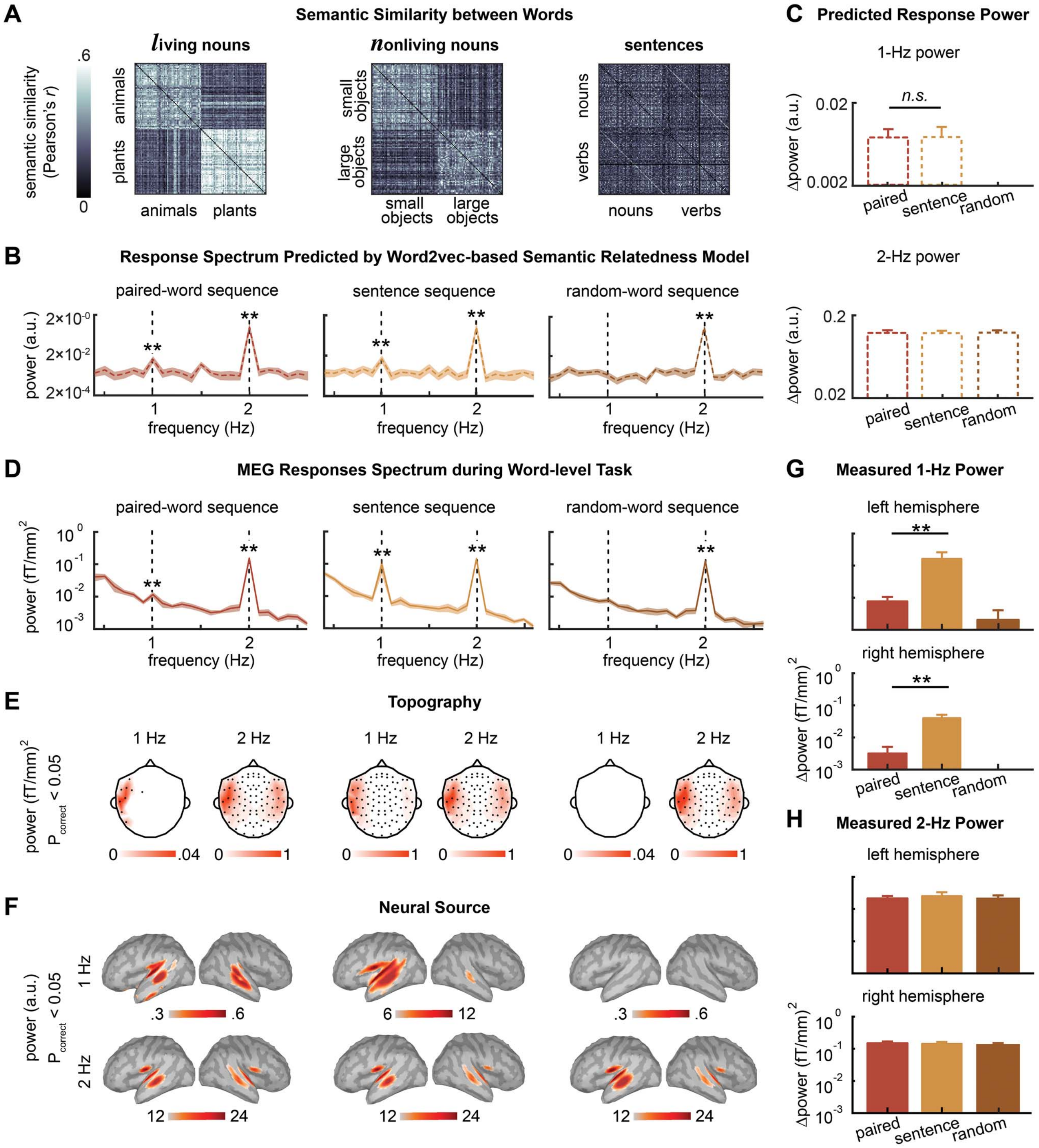

Fig. 3. Word2vec model predictions and measured MEG responses regarding differences in neural tracking among conditions. A) Correlation coefficient between the word2vec representations of words. Words in each subcategory, e.g. animals and plants, were correlated, and nouns and verbs in the same sentence also showed correlation. B) Predicted response spectrum. The semantic relatedness model predicted 1-Hz responses to the paired-word sequences and sentence sequences and 2-Hz responses to all 3 sequences. C) Predicted response power at 1 and 2 Hz. The semantic relatedness models predicted comparable 1-Hz responses to the paired-word sequences and the sentence sequences. D) Measured MEG responses to paired-word sequences, sentence sequences, and random-word sequences. The response spectrum was averaged over participants and MEG gradiometers. E and F) For the topography (E) and source localization (F) results, only statistically significant sensors (shown by black dots) and vertices ($Ps < 0.05$; bootstrap, FDR corrected) were shown. G and H), Normalized power at 1 Hz (G) and 2 Hz (H) in the left and right hemispheres. The 1- and 2-Hz power was normalized by subtracting the power in a neighboring frequency bin. The 1-Hz response to sentence sequences was significantly stronger than that to paired-word sequences in both hemispheres. The shaded areas and error bars represent 1 standard error of the mean (SEM) across subjects. ** $P < 0.005$.

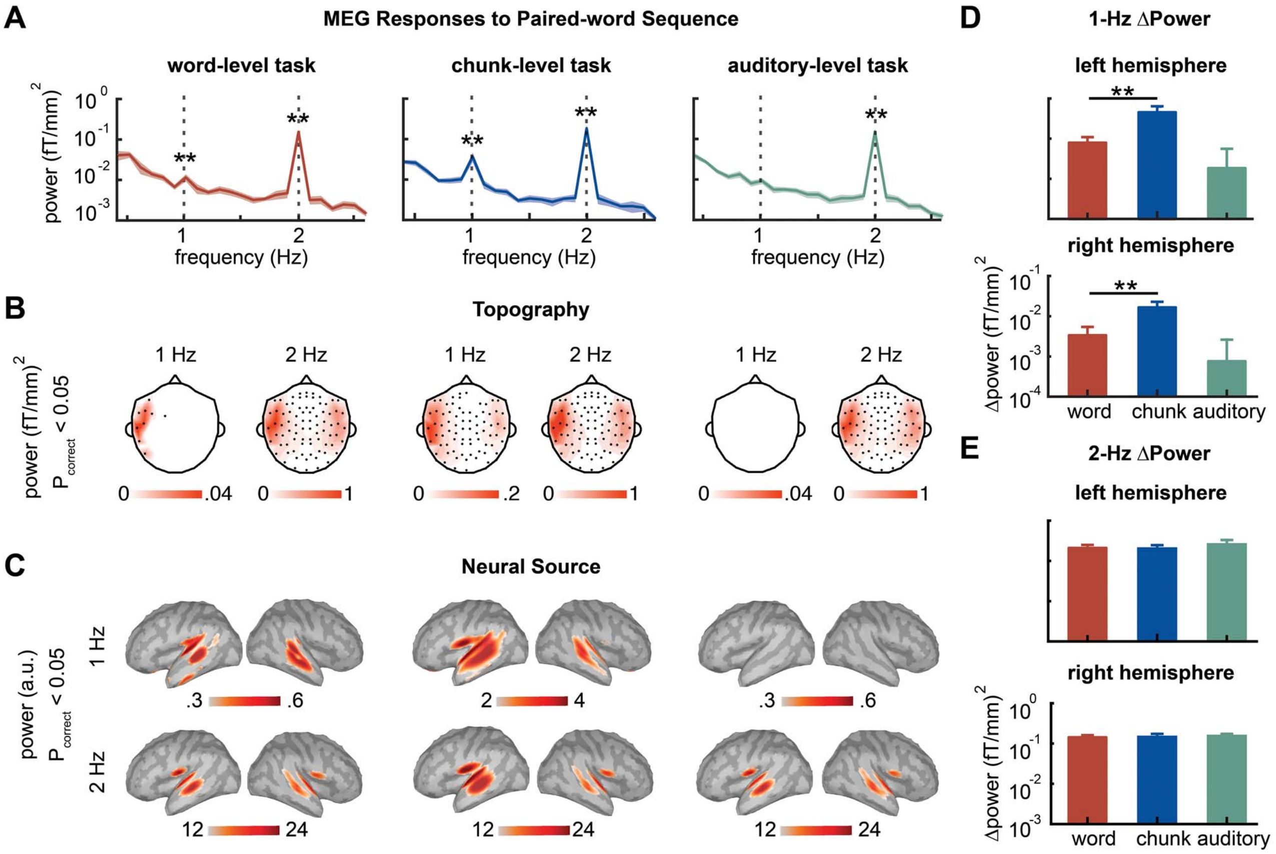

Fig. 4. MEG responses to paired-word sequences in word-level, chunk-level, and auditory-level tasks. A) The response spectrum was averaged over participants and MEG gradiometers. The shaded area covered 1 SEM over participants on each side. B and C) For the topography (B) and source localization (C) results, only statistically significant sensors (shown by black dots) and vertices ($P < 0.05$; bootstrap, FDR corrected) were shown. D and E) Normalized power at 1 Hz (D) and 2 Hz (E) in the left and right hemispheres. The 1-Hz power was normalized by subtracting the power in a neighboring frequency bin. The 1-Hz response was stronger during the chunk-level task than that during the word-level task in both hemispheres. Error bars represent 1 SEM across subjects. ** $P < 0.005$.

Neural encoding of semantic relatedness of words

Beyond single word analysis, the simplest form of word-level analysis is to analyze the semantic relations between adjacent words. Here, semantic relatedness refers to both semantic similarity (e.g. "travel"–"journey") and semantic associations (e.g. travel–plan). The semantic relatedness hypothesis is explainable for the neural representation of sentence structures, which is built on the priming effect in the psychological literature (Tulving and Schacter 1990) and the neural adaptation in neuroscience literature (Grill-Spector et al. 2006). The semantic relatedness between words can strongly affect single-sensor MEG/EEG responses and the global field power. For example, the classic N400 response is sensitive to semantic relatedness between words and can be observed in single MEG/EEG sensors (Kutas and Federmeier 2011). In the current study, we show that neural tracking of semantic relatedness between words is rather weaker than that of sentences during the same word-level task, which is inconsistent with predictions of the semantic relatedness model. Since the N400 is a robust neural marker of the semantic relatedness between words, it may seem surprising why it does not drive a strong 1-Hz response to the paired-word sequences. One potential reason is that the semantic categories used here are broad (e.g. animals and plants) and therefore the words within each category are generally not strongly related, which attenuates the N400 effect (Federmeier et al. 2003). Furthermore, the localization of the N400 effect has been found a strong left lateralization in neuroimaging and neurophysiological studies, e.g. left superior temporal areas (see a review, see Lau et al. 2008). The N400 effects have been hypothesized either as the results of integration (Hagoort et al. 2004) or semantic retrieval (Kutas and Hillyard 1980; Kutas and Federmeier 2011), similar to the current study of integration based on semantic relatedness. Therefore, the observations of left-lateralized neural tracking shown in the topography (Fig. 4B), are consistent with previous studies, suggesting possible left dominant neural structures that mediate semantic processing.

Besides the semantic relatedness hypothesis, it has been argued that other word properties, such as part-of-speech and lexical semantic information, can potentially explain neural tracking of phrases and sentences, without inferring phrase and sentence processing (Frank and Yang 2018). However, recent studies provide negative evidence for neural tracking of lexical semantics and part-of-speech information (Jin et al. 2020; Burroughs et al. 2021; Lo et al. 2022; Lu et al. 2022). For example, neural tracking of lexical semantics was weaker than that of sentences (Lu et al. 2022). After shuffling the order of words in sentences (e.g. "cotton sheep eat grass" to "sheep cotton grass eat"), the new stimuli that contain identical periodical part-of-speech information as the sentence but without correct syntactic structures did not induce neural responses at the rate of part-of-speech information (Lo et al. 2022). Taken together with the current study, it suggests that word-level features cannot fully explain neural tracking of phrases and sentences.

Mental construction of multi-word chunks

Neural tracking of multi-word chunks, e.g. phrases and sentences, is a widely-observed phenomenon. Functional MRI and stereo-EEG studies show that the mental construction of speech chunks is represented in bilateral temporal and frontal lobes (Ding et al. 2016; Bulut et al. 2017; Nelson et al. 2017; Keitel et al. 2018). Consistent with previous evidence, the current study shows that bilaterally distributed neural responses reliably track multi-word chunks defined by an explicit chunking rule (Fig. 4D) or implicit grammatic knowledge (Fig. 3G). Although the 2 kinds of multi-word chunks can both drive strong large-scale neural activity, it remains to be established whether they involve a similar neural mechanism. It has been hypothesized that the brain is sensitive to and encodes chunk boundaries following a domain-general computation principle. Electrophysiological studies have provided preliminary evidence supporting this hypothesis. For example, a closure positive shift response was observed at phrasal boundaries in speech (Steinhauer et al. 1999; Li and Yang 2009) and structural boundaries in music (Zhang et al. 2016). Transient neural activities are observed at event boundaries in speech (Ding et al. 2016), nonspeech sounds (Chait et al. 2007; Sohoglu and Chait 2016), and movies (Zacks et al. 2001).

Although a domain-general computation principle for chunk processing has been proposed, it is elusive whether the implementation of the principle in different contexts is mediated by a similar neural mechanism. First, neural tracking of sentences can be observed when listeners focus on semantic features of individual words, suggesting that sentential analysis is to some extent automatic, similar to what has been found for word-level processing (Ding et al. 2018). In contrast, a chunk-level task that requires listeners to explicitly parse sequences reflects a more controlled process that may engage sequential decision-making (O'Connell and Hofmann 2012; Barascud et al. 2016). Second, neuroimaging studies show that different brain regions are involved in language processing with different rule complexities. For instance, newly-learned artificial rules with simple-to-superordinate complexities activate posterior-to-anterior gradients in the frontal lobe (Koechlin and Summerfield 2007; Badre and Nee 2018), whereas syntactic rules with different complexities consistently activate Broca's area (Jeon and Friederici 2015). Taken together, these evidence suggests that explicit sequences chunking is likely to engage executive control processes that are not required for natural speech comprehension (Koechlin and Summerfield 2007). Therefore, it is possible that the conclusion based on the explicit sequences chunking task cannot generalize to other speech processing tasks that only implicitly or do not require sequences chunking.

Recent work has shown that syllabic rates across languages are generally within the range of 2–8 Hz (Ding et al. 2017c; Poeppel and Assaneo 2020), whereas neural responses to higher structures than syllables tend to occur below 4 Hz, falling in the delta-band range (usually defined as 1–3 Hz). It is still under debated whether the neural entrainment at a chunk rate is related to intrinsic neural oscillations. Neurophysiological studies show that the phrase of slow neural oscillations can mediate temporal attention and prediction (Arnal and Giraud 2012; Morillon et al. 2014), which associate with the neurolinguistic interpretation that attention is applied to the onset of each multi-word chunk (Astheimer and Sanders 2009). Therefore, the cortical dynamics on the delta-band provide a suitable timescale for multi-word chunk analysis (Keitel et al. 2018; Lo et al. 2022; Lu et al. 2022). Furthermore, within a multi-word chunk, the power of neural activity has been found a sustained increase or build upon other frequency scales, for example, beta (Bastiaansen et al. 2010; Bastiaansen and Hagoort 2015), gamma (Peña and Melloni 2012), and high gamma (Ding et al. 2016; Nelson et al. 2017). Neural tracking in different frequency scales may reflect strong coupling across those frequency bands (Lakatos et al. 2005; Giraud and Poeppel 2012) and processing of semantic and syntactic information (Bastiaansen and Hagoort 2015).

Task modulation of speech processing

It is well-established that neural response to speech is strongly modulated by tasks (Makov et al. 2017; Ding et al. 2018). First, the mental construction of multi-word chunks can overwhelm neural tracking of semantic relatedness between words. Jin et al. (2020) designed word sequences in which 2 kinds of semantically similar word pairs varied periodically (e.g. "teacup-button oyster-rabbit") and asked listeners to group pairs of words based on the experimental rules. The results show that neural activity prominently tracks multi-word chunks defined by the rules rather than the semantic relatedness between words. The current study builds on and extends this finding by demonstrating delta-band neural tracking of semantic relatedness is rather weak even when listeners have access to the semantic properties of individual words in the sentential context. Second, a previous study shows that listeners can recall the gender of a speaker of unattended speech (Cherry 1953), indicating that the gender detection task is an easy task and acoustic information can be processed even without top-down attention. The current study does not observe a 1-Hz response to paired-word sequences during the auditory-level task, which is also consistent with a previous electrophysiological study that the N400 component is attenuated only when semantic processing is required (Chwilla et al. 1995).

In summary, delta-band neural activity is more effectively driven by implicit syntactic processing and explicit sequence chunking, rather than semantic relatedness between words. Together with similar studies that distinguish the role of word properties and phrasal/sentential chunk structures (Jin et al. 2020; Kaufeld et al. 2020; Lo 2021; Lu et al. 2022), delta-band neural activity is better explained by chunk-level representations, rather than word-level representations.

Acknowledgments

We thank Cheng Cheng for excellent assistance in data collection, Jiajie Zou for discussion and generating the speech materials, Yuchunzi Wu, Stefan Frank, Lucia Melloni, Lang Qin, and David Poeppel for thoughtful comments on previous versions of the manuscript.

Supplementary material

Supplementary material is available at *Cerebral Cortex* online.

Funding

This work was supported by the National Natural Science Foundation of China (32071099 to X.T.), Key R&D Program of Zhejiang Lab (2022C03011 to N.D.), Natural Science Foundation of Shanghai (20ZR1472100 to X.T.), Program of Introducing Talents of Discipline to Universities (Base B16018 to X.T.), and NYU Shanghai Boost Fund (to X.T.).

Conflict of interest statement: The authors declare no competing financial interests.

Data and code availability

The computer code used for MEG analyses is available upon request.

Author contributions

N. D. designed the research; Y. L. and P. J. performed experiments and analyzed the data; Y. L., X.T., N. D.; and P. J. wrote and edited the paper.

References

Arnal LH, Giraud A-L. Cortical oscillations and sensory predictions. *Trends Cogn Sci*. 2012:16(7):390–398.

Astheimer LB, Sanders LD. Listeners modulate temporally selective attention during natural speech processing. *Biol Psychol*. 2009:80(1):23–34.

Badre D, Nee DE. Frontal cortex and the hierarchical control of behavior. *Trends Cogn Sci*. 2018:22(2):170–188.

Barascud N, Pearce MT, Griffiths TD, Friston KJ, Chait M. Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proc Natl Acad Sci U S A*. 2016:113(5):E616–E625.

Bastiaansen M, Hagoort P. Frequency-based segregation of syntactic and semantic unification during online sentence level language comprehension. *J Cogn Neurosci*. 2015:27(11):2095–2107.

Bastiaansen M, Magyari L, Hagoort P. Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *J Cogn Neurosci*. 2010:22(7):1333–1347.

Boersma P. *Praat: doing phonetics by computer*. 2006 http://www.praat.org/ (Accessed May 20, 2022).

Broderick MP, Anderson AJ, Di Liberto GM, Crosse MJ, Lalor EC. Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. *Curr Biol*. 2018:28(5):803–809.e3.

Bulut T, Hung YH, Tzeng O, Wu DH. Neural correlates of processing sentences and compound words in Chinese. *PLoS One*. 2017:12(12):e0188526.

Burroughs A, Kazanina N, Houghton C. Grammatical category and the neural processing of phrases. *Sci Rep*. 2021:11(1):2446.

Chait M, Poeppel D, de Cheveigné A, Simon JZ. Processing asymmetry of transitions between order and disorder in human auditory cortex. *J Neurosci*. 2007:27(19):5207–5214.

Cherry EC. Some experiments on the recognition of speech, with one and with two ears. *J Acoust Soc Am*. 1953:25(5):975–979.

Christiansen MH, Chater N. The now-or-never bottleneck: a fundamental constraint on language. *Behav Brain Sci*. 2016:39:e62.

Chwilla DJ, Brown CM, Hagoort P. The N400 as a function of the level of processing. *Psychophysiology*. 1995:32(3):274–285.

Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis: I. segmentation and surface reconstruction. *NeuroImage*. 1999:9(2):179–194.

Dale AM, Liu AK, Fischl BR, Buckner RL, Belliveau JW, Lewine JD, Halgren E. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*. 2000:26(1):55–67.

de Cheveigné A, Simon JZ. Denoising based on spatial filtering. *J Neurosci Methods*. 2008:171(2):331–339.

Dehaene S, Meyniel F, Wacongne C, Wang L, Pallier C. The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron*. 2015:88(1):2–19.

Ding N, Melloni L, Zhang H, Tian X, Poeppel D. Cortical tracking of hierarchical linguistic structures in connected speech. *Nat Neurosci*. 2016:19(1):158–164.

Ding N, Melloni L, Tian X, Poeppel D. Rule-based and word-level statistics-based processing of language: insights from neuroscience. *Language, Cognition and Neuroscience*. 2017a:32(5):570–575.

Ding N, Melloni L, Yang A, Wang Y, Zhang W, Poeppel D. Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Front Hum Neurosci*. 2017b:11:481.

Ding N, Patel AD, Chen L, Butler H, Luo C, Poeppel D. Temporal modulations in speech and music. *Neurosci Biobehav Rev*. 2017c:81 (Pt B):181–187.

Ding N, Pan X, Luo C, Su N, Zhang W, Zhang J. Attention is required for knowledge-based sequential grouping: insights from the integration of syllables into words. *J Neurosci*. 2018:38(5):1178–1188.

Efron B, Tibshirani RJ. In: Cox DR, Hinkley DV, Reid N, Rubin DB, and Silverman BW, editors. *An introduction to the bootstrap*. New York: CRC press; 1994.

Elman J. Finding structure in time. *Cogn Sci*. 1990:14(2):179–211.

Federmeier KD, Van Petten C, Schwartz TJ, Kutas M. Sounds, words, sentences: age-related changes across levels of language processing. *Psychol Aging*. 2003:18(4):858–872.

Fonov V, Evans AC, Botteron K, Almli CR, McKinstry R, Collins DL, Brain Development Cooperative Group. Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*. 2011:54(1):313–327.

Frank SL, Willems RM. Word predictability and semantic similarity show distinct patterns of brain activity during language comprehension. *Lang, Cogn Neurosci*. 2017:32(9):1192–1203.

Frank SL, Yang J. Lexical representation explains cortical entrainment during speech comprehension. *PLoS One*. 2018:13(5):e0197304.

Frank SL, Bod R, Christiansen MH. How hierarchical is language use? *Proc Biol Sci*. 2012:279(1747):4522–4531.

Giraud AL, Poeppel D. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci*. 2012:15(4):511–517.

Grill-Spector K, Henson R, Martin A. Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci*. 2006:10(1):14–23.

Hagoort P, Hald L, Bastiaansen M, Petersson KM. Integration of word meaning and world knowledge in language comprehension. *Science*. 2004:304(5669):438–441.

Halgren E, Dhond RP, Christensen N, Van Petten C, Marinkovic K, Lewine JD, Dale AM. N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *NeuroImage*. 2002:17(3):1101–1116.

Hämäläinen MS, Ilmoniemi RJ. Interpreting magnetic fields of the brain: minimum norm estimates. *Med Biol Eng Comput*. 1994:32(1):35–42.

Henke L, Meyer L. Endogenous oscillations time-constrain linguistic segmentation: cycling the garden path. *Cereb Cortex*. 2021:31(9):4289–4299.

Huang M, Mosher JC, Leahy R. A sensor-weighted overlapping-sphere head model and exhaustive head model comparison for MEG. *Phys Med Biol*. 1999:44(2):423–440.

Jeon H-A, Friederici AD. Degree of automaticity and the prefrontal cortex. *Trends Cogn Sci*. 2015:19(5):244–250.

Jin P, Zou J, Zhou T, Ding N. Eye activity tracks task-relevant structures during speech and auditory sequence perception. *Nat Commun*. 2018:9(1):5374.

Jin P, Lu Y, Ding N. Low-frequency neural activity reflects rule-based chunking during speech listening. *Elife*. 2020:9. https://doi.org/10.7554/eLife.55613.

Kaufeld G, Bosker HR, Ten Oever S, Alday PM, Meyer AS, Martin AE. Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. *J Neurosci*. 2020:40(49): 9467–9475.

Keitel A, Gross J, Kayser C. Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS Biol*. 2018:16(3):e2004473.

Koechlin E, Summerfield C. An information theoretical approach to prefrontal executive function. *Trends Cogn Sci*. 2007:11(6):229–235.

Kutas M, Federmeier KD. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol*. 2011:62:621–647.

Kutas M, Hillyard SA. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*. 1980:207(4427):203–205.

Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE. An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J Neurophysiol*. 2005:94(3): 1904–1911.

Lau EF, Phillips C, Poeppel D. A cortical network for semantics: (de)constructing the N400. *Nat Rev Neurosci*. 2008:9(12):920–933.

Lebedev M, Ossadtchi A, Adell Mill N, Urpí N, Cervera M, Nicolelis M. Analysis of neuronal ensemble activity reveals the pitfalls and shortcomings of rotation dynamics. *Sci Rep*. 2019:9:18978.

Li W-J, Yang Y. Perception of prosodic hierarchical boundaries in mandarin Chinese sentences. *Neuroscience*. 2009:158(4): 1416–1425.

Li S, Zhao Z, Hu R, Li W, Liu T, Du X. Analogical reasoning on chinese morphological and semantic relations. *arXivpreprint arXiv:1805.06504*. 2018.

Lo C-W. *Testing low-frequency neural activity in sentence understanding*. Michigan, USA: University of Michigan; 2021

Lo C-W, Tung T-Y, Ke AH, Brennan JR. Hierarchy, not lexical regularity, modulates low-frequency neural synchrony during language comprehension. *Neurobiol Lang*. 2022:1–41.

Lu Y, Jin P, Pan X, Ding N. Delta-band neural activity primarily tracks sentences instead of semantic properties of words. *NeuroImage*. 2022:251:118979.

Makov S, Sharon O, Ding N, Ben-Shachar M, Nir Y, Zion-Golumbic E. Sleep disrupts high-level speech parsing despite significant basic auditory processing. *J Neurosci*. 2017:37(32):7772–7781.

Martin AE, Doumas LA. A mechanism for the cortical computation of hierarchical linguistic structure. *PLoS Biol*. 2017:15(3):e2000663.

Meyer L, Gumbert M. Synchronization of electrophysiological responses with speech benefits syntactic information processing. *J Cogn Neurosci*. 2018:30(8):1066–1074.

Meyer L, Henry MJ, Gaston P, Schmuck N, Friederici AD. Linguistic bias modulates interpretation of speech via neural delta-band oscillations. *Cereb Cortex*. 2016:27(9):4293–4302.

Meyer AF, Williamson RS, Linden JF, Sahani M. Models of neuronal stimulus-response functions: elaboration, estimation, and evaluation. *Front Syst Neurosci*. 2017:10. https://doi.org/10.3389/fnsys.2016.00109.

Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. 2013: arXiv preprint arXiv:1301.3781.

Mitchell TM, Shinkareva SV, Carlson A, Chang K-M, Malave VL, Mason RA, Just MA. Predicting human brain activity associated with the meanings of nouns. *Science*. 2008:320(5880): 1191–1195.

Morillon B, Schroeder CE, Wyart V. Motor contributions to the temporal precision of auditory attention. *Nat Commun*. 2014:5: 5255.

Nelson MJ, El Karoui I, Giber K, Yang X, Cohen L, Koopman H, Cash SS, Naccache L, Hale JT, Pallier C et al. Neurophysiological dynamics of phrase-structure building during sentence processing. *Proc Natl Acad Sci U S A*. 2017:114(18):E3669–E3678.

O'Connell LA, Hofmann HA. Evolution of a vertebrate social decision-making network. *Science*. 2012:336(6085):1154–1157.

Peña M, Melloni L. Brain oscillations during spoken sentence processing. *J Cogn Neurosci*. 2012:24(5):1149–1164.

Poeppel D, Assaneo MF. Speech rhythms and their neural foundations. *Nat Rev Neurosci*. 2020:21(6):322–334.

Sohoglu E, Chait M. Detecting and representing predictable structure during auditory scene analysis. *Elife*. 2016:5. https://doi.org/10.7554/eLife.19113.

Steinhauer K, Alter K, Friederici AD. Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nat Neurosci*. 1999:2(2):191–196.

Tadel F, Baillet S, Mosher JC, Pantazis D, Leahy RM. Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput Intell Neurosci*. 2011:2011:8–13.

Taulu S, Hari R. Removal of magnetoencephalographic artifacts with temporal signal-space separation: demonstration with single-trial auditory-evoked responses. *Hum Brain Mapp*. 2009:30(5): 1524–1534.

Tulving E, Schacter DL. Priming and human memory systems. *Science*. 1990:247(4940):301–306.

Zacks JM, Braver TS, Sheridan MA, Donaldson DI, Snyder AZ, Ollinger JM, Buckner RL, Raichle ME. Human brain activity time-locked to perceptual event boundaries. *Nat Neurosci*. 2001:4(6): 651–655.

Zhan W, Guo R, Chang B, Chen Y, Chen L. *The building of the CCL corpus: its design and implementation*. Beijing, China: Corpus Linguistics; 2019. pp. 71–86

Zhang W, Ding N. Time-domain analysis of neural tracking of hierarchical linguistic structures. *NeuroImage*. 2017:146:333–340.

Zhang J, Jiang C, Zhou L, Yang Y. Perception of hierarchical boundaries in music and its modulation by expertise. *Neuropsychologia*. 2016:91:490–498.