Behavioral/Cognitive

Naturalistic Spoken Language Comprehension Is Supported by Alpha and Beta Oscillations

Ioanna Zioga,[1,2] **Hugo Weissbart,**[1] **Ashley G. Lewis,**[1,2] **Saskia Haegens,**[1,3,4] **and Andrea E. Martin**[1,2]

[1]Donders Institute for Brain, Cognition and Behaviour, Centre for Cognitive Neuroimaging, Radboud University, Nijmegen, 6525 EN, The Netherlands, [2]Max Planck Institute for Psycholinguistics, Nijmegen, 6525 XD, The Netherlands, [3]Department of Psychiatry, Columbia University, New York, New York 10032, and [4]Division of Systems Neuroscience, New York State Psychiatric Institute, New York, New York 10032

Brain oscillations are prevalent in all species and are involved in numerous perceptual operations. α oscillations are thought to facilitate processing through the inhibition of task-irrelevant networks, while β oscillations are linked to the putative reactivation of content representations. Can the proposed functional role of α and β oscillations be generalized from low-level operations to higher-level cognitive processes? Here we address this question focusing on naturalistic spoken language comprehension. Twenty-two (18 female) Dutch native speakers listened to stories in Dutch and French while MEG was recorded. We used dependency parsing to identify three dependency states at each word: the number of (1) newly opened dependencies, (2) dependencies that remained open, and (3) resolved dependencies. We then constructed forward models to predict α and β power from the dependency features. Results showed that dependency features predict α and β power in language-related regions beyond low-level linguistic features. Left temporal, fundamental language regions are involved in language comprehension in α, while frontal and parietal, higher-order language regions, and motor regions are involved in β. Critically, α- and β-band dynamics seem to subserve language comprehension tapping into syntactic structure building and semantic composition by providing low-level mechanistic operations for inhibition and reactivation processes. Because of the temporal similarity of the α-β responses, their potential functional dissociation remains to be elucidated. Overall, this study sheds light on the role of α and β oscillations during naturalistic spoken language comprehension, providing evidence for the generalizability of these dynamics from perceptual to complex linguistic processes.

Key words: α oscillations; β oscillations; dependency parsing; MEG; naturalistic spoken language comprehension; TRF

Significance Statement

It remains unclear whether the proposed functional role of α and β oscillations in perceptual and motor function is generalizable to higher-level cognitive processes, such as spoken language comprehension. We found that syntactic features predict α and β power in language-related regions beyond low-level linguistic features when listening to naturalistic speech in a known language. We offer experimental findings that integrate a neuroscientific framework on the role of brain oscillations as "building blocks" with spoken language comprehension. This supports the view of a domain-general role of oscillations across the hierarchy of cognitive functions, from low-level sensory operations to abstract linguistic processes.

Introduction

Out of the many neural phenomena that exist, brain oscillations are prevalent in all species and are involved in numerous perceptual operations. Are brain oscillations the "building blocks" of cognitive function, from low-level sensory to higher-level processes? Extensive prior research focused on the role of α and β oscillations in basic perceptual and motor functions. Here, we asked whether their proposed role in low-level operations generalizes to higher-level cognitive functions, in particular language comprehension. We adopted a forward-modeling approach predicting brain responses from high-level, syntactic features to test the role of α and β oscillations during naturalistic spoken language comprehension.

Alpha oscillations (8-12 Hz) are thought to reflect a mechanism of active inhibition, which fine-tunes sensory processing by

guiding attention and suppressing distracting input (Klimesch et al., 2007; Jensen and Mazaheri, 2010). α power increases with working memory load during retention, reflecting inhibition of task-irrelevant regions (Jensen et al., 2002; Tuladhar et al., 2007; Scheeringa et al., 2009; Pan et al., 2018), and is associated with behavioral performance (Haegens et al., 2010, 2011b, 2012). Traditionally, β oscillations (15-30 Hz) are considered a motor rhythm (Kilavik et al., 2013), associated with top-down processing and long-distance network communication (Varela et al., 2001). β power increases during information retention, attributed to active maintenance of the current cognitive set (Engel and Fries, 2010). Spitzer and Haegens (2017) proposed that β oscillations support the reactivation of content representations, via the transitioning of latent items into active working memory (Rose et al., 2016). To date, there is extensive research investigating the oscillatory correlates of language processing (Bastiaansen et al., 2010; Obleser and Weisz, 2012; Lewis et al., 2015; Zoefel and VanRullen, 2015; Ding et al., 2016; Kösem et al., 2016; Martin and Doumas, 2017; L. Meyer, 2018; Brennan and Martin, 2020; Kaufeld et al., 2020; L. Meyer et al., 2020; van Bree et al., 2021; Bai et al., 2022; Coopmans et al., 2022; Hauswald et al., 2022; ten Oever et al., 2022a). However, few steps have been taken to begin to link investigations of α and β oscillations in sensory and perceptual neuroscience to cognitive functions that are fundamentally derivative of sensory processing, such as language comprehension from speech or sign perception.

Here, we operationalized high-level linguistic processing using attributes from dependency parsing, describing syntactic sentence structure as relations between pairs of words (Mel'cuk, 1988; Tesnière, 2015). Dependencies are created when a nonunified word ("dependent constituent") is encountered. Processing load increases with the number of dependencies being processed (Vos et al., 2001; Demberg and Keller, 2008). Dependencies are resolved once the linking word ("dependent") is encountered, recruiting unification or integration processes (Hagoort, 2005; Martin, 2016, 2020; Kapteijns and Hintz, 2021). Integration is thought to require reactivation of any dependent constituent (McElree et al., 2003; Martin and McElree, 2008; Foraker and McElree, 2011). As resolving linguistic dependencies is crucial for language comprehension, it can be argued that dependency resolution exemplifies higher-level cognitive operations. We therefore hypothesized that α and β power would be modulated by dependencies, and used dependency parsing as a proxy for this processing.

Specifically, we compared MEG responses while participants listened to comprehended (Dutch) versus uncomprehended (French) spoken stories. We identified three states at each word: number of opened/remained open/resolved dependencies. We constructed forward models to predict α and β power from these dependency features, controlling for low-level linguistic features. We predicted that Dutch compared with French stories would elicit stronger modulations of α and β power in typical language-processing brain regions. Additionally, we hypothesized that our dependency features would predict α and β power in language processing-related regions beyond low-level linguistic features. We further hypothesized that the opening of dependencies would be associated with α increases/decreases at task-irrelevant/-relevant areas linked to inhibition, while "maintenance" of dependencies would be associated with β power increases, attributed to anticipatory and active ongoing processes. Finally, we predicted that β power would increase during dependency resolution, related to content reactivation.

Materials and Methods

Participants

Twenty-two adults (18 female) aged between 18 and 63 years old (mean $\pm$ SD age, 34 $\pm$ 15 years) took part in the experiment. Prescreening required that participants were monolingual Dutch native speakers, right-handed, without hearing problems, reading problems, or epilepsy. All participants self-reported zero use and minimal understanding of French at the level of isolated words but not whole sentences before taking part in the experiment. Before the experiment, participants were provided with written and verbal information about the MEG system and safety regulations and gave written informed consent. They received monetary reimbursement after participation. The study falls under the general ethics approval (CMO 2014/288 "Imaging Human Cognition") in accordance with the Declaration of Helsinki.

Stories

In order to tap into language comprehension, we compared brain responses recorded with MEG while participants listened to spoken stories in a language they comprehend (Dutch, mother tongue) versus a language they do not comprehend (French, a familiar but uncomprehended language). Stories in French were selected as a control to confirm that our effects are because of comprehension and not acoustic properties of speech, with which participants would be familiar because of regional proximity. Behavioral performance and debriefing demonstrated that Dutch native speakers were not able to understand the French narratives, despite their familiarity with the acoustic properties and some common words in French. French thus constituted a stronger control than a language with which participants would be completely unfamiliar. Critically, compared with traditional studies using artificial word or sentence stimuli, the use of natural speech in prerecorded stories allowed for a more ecologically valid approach as (1) the natural prosody of the voice recording guides comprehension via auditory cues, (2) processing requires constant effortful attention throughout, and (3) it lacks the brain responses induced by certain properties of artificial stimuli, such as abrupt voice modulations or unnatural syllable timing.

The following three Dutch (NL) stories were used: *Het Lelijke Jonge Eendje* by H.C. Andersen, *De Ransel, het Hoedje en het Hoorntje* and *De Gouden Vogel* by the Grimm brothers. All NL stories were spoken by female voices. The following three French (FR) stories were used: *L'eau de la vie* by the Grimm brothers (male voice), *L'ange* by H.C. Andersen (female voice), and an excerpt from *Le Canard Ballon* by E.A. Poe (female voice). The NL stories and the last FR story were retrieved from www.librivox.org, and the rest from www.litteratureaudio.com. In order to reduce fatigue, stories were split into parts of short duration (NL: 9 story parts, mean $\pm$ SD, 5.5 $\pm$ 0.6 min; FR: 4 story parts, duration 5.3 $\pm$ 0.7 min). Stories that were already <6 min were not split further. All audio files were normalized to an equal perceived loudness.

Five MCQs with four choices each were included after each story part (65 in total) to (1) ensure that participants paid attention to the spoken stories and (2) confirm the lack of understanding of the French stories. A Dutch and a French native speaker composed the questions for the Dutch and French stories, respectively. All were content questions, for example: *Who lives in the old house? A. An old man B. An old lady C. Nobody D. A family; What did the traveler take from the table? A. The tablecloth B. The bread C. The potatoes D. The wine.*

Procedure

Participants were seated in the MEG system in a dimly lit room. They were informed that they would listen to stories in Dutch and French during MEG recording. Further, they were instructed to pay attention to the stories as they would be prompted to answer MCQs after each story part. Responding to the MCQs was done by pressing four keys of a response box in a self-paced manner. Resting-state MEG was recorded for 10 s before the onset of each story part but was not included in the analysis. The presentation order of the story parts was pseudorandomized across participants: NL and FR story parts were interleaved but care was taken so that their order remained intact (e.g., the second part of a story could be presented only if the first part of that same story was

previously heard). The overall procedure lasted for ~1.5 h. The experiment was programmed with custom MATLAB (The MathWorks) scripts using Psychtoolbox (Brainard and Vision, 1997).

Data acquisition and MEG preprocessing

MEG data were recorded at a sampling rate of 1200 Hz using a 275 channel axial gradiometer system (CTF MEG systems, VSM MedTech) located in a magnetically shielded room. Eight sensors were excluded because of permanent malfunction, leaving a total of 267 usable sensors. Three fiducial localization coils were placed at the participant's nasion and left and right ear canals to (1) allow for real-time monitoring of the participant's head position and adjustment in between story parts if necessary, and (2) provide anatomic landmarks for offline coregistration of the MEG data with T1-weighted MRI images for source reconstruction. After completion of the task, the x, y, z coordinates of the three fiducial points as well as the participant's head shape were digitized using a Polhemus 3D tracking device. Furthermore, individual structural MRI scans were acquired in a 3T Siemens Magnetom Skyra MR scanner using earplugs with a drop of vitamin E at the subject's ear canals to facilitate subsequent alignment with the MEG data.

Continuous MEG data were downsampled to 100 Hz and epoched from the onset until the offset of each story part. Data from sensors with consistently poor signal quality, as observed by visual inspection, were removed and interpolated based on neighboring sensors. Finally, independent component analysis was performed to correct for eye-blinks and heartbeat artifacts. Custom-written scripts in MATLAB and the FieldTrip toolbox (Oostenveld et al., 2011) were used for analysis of the MEG data.

Data analysis

Behavioral data

To assess participants' understanding of the stories, we calculated the percentage of correct responses in the MCQs separately for NL and FR stories. A paired t test was used to compare the two conditions and a one-sample t test to compare performance accuracy to chance level at 25%.

Source reconstruction of MEG data

MRI preprocessing. First, coregistration of the MRI with the CTF and Polhemus fiducials was performed. Individual MRIs were normalized in MNI space and segmented. Realistic volume conduction models were created for each participant based on the single-shell model of their MRIs (Nolte, 2003). For each participant, 5798 dipole positions were defined with an 8 mm resolution.

Spatial filters. A spatial filter for the source reconstruction analysis was calculated for each participant. Covariance matrices were computed over single trials (13 Dutch and French story parts) and then averaged. Leadfields for all grid points, combined with the covariance matrices, were used to compute a spatial filter with the Linearly Constrained Minimum Variance (Veen et al., 1997) method. The source orientation was fixed to the dipole orientation with the highest strength.

Forward models predicting α and β power from linguistic features

We attempted to quantify higher-level operations during spoken language comprehension in response to the processing of dependencies that opened/remained open/were resolved at each word. We then constructed forward models predicting α and β power from the dependency features controlling for low-level linguistic features (acoustic edges, word onset, and word frequency).

To investigate the relationship between linguistic features and α and β power, we constructed a time series for each feature. Each word feature was time-aligned with the auditory stimulus using the forced-alignment function of the web-service MAUS (Kisler et al., 2017). In order to align the linguistic features with the auditory stimuli, a single impulse-like value representing the magnitude of the feature was assigned at the onset of each word (except for acoustic edges where the impulses could be at different time points, see below).

Dependency features.

Dependency parsing.
We operationalized high-level linguistic processing during spoken language comprehension using attributes from dependency parses. This is mainly motivated by the trade-off between coverage of features and accessibility in parsing models; it is not a strong theoretical commitment to one parsing framework over another. Dependency grammars describe the syntactic structure of a sentence as a set of relations between two words (Mel'cuk, 1988; Tesnière, 2015). The links begin from the head and end on the dependent word and are assigned a label representing the type of dependency (e.g., subject, object, determinant, etc.). Each sentence has a root, usually the verb, which is the head of the entire structure (for an example of dependency parsing, see the graph in Fig. 1A, top). Dependency grammars often reveal nonadjacent, complex dependencies. Previous work has used dependency structures as a measure of or proxy for syntactic complexity, as words that form dependencies often appear in nonadjacent positions (e.g., Wilson et al., 2020).

We used an automated parser (Stanford parser "Stanza") (Qi et al., 2020) to generate dependency graphs for each sentence in the stories. Stanza uses universal dependencies (Nivre et al., 2016), which is a set of dependency relations that are cross-linguistically applicable (for the types of universal dependencies in our stories, see Table 1). Based on those, three dependency measures were extracted for each word using custom-written scripts: (1) number of opened dependencies, (2) number of dependencies that remained open, and (3) number of resolved dependencies. As we did not have any hypothesis about left- versus right-branching dependencies, we summed over both directions (Fig. 1A). Dependency features were represented as valued impulses at the word onsets of the respective words where the dependency took place.

(1) Opened dependencies: the number of dependencies that open at a given word. In the example of Figure 1A, one dependency opens at each word (nsubj, obj, det, amod, respectively) except for the last one (zero opened dependencies).

(2) Remained open dependencies: the number of dependencies that are already open but remain unresolved. In Figure 1A, the obj dependency is still unresolved at word "the," while both the obj and det dependencies are unresolved at word "big."

(3) Resolved dependencies: the number of dependencies that are resolved at this word. In Figure 1A, the nsubj relation is resolved at word "opened"; the rest of the dependencies are resolved at the last word "presents."

As mentioned earlier, we were interested in investigating high-level cognitive processing associated with comprehension. Content words (nouns, verbs, adjectives, adverbs) are known to have a lexical semantic content, whereas function words (pronouns, articles, prepositions, auxiliary verbs) contribute mostly to the grammatical structure and have a relatively less lexical semantic meaning (Corver and van Riemsdijk, 2013), although they clearly have consequences for syntactic and semantic compositional meaning. Therefore, we focused our analysis on dependency relations between content words only. This was done by using relations that were comprised of two content words, while excluding relations in which at least one of the two words was a function word. For instance, in Figure 1A, three dependency relations are resolved at the last word of the sentence "presents": "opened presents," "the presents," and "big presents." However, only two of those relations are comprised of two content words ("opened presents" and "big presents"), as the relation "the presents" contains a function word, the article "the." By subtracting the number of relations containing function words, we are left with two resolved dependencies instead of three. This process is performed separately for the construction of each of the three dependency features.

Low-level features: control variables. To make sure that potential dependency effects are not because of low-level linguistic properties, we considered the following features as our base model:

Acoustic edges.
Abrupt changes in the acoustics are tracked by neural activity. It is possible that, through cross-frequency coupling, the neural tracking of syllables in low frequencies (δ-θ band) (Doelling et al., 2014) modulates the α and β frequency bands. Notably, it is still a matter of debate whether the tracking of low frequencies is implemented as endogenous oscillations or as a series of evoked responses to acoustic landmarks (Kojima et al., 2020). We thus
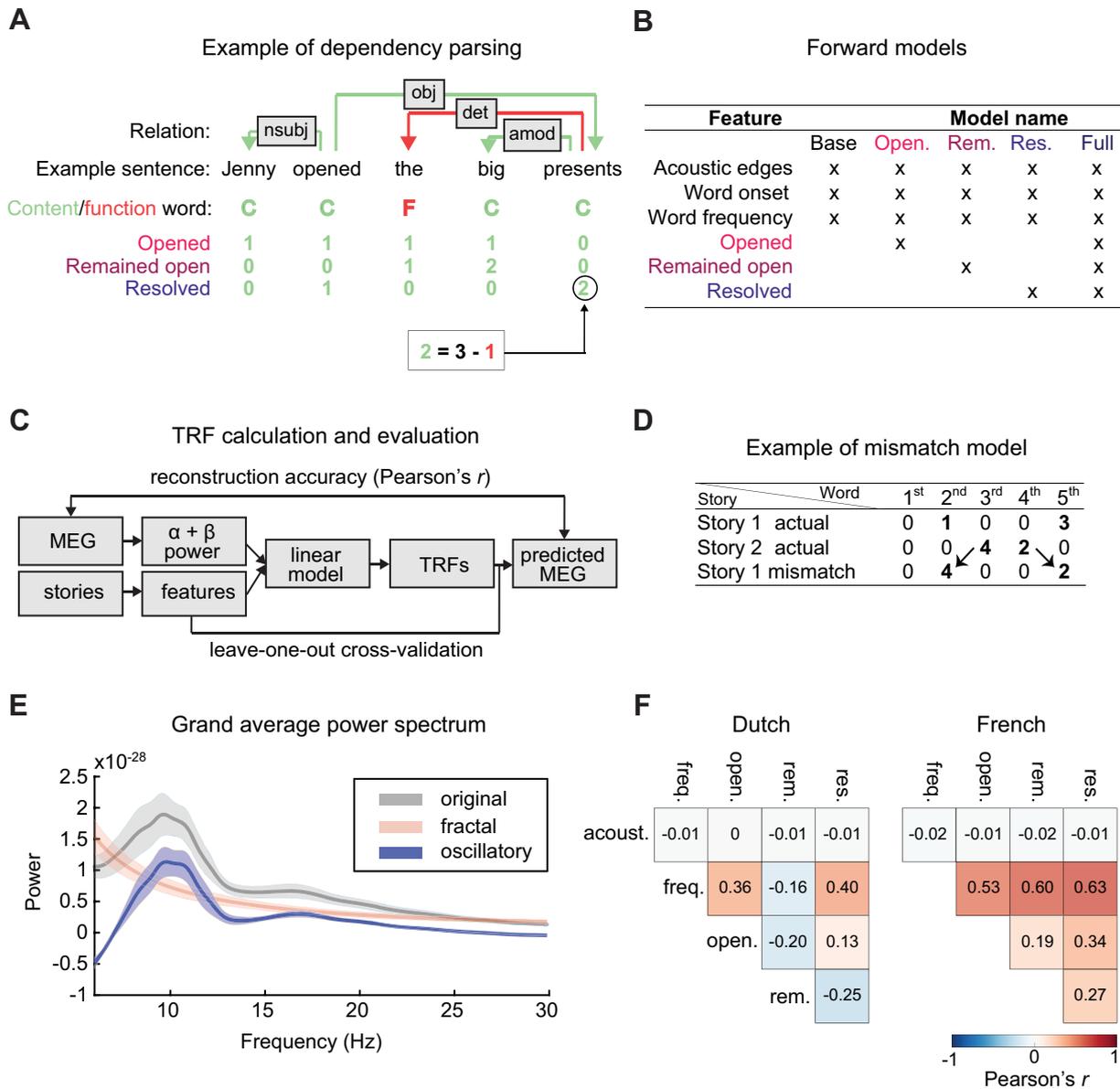
Figure 1. Methodological aspects of the TRF analysis using linguistic features as predictors for α and β power during naturalistic story listening. *A*, Example of dependency parsing (Mel'cuk, 1988) and the extracted dependency features (number of opened/remained open/resolved dependencies). The automated Stanford parser "Stanza" (Qi et al., 2020) generated dependency graphs for each sentence. Green arrows represent relations between two content words (nouns, verbs, adjectives, adverbs). Red arrows represent relations containing at least one function word (pronouns, articles, prepositions, auxiliary verbs). The dependency features were constructed based on relations comprised of two content words (green values), excluding relations containing function words (red values). *B*, Model construction. The base model includes low-level linguistic features (acoustic edges, word onset, word frequency), which are included in the dependency models. *C*, Schematic of the TRF analysis pipeline. *D*, Example of mismatch model construction. The actual feature values are replaced by those of another story while keeping the initial positions. *E*, Grand average power spectrum over all data, participants, and sensors. The FOOOF algorithm (Donoghue et al., 2020) was used to separate the fractal from the oscillatory components of the original signal. *F*, Correlation plots between all features (except word onset as it is a constant) separately for NL and FR stories.

controlled for low-level acoustic properties by incorporating acoustic edges in our feature set, extracted from the speech envelope. First, we generated broadband envelopes of the audio files using gammatone filter banks (method following Fishbach et al., 2001). Then, we calculated the derivative of the envelope and defined as acoustic edges the points when the derivative exceeds its 97.5th percentile. Acoustic edges were represented as nonzero, equally valued pulses.

Word onset. Neural activity has been found to track the onset of words because of the brain's parsing of the acoustic input to form discrete meaningful units (Ding and Simon, 2014). Word onsets were represented as nonzero, equally valued pulses at the time points defined by the forced alignment procedure.

Word frequency. The frequency of a word outside the sentential context has been shown to modulate neural responses (Brodbeck et al., 2018). Two online databases were used to calculate word frequency,

SUBTLEX-NL for Dutch (Keuleers et al., 2010) and Lexique for French (New et al., 2004). The number of instances of each word was divided by the total number of instances of all words. Word frequency was defined as the negative logarithm of that number, so that the higher the value, the lower the frequency.

Before using the above features in the linear regression analysis, we examined the feature inter-correlations by calculating Pearson's r coefficient between features. All correlations were low to moderate (Fig. 1*F*). To detect multicollinearity between features, the variance inflation factor was computed, which indicates whether the variation of one feature is largely explained by a linear combination of the other features. Variance inflation factor was low for all features ($NL_{ac\_edges} = 1$, $NL_{freq} = 1.34$, $NL_{opened} = 1.18$, $NL_{remained\_open} = 1.10$, $NL_{resolved} = 1.24$; $FR_{ac\_edges} = 1$, $FR_{freq} = 3.09$, $FR_{opened} = 1.44$, $FR_{remained\_open} = 1.66$, $FR_{resolved} = 1.70$), indicating no concern for multicollinearity (for feature descriptives, see

Table 1. Universal dependency types in NL and FR stories[a]

Universal dependency type	NL	FR
Adverbial modifier	13.72	8.73
Nominal subject	10.63	8.53
Conjunct	9.25	8.03
Oblique nominal	8.72	9.90
Object	7.21	9.23
Verb	7.04	5.47
Determiner	6.34	9.34
Adjectival modifier	5.41	5.88
Case marking	4.78	6.19
Coordinating conjunction	4.12	3.34
Parataxis	3.00	1.04
Nominal modifier	2.81	5.84
Adverbial clause modifier	2.76	2.63
Open clausal complement	2.73	4.08
Marker	2.58	2.04
Auxiliary	2.09	1.45

[a]Percentage of trials belonging in each type of Universal Dependency Relation (for more information, see https://universaldependencies.org/u/dep/index.html), separately for NL and FR stories, showing only types occurring >2% of the time.

Table 2. Descriptive statistics of features used in the models[a]

	NL			FR		
Feature	Mean	SD	Range	Mean	SD	Range
Word frequency	3.13	1.39	1.39-7.64	3.36	1.37	1.45-8.26
Opened	1.29	0.66	1-5	1.33	0.62	1-5
Remained open	1.70	0.86	1-7	1.53	0.74	1-5
Resolved	1.46	0.79	1-6	1.20	0.47	1-4

[a]Mean, SD, and range of word frequency, opened, remained open, and resolved features. Acoustic edges and word onset are not presented as they have a single value.

Table 2). Finally, for each linguistic feature, values were standardized to have unit variance and zero mean.

α and β power estimation
Following our hypotheses focusing on genuine brain oscillations, we used spectral analysis of the MEG data to confirm the presence of two distinct peaks separately for α and β, as an index of oscillatory activity. Welch's method was used to compute the power spectra. Subsequently, the Fitting Oscillations & One Over F (FOOOF) algorithm (Donoghue et al., 2020) was applied to confirm the presence of peaks with power over and above the aperiodic 1/f signal (Fig. 1E).

Sensor-level. The time course of the α (mean of 8–12 Hz) and β (15–30 Hz) power was estimated throughout all story parts. Preprocessed MEG data were convolved with a sliding window Hanning taper (adaptive window length). The time-frequency representation was calculated with 1 Hz steps using 6-cycle wavelets over the course of each story part. Then, the wavelet convolved values were averaged over the frequency band.

Source-level. First, the complex Fourier coefficients were estimated separately for the α and β bands with same parameters as in the sensor-level analysis. Then, the coefficients were multiplied with the participant's spatial filter, and, finally, the power of that product was calculated.

Finally, the spectral data were normalized by subtracting the mean α and β power over all time points of all stories from each time point, separately for each sensor/source, before estimation of the temporal response functions (TRFs).

TRF analysis
We constructed linear forward models (TRFs) (Crosse et al., 2016) to predict α and β power from these dependency features, controlling for low-level linguistic features (Fig. 1C). TRF analysis is capable of disentangling overlapping neural responses because of consecutive events with high temporal proximity, and can handle confounding covariates. TRFs are forward or encoding models based on the assumption that the output of a system relates to the input via a linear convolution (Ding and

Simon, 2012; Broderick et al., 2018). Here we assume that the neural responses (α and β power) at each sensor can be expressed as a linear combination of linguistic features shifted by different latencies (for a schematic of the TRF analysis pipeline, see Fig. 1C). Specifically, the instantaneous MEG response $r(t, n)$ of times $t = 1...T$ at channel n is expressed by the convolution of a linguistic feature, $s_k(t)$, with a kernel or TRF, $w_k(\tau, n)$. The TRF covers a specified range of time lags, τ, relative to time t. The forward model can be represented by the following equation:

$$r(t, n) = \sum_k \sum_\tau w_k(\tau, n) s_k(t - \tau) + \varepsilon(t, n) \tag{1}$$

where $\varepsilon(t, n)$ is a white noise process, capturing part of the signal unrelated to the stimulus. Contributions from each feature are linearly combined. The TRF is estimated by minimizing the mean-squared error between the MEG response, $r(t, n)$, and the predicted MEG response, $\hat{r}(t, n)$ as follows:

$$\operatorname{argmin}_w \varepsilon(t, n) = \sum_t \left[r(t, n) - \hat{r}(t, n) \right]^2 \tag{2}$$

The solution to (2) can be computed in closed-form using the pseudo inverse as follows:

$$w = (S^T S)^{-1} S^T r \tag{3}$$

where S is the concatenation of the lagged time series of each linguistic feature, s_k.

TRF analysis was conducted using the MATLAB mTRF Toolbox (Crosse et al., 2016). Here we used the function *mTRFtrain* to estimate the TRF coefficients for each linguistic feature, separately for NL and FR stories. By visual inspection of the TRF coefficients, the time lags over which TRFs were analyzed were from −1 to 1.5 s. The TRF at time t indexes how a unit change in a given linguistic feature affects the MEG response t seconds later. For Ridge regression, a regularization term is added to leverage the fact that the inversion of $S^T S$ is unstable, and thus prevent overfitting because of fitting high-frequency noise. This happens when the columns of S are correlated. With continuous regressors, the lagged time series forming the columns of S comprise a highly autocorrelated signal. However, in our case, all columns of the lagged time series are independent, as they are not continuous and contain nonzero values only at word onsets. The lagged time series is thus not correlated, hence adding a regularization term was not necessary and would lead to underfitting.

Model validation. Validation of the TRF models was performed by comparing the Pearson's r correlation between the actual MEG and the reconstructed MEG response. This was implemented using the function *mTRFcrossval* of the mTRF Toolbox following a leave-one-out cross-validation approach. Specifically, a story part was used as the test set and the remaining M-1 story parts were used as the training set. The TRF model was then estimated for each story part of the training set, and their average TRF is computed. Subsequently, the averaged model was convolved with the test set to predict the MEG responses. Pearson's r was computed between the actual MEG and the reconstructed MEG responses of the test set from −1 to 1.5 s. The aforementioned process was repeated M times, so that all story parts were assigned to the test set once. The Pearson's r values were then averaged over all M validations. This procedure was done separately for NL and FR stories.

Statistical evaluation
Model comparison
We first assessed the contribution of the low-level linguistic features to α and β power modulations by evaluating their reconstruction accuracy. Results showed that all features (acoustic edges, word onset, word frequency) explain a substantial amount of variance of the MEG response (Fig. 2A). As there is evidence that word surprisal affects neural responses (Weissbart et al., 2020), we also evaluated the contribution of surprisal. Surprisal values were estimated from the GPT2 language model (from huggingface transformer model,

available at https://huggingface.co/GroNLP/gpt2-small-dutch) (de Vries and Nissim, 2020). Results showed that (1) surprisal did not explain a substantial amount of variance, neither in the α nor in the β band (i.e., adding surprisal to a model with acoustic edges and word onset or to a model with acoustic edges, word onset and word frequency did not significantly improve reconstruction accuracy of the model); and (2) word frequency and surprisal are highly correlated (Pearson's $r = 0.574$, $p < 0.001$); therefore, we decided to include only acoustic edges, word onset, and word frequency in the base model. The respective model coefficients of the low-level features are shown in Figure 2C.

Comparison with base model. In order to test whether the dependency features predict the neural data over and beyond the base model, we compared reconstruction accuracy between the base model (including only the low-level linguistic features) against the base model augmented with each and all of the dependency features (opened/remained open/resolved/full) (for model construction, see Fig. 1B).

As dependencies do not occur at every word instance, dependency features had a substantially lower number of nonzero values ("trials" from now on) compared with the base features (base features: $N_{word\_onset, word\_frequency} = 8535$; dependency features: $N_{opened} = 2691$; $N_{remained\_open} = 5596$; $N_{resolved} = 2381$). This would affect the signal-to-noise ratio in the estimated TRF and, therefore, the associated reconstruction accuracy. Because of the different number of trials between features, we needed to equalize the number of trials of the features of the two contrasting models at each time (i.e., a dependency model vs the base model) by randomly selecting an equal number of trials across features. To make sure this random selection was not particular in any way, we followed a bootstrapping procedure. More specifically, the feature with the smallest number of trials was first identified. Then, an equal number of trials was randomly selected in the rest of the features of the two models being compared at that time, while the excessive trials were converted to zero. This was performed for every feature except acoustic edges, as those trials were not aligned with word onset and were therefore relatively independent to the rest of the features. To make sure that our effects would not be because of a certain random selection during bootstrapping, we repeated this procedure 10,000 times. Subsequently, we tested in how many of these iterations reconstruction accuracy of the dependency model was significantly higher than the base model. To do this, reconstruction accuracy was first averaged over sensors exhibiting improved accuracy over the base model (i.e., where the z-scored difference between models exceeded 1 SD). Then, paired t tests were conducted between models. Results showed that all models significantly improved reconstruction accuracy compared with the base model across the 10,000 iterations (percentage of significant iterations >95%) of the bootstrapping procedure (for the t value distributions, see Fig. 2B). Therefore, reconstruction accuracy was averaged over all iterations to perform the final statistical evaluation. Reconstruction accuracy was then averaged over the sensors of which the z score difference exceeded 1 SD in >50% of the iterations.

Here is a summary of the bootstrapping pipeline for dependency model versus base model comparisons:

1. Among the two contrasting models, find the feature with the smallest number of trials ("trials" defined as nonzero feature values).
2. For all features, randomly select an equal number of trials and set the remaining trials to zero.
3. Compare average reconstruction accuracy between the two models.
4. Repeat Steps 2 and 3 for 10,000 iterations.
5. Calculate the percentage of times Step 3 was significant and compute improvement by subtracting the average reconstruction accuracy of the base model from the dependency model.

We performed a 3 (model: opened/remained open/resolved) × 2 (frequency band: α vs β) repeated-measures ANOVA with reconstruction accuracy improvement (dependency model – base model) as the dependent variable. Improvement was also compared to zero with one-sample t tests, for each model. As results showed that all three dependency features were significant, the full model was evaluated in a 4 (model: opened/remained open/resolved/full) × 2 (frequency band: α vs β) repeated-measures ANOVA, and was compared to zero. All *post hoc* contrasts were Bonferroni-corrected for multiple comparisons.

Comparison with mismatch model. In order to perform comparisons to chance levels of reconstruction accuracy, we constructed null models to which the full dependency model was compared. To confirm that reconstruction accuracy improvement with the dependency features is not merely because of (1) the addition of features or (2) the existence or not of a dependency state independent of its value, we compared the full model with mismatch models. Mismatch models are models of which the feature values of one of the features is replaced by those from another story, while keeping the value positions of the actual story (Fig. 1D). This allows to compare models with matching number of predictors.

As the mismatch models have the same number of trials per feature with the actual models, there was no need for a bootstrapping approach here. We performed a 4 (model: opened/remained open/resolved/full) × 2 (frequency band: α vs β) repeated-measures ANOVA with reconstruction accuracy improvement as the dependent variable, averaged over sensors with z-scored difference between mismatch – actual exceeding 1 SD. One-sample t tests compared reconstruction accuracy improvement from zero.

Control analysis: comparison with mismatch model in French stories. As participants did not understand French, we used the French stories as a control condition to confirm that the α and β power modulations by the dependency features in Dutch is linked to comprehension rather than acoustic or speech properties. Similar to the above analysis, we performed one-sample t tests (compare to zero) and a 3 × 2 ANOVA on the reconstruction accuracy difference between actual versus mismatch models, averaged over the identified sensors with maximal improvement.

Reconstruction accuracy between NL versus FR stories

Considering the multiple comparisons problem and the lack of a specific hypothesis about the location of the effects, we used a nonparametric cluster permutation approach (Maris and Oostenveld, 2007) to compare the reconstruction accuracy in NL versus FR on source level. As there were nine story parts in NL, but only four in FR, we performed a bootstrapping procedure with replacement by randomly selecting four of the NL story parts over which we compared the reconstruction accuracy with the FR. This was done over 50 iterations, all showing significant differences between conditions. Reconstruction accuracy was then averaged over all iterations to perform the final statistical evaluation. The cluster permutation procedure addresses the multiple comparison problem by combining neighboring source points that show the same effect into clusters and comparing those with the null distribution. Paired samples t tests were computed for each source point, testing NL versus FR conditions. Spatially adjacent source points whose t values exceeded an *a priori* threshold (uncorrected p value < 0.05) were combined into the same cluster, with the cluster-level statistic calculated as the sum of the t values of the cluster. Finally, the values of the cluster-level statistic were evaluated by calculating the probability that it would be observed under the assumption that the two compared conditions are not significantly different ($\alpha = 0.05$, two-tailed). To obtain a null distribution to evaluate the statistic of the actual data, values were randomly assigned to the two conditions and the statistics recomputed 1,000 times (Monte-Carlo permutation).

α and β band modulations by dependency features

Comparison of dependency models with base model: reconstruction accuracy. To identify the neural sources of the dependency feature contributions, we used beamformer source reconstruction in the α and β bands (see Materials and Methods sections "Source reconstruction of MEG data" and "α and β power estimation"). The reconstruction accuracy of each dependency model was compared with the base model. A bootstrapping approach with replacement was used as described in section "Model comparison", so that all features of the two models under comparison had the same number of trials. Reconstruction accuracy at each source point was averaged over iterations. A cluster-based permutation approach was used for statistical evaluation as described in Reconstruction accuracy between NL versus FR stories ($\alpha = 0.0001$, two-tailed; 1,000 permutations), by shuffling the labels between dependency and base
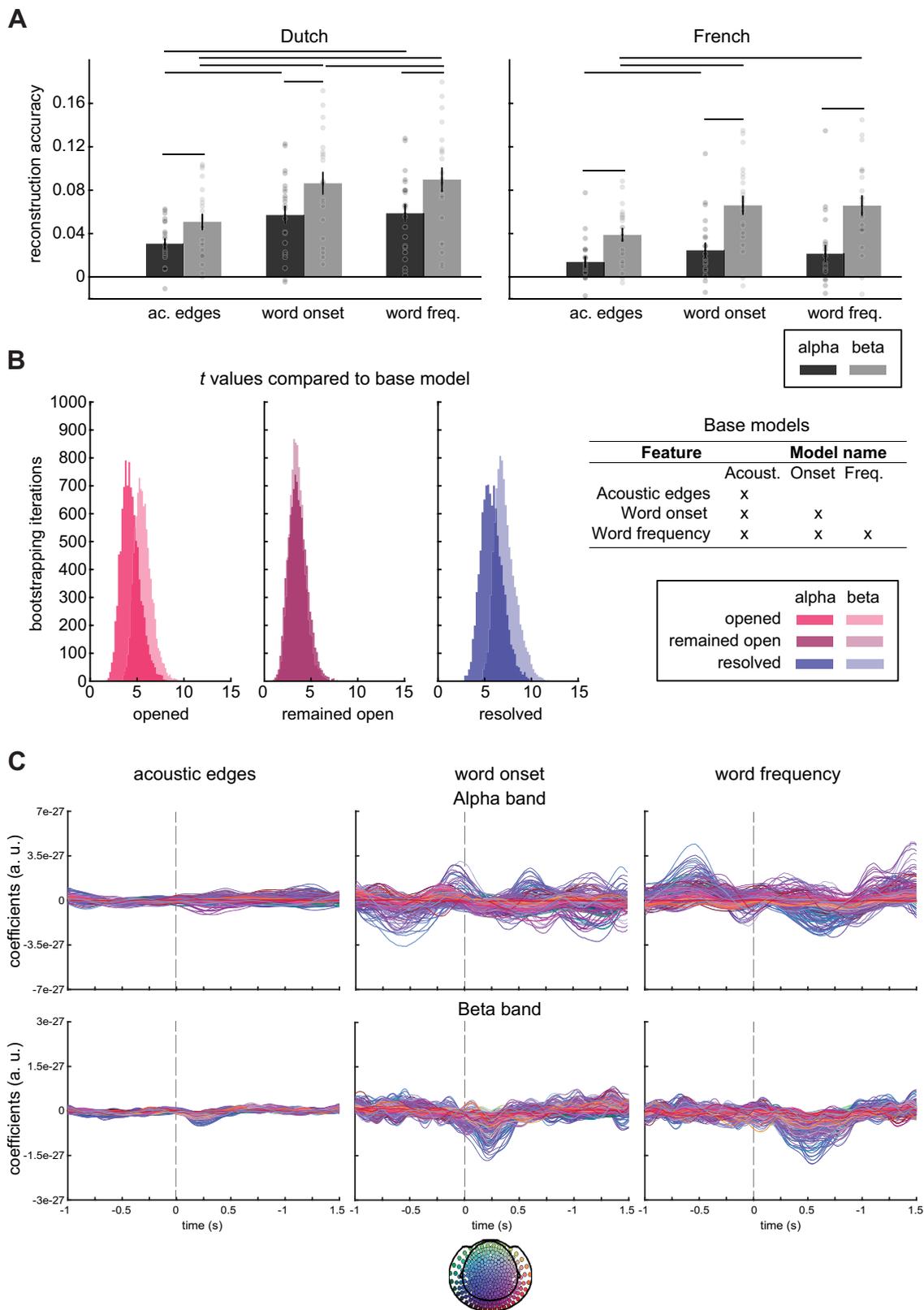
Figure 2. Evaluation of features based on reconstruction accuracy (Pearson's *r*). **A**, Evaluation of the effect of speech features of the base models (see Table 2 at the right side in Fig. 2*B*.): reconstruction accuracy of each model (averaged over sensors >1 SD) separately for the α (black) and β (gray) bands, and for NL (left) and FR (right) stories. Error bars indicate ±1 SEM. Horizontal lines indicate statistical significance ($p < 0.005$). **B**, Histograms of *t* values between the reconstruction accuracy of the models opened/remained open/resolved versus the base model over 10,000 iterations with random trial subselection (for details on the bootstrapping procedure, see Comparison with base model). **C**, Time courses of the TRF coefficients of all sensors for each low-level feature (acoustic edges, word onset, and word frequency), separately for the α and β band. The colors of the sensors are denoted in the topography at the bottom.

model. This analysis was performed separately for the α and β bands.

TRF coefficients. In order to define the contribution of the dependency features in time, the TRF coefficients were analyzed. Specifically, the coefficients of each dependency feature were averaged over significant sources, as identified from the analysis above; therefore, the comparison was done over the temporal, but not spatial, dimension. Subsequently, paired-sample t tests were performed to identify the time instances at which the TRF coefficients were significantly different from those of a mismatch model (feature values replaced by those from another story) (FDR-corrected at $p = 0.05$).

The Brainnetome Atlas (Fan et al., 2016) was used to identify the regions (parcel labels) where the effects were found. According to this atlas, each hemisphere is divided into 123 parcels, while the parcellation is based on both structural and functional connectivity.

Control analysis in the δ, θ, and γ band. To test the frequency specificity of our effects, we constructed similar TRF models in other frequency bands (δ, θ, and γ). Specifically, we filtered the signal in the δ band (bandpass at 0.5–4 Hz, filter order 206) and θ band (bandpass at 4–6.5 Hz, filter order 206), to get the respective phase-locked brain responses. Furthermore, the time course of θ (4–6.5 Hz) and γ power (40–80 Hz) was estimated by convolving the signal with a sliding window Hanning taper (adaptive window length) as described in Sensor-level. We note that it is questionable whether activity in the γ band here is truly oscillatory, however, it was analyzed for completeness. A nonparametric cluster-based permutation procedure (described in Reconstruction accuracy between NL versus FR stories; here $\alpha = 0.05$, two-tailed; 1,000 permutations) was performed at the temporal and spatial dimension between the TRF coefficients of the actual versus the mismatch models, by shuffling the labels between dependency and mismatch models. This analysis was performed separately for each dependency feature and frequency band.

Data availability. Data and code used in the main analyses are available from the corresponding author on reasonable request.

Results

Comprehension of Dutch but not French stories
To confirm that participants paid attention to and understood the NL stories but not the FR stories, we calculated performance accuracy as the percentage of correct answers in the multiple-choice comprehension questions. Results revealed that participants comprehended NL stories (mean = 89.0, SD = 6.3) significantly better than FR stories (mean = 25.7, SD = 10.7) ($t(21) = 23.907$, $p < 0.001$, Cohen's $d = 7.189$). Performance was significantly higher than chance level (25%) only for NL ($t(21) = 47.337$, $p < 0.001$, Cohen's $d = 10.092$), but not for FR ($t(21) = 0.298$, $p = 0.768$, Cohen's $d = 0.063$).

Dependency features predict α and β power beyond low-level linguistic features
Dependency features improve reconstruction accuracy from base model
To evaluate whether each of the dependency features explains variance of α and β power over and beyond the base model (i. e., based on acoustic edges, word onset, word frequency), we compared the reconstruction accuracy (Pearson's r) improvement by adding opened/remained open/resolved/all features to the base model (Fig. 3A). Reconstruction accuracy was averaged over sensors exhibiting improved accuracy over the base model (z-scored improvement > 1 SD). A bootstrapping method with replacement was used to ensure equal number of trials between features of each model (for more details on the bootstrapping method, see section "Model comparison").

Reconstruction accuracy improvement was significantly higher than zero in all models and bands (α-opened: $t(21) = 2.915$, $p = 0.008$; remained open: $t(21) = 4.903$, $p < 0.001$; resolved: $t(21) = 5.889$, $p < 0.001$; β-opened: $t(21) = 6.205$, $p < 0.001$; remained open: $t(21) = 5.746$, $p < 0.001$; resolved: $t(21) = 7.418$, $p < 0.001$).

A 3 (model: opened/remained open/resolved) $\times$ 2 (band: α/β) repeated-measures ANOVA revealed a significant *model $\times$ band* interaction ($F(2,42) = 3.254$, $p = 0.049$, $\eta^2 = 0.134$). Planned contrasts in the β band were significant: resolved was higher than opened ($t(21) = 4.465$, $p < 0.001$, Cohen's $d = 1.030$) and remained open ($t(21) = 6.170$, $p = 0.001$, Cohen's $d = 1.708$), and opened was higher than remained open ($t(21) = 4.001$, $p = 0.001$, Cohen's $d = 0.984$). In the α band, resolved was higher than remained open, but this did not survive Bonferroni correction at $\alpha = 0.005$ ($t(21) = 2.257$, $p = 0.035$, Cohen's $d = 0.491$). Resolved was also higher for β compared with α, but this was not significant after multiple comparison correction either ($t(21) = 2.552$, $p = 0.019$, Cohen's $d = 0.549$). None of the other contrasts was significant ($p > 0.2$).

The ANOVA also revealed a significant effect of *model*, which was because of resolved being higher than remained open ($t(21) = 4.776$, $p < 0.001$, Cohen's $d = 1.151$), while there was a trend for resolved being higher than opened ($t(21) = 2.190$, $p = 0.040$, Cohen's $d = 0.470$) and the same for opened compared with remained open ($t(21) = 2.096$, $p = 0.048$, Cohen's $d = 0.522$). There was no significant main effect of frequency band ($p = 0.131$).

As all three dependency features contributed to explained variance of the α and β power, we created a full model with all features included. A 4 (model: opened/remained open/resolved/full) $\times$ 2 (band: α/β) repeated-measures ANOVA revealed a significant main effect of *model* ($F(3,63) = 6.004$, $p = 0.001$, $\eta^2 = 0.222$), which was because of the full model being higher than the remained open ($t(21) = 3.269$, $p = 0.004$, Cohen's $d = 0.981$) and a trend for the full being higher than the opened ($t(21) = 2.190$, $p = 0.040$, Cohen's $d = 0.508$). The full model was not significantly different from the resolved model ($t(21) = 1.385$, $p = 0.181$, Cohen's $d = 0.324$). There was no main effect of band or interaction between the variables ($p < 0.4$). Finally, reconstruction accuracy improvement of the full model was significantly higher than zero (α: $t(21) = 2.831$, $p = 0.010$; β: $t(21) = 6.710$, $p < 0.001$).

Dependency features improve reconstruction accuracy from mismatch model
To test whether reconstruction accuracy improved merely because of the addition of extra features, we compared the full model with mismatch models (Fig. 3A) (for details see section "Model comparison"). Reconstruction accuracy improvement was significantly higher than zero in all models and bands (α-opened: $t(21) = 5.432$, $p < 0.001$; remained open: $t(21) = 2.784$, $p = 0.011$; resolved: $t(21) = 4.074$, $p = 0.001$; full: $t(21) = 4.340$, $p < 0.001$; β-opened: $t(21) = 2.726$, $p = 0.013$; remained open: $t(21) = 2.688$, $p = 0.014$; resolved: $t(21) = 5.352$, $p < 0.001$; full: $t(21) = 4.232$, $p < 0.001$). A 4 (model: opened/remained open/resolved/full) $\times$ 2 (band: α/β) repeated-measures ANOVA revealed a significant main effect of model ($F(3,63) = 4.514$, $p = 0.006$, $\eta^2 = 0.177$). This was because of the full model showing higher improvement than both opened ($t(21) = 2.797$, $p = 0.011$, Cohen's $d = 0.757$) and remained open ($t(21) = 3.484$, $p = 0.002$, Cohen's $d = 0.634$). There was no other significant main effect or interaction between the variables ($p > 0.06$).

A

Dutch stories

reconstruction accuracy improvement (dependency minus base model)

reconstruction accuracy improvement (actual minus mismatch model)

B

French stories

reconstruction accuracy improvement (actual minus mismatch model)

	alpha	beta
opened		
remained open		
resolved		
full		
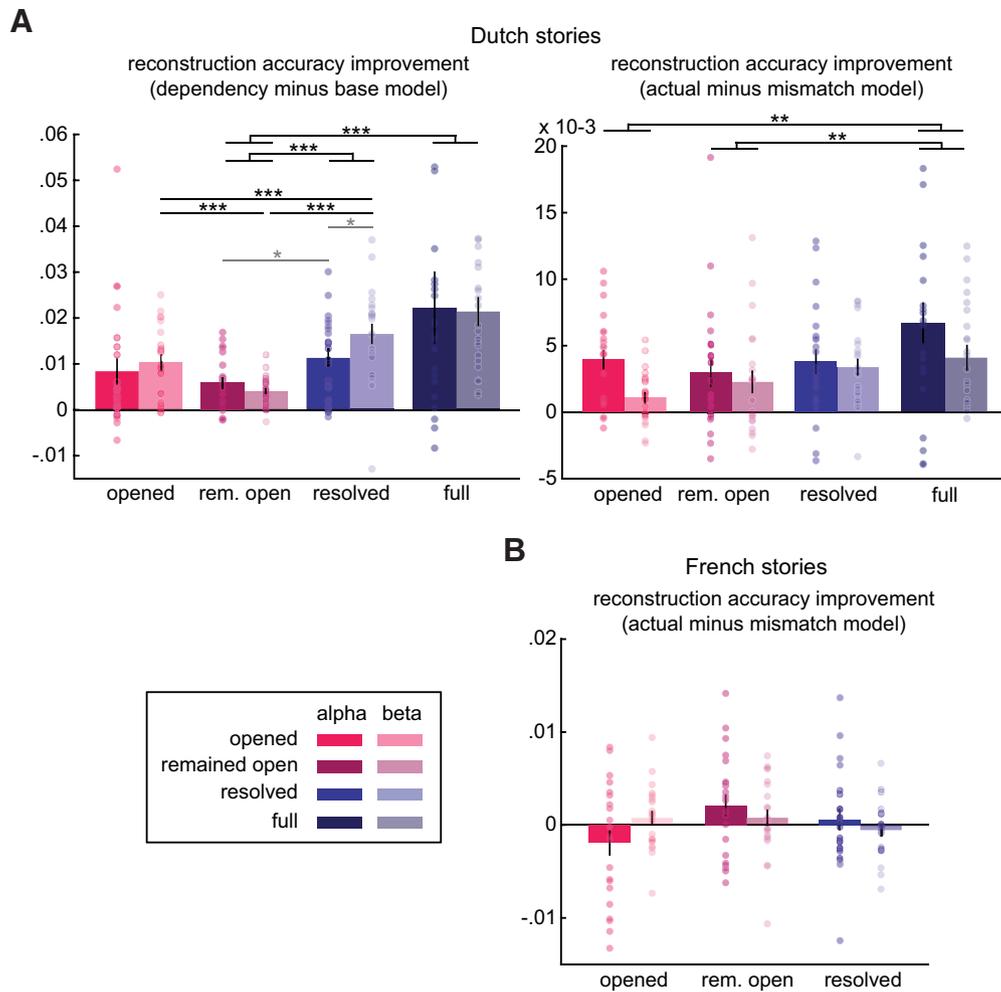
Figure 3. Model comparison. Reconstruction accuracy (Pearson's r) improvement using dependency features. **A**, Left, Reconstruction accuracy difference between dependency models (including opened/remained open/resolved/all dependency features) minus the base model (acoustic edges, word onset, word frequency features), averaged over sensors exhibiting improved accuracy over the base model (z-scored improvement >1 SD), for α (opaque) and β bands (transparent). Right, Reconstruction accuracy improvement from mismatch models (i.e., models in which the feature values of the respective feature are replaced by those from another story), averaged over sensors exhibiting improved accuracy, for α and β bands. For instance, the first bar group entitled "opened" refers to the comparison of the full model with a model in which the feature opened is mismatched (i.e., from a different story). The "full" bar group represents the comparison of the full model with a model in which all dependency features come from other stories. **B**, Reconstruction accuracy improvement from mismatch models in FR stories. Error bars indicate ± 1 SEM. *$p < 0.05$ (gray for uncorrected). **$p < 0.01$. ***$p < 0.001$.

Control analysis: reconstruction accuracy did not improve in French stories

In order to confirm that the effect of the dependency was because of language comprehension rather than any speech properties, we tested whether reconstruction accuracy was significantly higher in dependency versus mismatch models in the French stories. Conducting the same analysis as in Dutch stories, but for French, we found that reconstruction accuracy was not significantly higher than zero in any of the three dependency models ($p > 0.1$). There was no significant main effect or interaction between the variables ($p > 0.1$; Fig. 3B).

Dependency features modulate α power peaking in left temporal regions

Effect of comprehension: NL versus FR stories. First, we wanted to investigate the effect of comprehension by comparing the reconstruction accuracies of the full model (acoustic edges, word onset, word frequency, opened, remained open, resolved dependencies) in NL versus FR stories. A nonparametric cluster permutation test on source level revealed a significant cluster

mostly located in left temporal areas (cluster-corrected $p = 0.006$) for which NL exhibited higher reconstruction accuracy compared with FR (Fig. 4A). The mean t values of the significant cluster as well as the percentage of significant source points (voxels) per parcel (based on an anatomic atlas) are shown in Table 3.

Effect of dependency features. In order to identify the contributions of each dependency feature, we performed a cluster permutation test of the reconstruction accuracies in each dependency model (opened/remained open/resolved) versus the base model (Fig. 4B, bottom). We found significant clusters for opened versus base model ($p < 0.001$), remained open versus base ($p = 0.022$), and, finally, resolved versus base ($p < 0.001$). The mean t values of the significant clusters as well as the percentage of significant sources per parcel are shown in Table 3 (parcels including cortical areas only, subcortical excluded).

Subsequently, we analyzed the temporal profiles of the aforementioned effects, as demonstrated by their respective TRFs (Fig. 4B, top). The arbitrary units in Figure 4B,D represent the TRF model coefficients, as the degree of change in α/β power for every unit of change in the features. A positive value thus
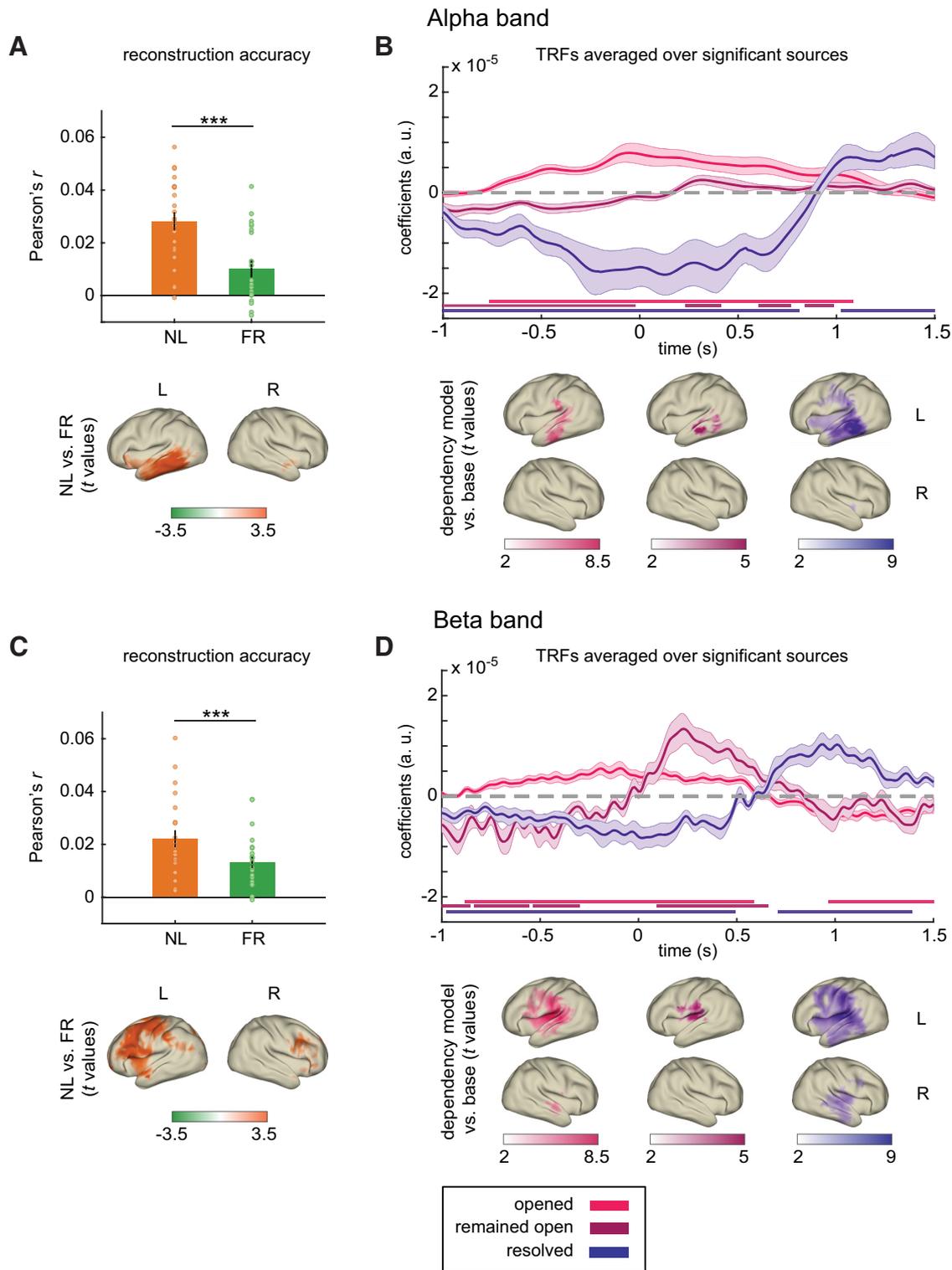
Figure 4. Results of the TRF analysis separately for the α and β bands. **A**, Top, Reconstruction accuracy of the full model in NL versus FR stories averaged over the significant sources. Bottom, Significant sources for the reconstruction accuracy between NL versus FR stories based on cluster permutation test. **B**, Top, Time courses of the TRF coefficients for each feature, averaged over significant sources shown below. Horizontal lines indicate time instances significantly different from a mismatch model (feature values replaced by those from another story) (FDR-corrected at $p = 0.05$). Bottom, Significant sources for the reconstruction accuracy of models opened/remained open/resolved versus the base model based on a cluster permutation test (cluster threshold at $t = 4.78$). **C**, **D**, Same as in **A**, **B** for the β band. Error bars indicate ± 1 SEM. ***$p < 0.001$.

represents power increase, while a negative value represents power decrease relative to the onset of the respective feature event. Specifically, we averaged the TRF coefficients over the significant sources as identified above. Paired-sample t tests identified the time instances at which the TRF coefficients were

significantly different from those of a mismatch model (FDR-corrected at $p = 0.05$). With regard to the effect of the opened feature, results revealed a long-lasting positive-going wave from $\sim\!-0.75$ to 1.1 s, peaking at word onsets (0 s). The remained open feature modulated α power negatively $\sim\!-1$ to 0 s, and

Table 3. Results of source localization analysis of dependency features[a]

Parcel	NL vs FR		Opened		Remained open		Resolved	
	t	%	t	%	t	%	t	%
Alpha band								
Middle frontal gyrus	0	0	2.23	8.82	0	0	2.85	65.86
Inferior frontal gyrus	1.77	1.70	2.71	91.15	0	0	4.47	100[b]
Precentral gyrus	0	0	3.38	40.38	0	0	4.43	67.63
Superior temporal gyrus	2.25	37.80	3.83	95.55	2.45[b]	6.93	5.07	100[b]
Middle temporal gyrus	3.09[b]	92.15[b]	4.60[b]	99.98[b]	3.40[b]	42.37[b]	6.53[b]	100[b]
Inferior temporal gyrus	3.81[b]	98.35[b]	4.76[b]	99.99[b]	3.15[b]	50.68[b]	6.93[b]	100[b]
Posterior superior temporal sulcus	2.71[b]	61.75[b]	5.61[b]	100.00[b]	2.32	11.80[b]	7.26[b]	100[b]
Superior parietal lobe	0	0	0	0	0	0	1.98	23.25
Inferior parietal lobe	2.03	2.24	3.99	57.21	0	0	4.24	78.68
Postcentral gyrus	0	0	3.75	43.47	0	0	4.49	68.61
Beta band								
Middle frontal gyrus	2.94	50.39	3.09	36.90	1.66	0.01	3.93	60.36
Inferior frontal gyrus	3.16[b]	75.21[b]	4.12	68.62	2.13	4.02	4.62	100[b]
Precentral gyrus	3.27[b]	83.51[b]	4.73	69.25	2.42	5.75	4.81	99.66
Superior temporal gyrus	2.51	35.62	5.21[b]	75.83	2.99[b]	5.86	5.22	94.86
Middle temporal gyrus	2.75	20.58	4.78	93.47	0	0	5.68[b]	99.87[b]
Inferior temporal gyrus	1.59	0.01	3.94	99.92[b]	0	0	5.85[b]	100[b]
Posterior superior temporal sulcus	0	0	6.95[b]	100[b]	2.07	7.49[b]	7.77	100[b]
Superior parietal lobe	2.38	15.27	1.94	1.29	0	0	2.86	80.88
Inferior parietal lobe	2.48	33.58	4.45	94.04[b]	3.23[b]	11.79[b]	5.13	98.01
Postcentral gyrus	2.96[b]	74.09[b]	4.81[b]	72.59	3.03[b]	14.78[b]	5.08	96.81

[a]Descriptive statistics of source localization of dependency features effects in the left hemisphere (Fig. 4). Columns "t" show the mean t value over significant voxels for NL versus FR and opened/remained open/resolved versus base models, while columns "%" show the percentage of significant voxels in the given parcel.
[b]Highest values per column.

positively between 0 and 1 s. Resolved dependencies showed a long-lasting negativity from ∼−1 to 0.75 s and a later positivity ∼1 to 1.5 s.

Dependency features modulate β power peaking in left frontal, parietal, and temporal regions

Effect of comprehension: NL versus FR stories. We also wanted to investigate the effect of comprehension in the β band by comparing the reconstruction accuracies of the full model in NL versus FR stories. A nonparametric cluster permutation test on source level revealed significant clusters, mostly located in left frontal areas ($p = 0.002$), for which NL exhibited higher reconstruction accuracy compared with FR (Fig. 4C; Table 3).

Effect of dependency features. A cluster permutation test of the reconstruction accuracies in each dependency model (opened/remained open/resolved) versus the base model (Fig. 4D, bottom; Table 3) revealed significant clusters for opened versus base model ($p < 0.001$), remained open versus base ($p = 0.025$), and resolved versus base ($p < 0.001$).

Further, we averaged over the TRF coefficients of the aforementioned significant sources and compared those to a mismatch model (Fig. 4D, top). Similar to the α band, in the β band, dependency opening was associated with an early positive-going wave from ∼−0.80 to 0.70 s, but also a negativity after ∼1 s. Remained open started with a negativity ∼−1 to −0.30, and showed a sharp positivity just after word onset, until 0.70 s. Finally, the resolved feature showed an early negativity up until ∼0.50 s, and a positive rebound after word onset at ∼1-1.5 s.

Control analysis in the δ, θ, and γ band
To test the frequency specificity of our effects, we constructed similar TRF models in δ, θ, and γ. A nonparametric cluster-based permutation test at sensor level between the reconstruction accuracy of each dependency model versus the base model was performed ($\alpha = 0.05$, two-tailed; 1,000 permutations). Results

revealed a positive significant cluster for the δ band (opened: $p = 0.004$; remained open: $p = 0.002$; resolved: $p = 0.002$) and θ power (opened: $p = 0.030$; remained open: $p = 0.032$; resolved: $p = 0.002$), but no clusters for θ band or γ power (no cluster or $p > 0.4$) (Fig. 5, bottom). TRF coefficients were averaged over significant sensors and compared with a mismatch model using paired-samples t tests (FDR correction at $p = 0.05$) (Fig. 5, top). TRFs were significantly different from the mismatch model for remained open (∼ −1 to 0.30 s and 0.80 to 1.30 s) and for resolved (0.10 to 1 s). There were no significant time instances in the δ band.

Discussion

In this study, we tested whether the functional role that α and β oscillations play in low-level perceptual processing can be generalized to naturalistic spoken language processing. Dutch native speakers listened to stories in Dutch and French while MEG was recorded. We identified three states at each word: number of opened/remained open/resolved dependencies. We then constructed forward models to predict α and β power from the dependency features, controlling for low-level linguistic features. We report the following key findings: (1) high-level syntactic features predict α and β power beyond acoustic, lexical, low-level linguistic features; (2) left temporal language-related regions are involved in language comprehension for α, while frontal and parietal, higher-order language regions, and motor regions play a critical role for β; and (3) α and β band dynamics subserve comprehension by contributing to higher-level operations, potentially associated with inhibition and reactivation or propagation processes, during structured meaning composition. Contrary to our expectations, the temporal profiles of α and β responses were highly similar, and dependency features also modulated lower-frequency bands, findings that do not align with functional specificity or dissociation of α and β in language comprehension.
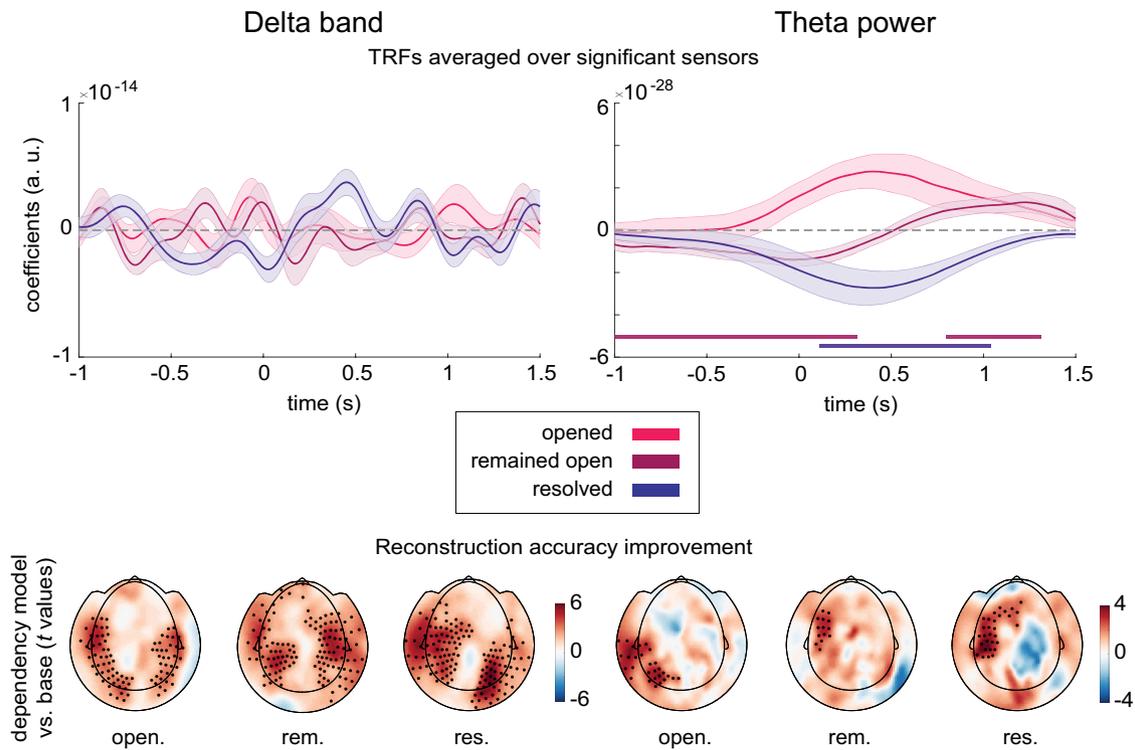
Figure 5. Results of the TRF analysis separately for δ band and θ power. Top, Time courses of the TRF coefficients for each feature, averaged over significant sensors shown below. Horizontal lines indicate significant time instances compared with a mismatch model (feature values replaced by those from another story) (FDR-corrected at $p = 0.05$). Bottom, t value topographies comparing the reconstruction accuracy of models opened/remained open/resolved versus the base model based on cluster permutation tests. Significant sensor clusters are marked on the topographies.

As expected, high-level syntactic features predicted α and β power beyond low-level linguistic features for NL, but not for FR stories. Our results provide evidence for the following: (1) dependency features are related to α and β modulations in comprehended but not in an uncomprehended spoken language, thus tapping into comprehension processes; and (2) α and β oscillations are modulated by higher-level operations associated with dependency formation and resolution in spoken language processing, beyond low-level features.

Consistent with our hypothesis, language comprehension seems to involve left temporal areas in α. Previous research demonstrated the role of those regions in lexical retrieval and creation of syntactic hierarchies (den Ouden et al., 2012; Klaus et al., 2020). The pMTG is argued to receive input from phonological networks and convert sequences of morphemes into nonlinear hierarchical structures, which are then mapped onto semantic networks (Matchin and Hickok, 2020). The anterior temporal lobe is associated with syntactic processing (Matchin et al., 2017), although mostly with semantic composition during combinatorial operations (Hagoort, 2013; Del Prato and Pylkkänen, 2014; Westerlund and Pylkkänen, 2014; Murphy, 2015; Segaert et al., 2018). Schoffelen et al. (2017) found that MTG exhibited a high degree of causal outflow to anterior temporal areas and the IFG, propagating information about lexical items to areas performing integration during sentence reading. In our study, dependency resolution was related to integration and unification of the dependent in the sentential context, while dependency formation was responsible for meaning construction during sentence evolution.

On the other hand, linguistic features modulated β band dynamics in a range of temporal, parietal, and frontal regions. The role of these regions in syntactic hierarchical structure and

semantic composition, as well as linguistic unification and integration processes, has been demonstrated (Friederici and von Cramon, 2000; M. Meyer et al., 2002; Dronkers et al., 2004; Rodd et al., 2005; Berwick et al., 2013; Zaccarella et al., 2017). The IFG is linked to binding words together into syntactic hierarchies, integration of abstract linguistic features with the existing context (Zaccarella et al., 2017; Friederici and von Cramon, 2000; ten Oever et al., 2022b), as well as encoding syntactic predictions (Matchin et al., 2017). Connections originating from temporal areas peak at α, whereas connections originating from parietal or frontal regions peak at β (Schoffelen et al., 2017). Finally, the contributions of motor and somatosensory areas might be related to motor-auditory system interactions for efficient speech comprehension (Morillon et al., 2014; Park et al., 2015, 2018; Morillon and Baillet, 2017; Assaneo and Poeppel, 2018; Keitel et al., 2018; Rimmele et al., 2018; Terporten et al., 2019; Abbasi and Gross, 2020; Poeppel and Assaneo, 2020; ten Oever and Martin, 2021).

Surprisingly, α and β power exhibited a similar time course of activation. Power increased before and decreased after the opening of new dependencies, especially β. The effects were widespread, both bands peaking in posterior superior temporal sulcus, and to a lesser extent in MTG and ITG for α, but in STG and PCG for β. The time courses are difficult to interpret; nevertheless, potential explanations of the power increases before dependency opening could lie in reduced interference from already open dependencies (Worden et al., 2000; Sauseng et al., 2009) or the motor system preparing the auditory system for processing of upcoming information (Abbasi and Gross, 2020). Encoding of new dependencies might be reflected in α and β decrease following the opening of dependencies signaling increased cortical excitability and active processing (Palva and Palva, 2007;

Sauseng et al., 2009) or, speculatively, reflecting coordinate transform from sensory information across the linguistic hierarchy (Martin, 2016, 2020).

A power increase followed unresolved dependencies for β, peaking in temporo-parietal regions, and to a lesser extent for α, in temporal regions. Processing difficulty increases with open dependencies (Vos et al., 2001; Demberg and Keller, 2008), while increased load is typically associated with power decreases in task-relevant regions (Jensen and Mazaheri, 2010). Therefore, the observed power increases were surprising. Nevertheless, there is evidence for higher β power in long- than short-distance dependencies, interpreted as active "maintenance" of the current processing mode (Meyer et al., 2013). Considering that low-level features were associated with faster power modulations after word onset, the temporally diffuse effects for dependency features are possibly because of syntactic and anticipatory processes, linked to incremental effects spanning beyond the single-word level (Bastiaansen et al., 2010).

Finally, we observed power decreases before and a positive rebound following dependency resolution. These peaked in temporal regions, while frontal and parietal modulations were stronger in β than in α. Typically, α and β power decreases are associated with upregulation of the cortex, making it more susceptible to processing upcoming information (Schubert et al., 2009; Haegens et al., 2011a; Ede et al., 2011) via increased neuronal excitability (Gastaldon et al., 2020; Iemi et al., 2022). The observed power decreases could reflect preparation for dependency resolution processing, by a release of language-related regions from inhibition. The subsequent β increase is in line with the proposition on β oscillations supporting reactivation of latent content representations (Antzoulatos and Miller, 2016; Spitzer and Haegens, 2017), based on studies showing content-specific β modulations during information recall (Haegens et al., 2011a, 2017; Spitzer and Blankenburg, 2011; Spitzer et al., 2014; Wimmer et al., 2016). Importantly, the dependent constituent's representation is retrieved from memory and reactivated during dependency resolution, to be integrated with the current context (Bever and McElree, 1988; Nicol and Swinney, 1989; McElree et al., 2003; Martin and McElree, 2008). We therefore speculate that the power increases reflect reactivation of the dependent constituent linked to integration and interpretation.

Previous studies found distinct roles of α and β in language processing. Röhm et al. (2001) and Scharinger et al. (2015) found reduced α power during sentence processing with higher processing load compared with baseline reading. Furthermore, α desynchronization was linked to syntactic/semantic violations (Davidson and Indefrey, 2007). Meyer et al. (2013) interpreted increased α power in sentences with long- than short-distance dependencies as suppression of interference, and increased β power at dependency resolution in long-dependency sentences as syntactic integration. A gradual β increase in syntactically structured sentences was observed in β but not in other bands (Bastiaansen et al., 2010). Lewis and Bastiaansen (2015) proposed that β may carry top-down predictive information based on the sentence context to regions responsible for hierarchically "lower" sensory and perceptual processing.

We acknowledge our findings should be interpreted with caution, especially with regard to the frequency specificity of the effects and the distinct role of α and β oscillations in language comprehension. First, in contrast with our hypothesis, the observed effects were highly similar between α and β in the temporal domain. Therefore, our findings do not provide robust evidence for distinct functional roles of α and β.

However, there was a spatial distinction in the Dutch versus French contrast tapping into the neural correlates of comprehension. It could thus be that dependency features recruited α and β oscillations in a time-locked manner, engaging different brain systems with the same timing. Second, control analysis showed that the dependency features also modulate lower-frequency bands (although we did not observe distinct spectral peaks for these lower bands). This finding does not align with the idea that dependency features specifically modulate α and β. Future studies using controlled-trial paradigms are needed to elucidate the effects of high-level linguistic features on α and β.

Overall, our study offers novel contributions for studying α and β dynamics "in the wild" using naturalistic stimuli and TRFs. Our findings are consistent with an account where oscillatory building blocks are reconfigured as a function of language processing. Assuming that dependency features tap into domain-general operations, our results provide support for the generalization of the α and β functional roles from sensory operations to complex linguistic processing. However, the specificity of α and β modulations by high-level features and the functional distinction between the two bands were only partially supported by the findings.

References

Abbasi O, Gross J (2020) Beta-band oscillations play an essential role in motor–auditory interactions. Hum Brain Mapp 41:656–665.

Antzoulatos E, Miller E (2016) Synchronous beta rhythms of frontoparietal networks support only behaviorally relevant representations. Elife 5: e17822.

Assaneo MF, Poeppel D (2018) The coupling between auditory and motor cortices is rate-restricted: evidence for an intrinsic speech-motor rhythm. Sci Adv 4:eaao3842.

Bai F, Meyer AS, Martin AE (2022) Neural dynamics differentially encode phrases and sentences during spoken language comprehension. PLoS Biol 20:e3001713.

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

Berwick R, Friederici A, Chomsky N, Bolhuis J (2013) Evolution, brain, and the nature of language. Trends Cogn Sci 17:89–98.

Bever T, McElree B (1988) Empty categories access their antecedents during comprehension. Linguistic Inquiry 19:35–43.

Brainard D, Vision S (1997) The psychophysics toolbox. Spat Vis 10:433–436.

Brennan J, Martin A (2020) Phase synchronization varies systematically with linguistic structure composition. Philos Trans R Soc Lond B Biol Sci 375:20190305.

Brodbeck C, Presacco A, Simon J (2018) Neural source dynamics of brain responses to continuous stimuli: speech processing from acoustics to comprehension. Neuroimage 172:162–174.

Broderick M, Anderson A, Liberto G, Di Crosse M, Lalor E (2018) Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. Curr Biol 28:803–809.e3.

Coopmans CW, de Hoop H, Hagoort P, Martin AE (2022) Effects of structure and meaning on cortical tracking of linguistic units in naturalistic speech. Neurobiol Lang 3:386–412.

Corver N, van Riemsdijk H (2013) Semi-lexical categories. In: The function of content words and the content of function words (Corver N, van Riemsdijk H, eds). Berlin: Walter de Gruyter.

Crosse MJ, Di Liberto GM, Bednar A, Lalor EC (2016) The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. Front Hum Neurosci 10:604.

de Vries W, Nissim M (2020) As good as new: how to successfully recycle English GPT-2 to make models for other languages. ArXiv 836–846.

Davidson D, Indefrey P (2007) An inverse relation between event-related and time -frequency violation responses in sentence processing. Brain Research 1158:81–92.

Del Prato P, Pylkkänen L (2014) MEG evidence for conceptual combination but not numeral quantification in the left anterior temporal lobe during language production. Front Psychol 5:524.

Demberg V, Keller F (2008) Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. Cognition 109:193–210.

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

Ding N, Simon JZ (2012) Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. J Neurophysiol 107:78–89.

Ding N, Simon JZ (2014) Cortical entrainment to continuous speech: functional roles and interpretations. Front Hum Neurosci 8:311.

Doelling K, Arnal L, Ghitza O, Poeppel D (2014) Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. Neuroimage 85:761–768.

Donoghue T, Haller M, Peterson EJ, Varma P, Sebastian P, Gao R, Noto T, Lara AH, Wallis JD, Knight RT, Shestyuk A, Voytek B (2020) Parameterizing neural power spectra into periodic and aperiodic components. Nat Neurosci 23:1655–1665.

Dronkers N, Wilkins D, Van Valin R Jr, Redfern B, Jaeger J (2004) Lesion analysis of the brain areas involved in language comprehension. Cognition 92:145–177.

Ede FV, Lange FD, Jensen O, Maris E (2011) Orienting attention to an upcoming tactile event involves a spatially and temporally specific modulation of sensorimotor alpha-and beta-band oscillations. J Neurosci 31:2016–2024.

Engel AK, Fries P (2010) Beta-band oscillations: signalling the status quo? Curr Opin Neurobiol 20:156–165.

Fan L, Li H, Zhuo J, Zhang Y, Wang J, Chen L, Yang Z, Chu C, Xie S, Laird AR, Fox PT, Eickhoff SB, Yu C, Jiang T (2016) The human brainnetome atlas: a new brain atlas based on connectional architecture. Cereb Cortex 26:3508–3526.

Fishbach A, Nelken I, Yeshurun Y (2001) Auditory edge detection: a neural model for physiological and psychoacoustical responses to amplitude transients. J Neurophysiol 85:2303–2323.

Foraker S, McElree B (2011) Comprehension of linguistic dependencies: speed-accuracy tradeoff evidence for direct-access retrieval from memory. Lang Linguistics Compass 5:764–783.

Friederici A, von Cramon D (2000) Syntax in the brain: linguistic versus neuroanatomical specificity. Behav Brain Sci 23:32–33.

Gastaldon S, Arcara G, Navarrete E, Peressotti F (2020) Commonalities in alpha and beta neural desynchronizations during prediction in language comprehension and production. Cortex 133:328–345.

Haegens S, Osipova D, Oostenveld R, Jensen O (2010) Somatosensory working memory performance in humans depends on both engagement and disengagement of regions in a distributed network. Hum Brain Mapp 31:26–35.

Haegens S, Nácher V, Hernández A, Luna R, Jensen O, Romo R (2011a) Beta oscillations in the monkey sensorimotor network reflect somatosensory decision making. Proc Natl Acad Sci USA 108:10708–10713.

Haegens S, Händel BF, Jensen O (2011b) Top-down controlled alpha band activity in somatosensory areas determines behavioral performance in a discrimination task. J Neurosci 31:5197–5204.

Haegens S, Luther L, Jensen O (2012) Somatosensory anticipatory alpha activity increases to suppress distracting input. J Cogn Neurosci 24:677–685.

Haegens S, Vergara J, Rossi-Pool R, Lemus L, Romo R (2017) Beta oscillations reflect supramodal information during perceptual judgment. Proc Natl Acad Sci USA 114:13810–13815.

Hagoort P (2005) On Broca, brain, and binding: a new framework. Trends Cogn Sci 9:416–423.

Hagoort P (2013) MUC (memory, unification, control) and beyond. Front Psychol 4:416.

Hauswald A, Keitel A, Chen YP, Rösch S, Weisz N (2022) Degradation levels of continuous speech affect neural speech tracking and alpha power differently. Eur J Neurosci 55:3288–3302.

Iemi L, Gwilliams L, Samaha J, Auksztulewicz R, Cycowicz YM, King JR, Nikulin VV, Thesen T, Doyle W, Devinsky O, Schroeder CE, Melloni L, Haegens S (2022) Ongoing neural oscillations influence behavior and sensory representations by suppressing neuronal excitability. Neuroimage 247:118746.

Jensen O, Gelfand J, Kounios J, Lisman J (2002) Oscillations in the alpha band (9-12 Hz) increase with memory load during retention in a short-term memory task. Cereb Cortex 12:877–882.

Jensen O, Mazaheri A (2010) Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Front Hum Neurosci 4:186.

Kapteijns B, Hintz F (2021) Comparing predictors of sentence self-paced reading times: syntactic complexity versus transitional probability metrics. PLoS One 16:e0254546.

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

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

Keuleers E, Brysbaert M, New B (2010) SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles. Behav Res Methods 42:643–650.

Kilavik BE, Zaepffel M, Brovelli A, MacKay WA, Riehle A (2013) The ups and downs of beta oscillations in sensorimotor cortex. Exp Neurol 245:15–26.

Kisler T, Reichel U, Schiel F (2017) Multilingual processing of speech via web services. Comput Speech Lang 45:326–347.

Klaus J, Schutter D, Piai V (2020) Transient perturbation of the left temporal cortex evokes plasticity-related reconfiguration of the lexical network. Hum Brain Mapp 41:1061–1071.

Klimesch W, Sauseng P, Hanslmayr S (2007) EEG alpha oscillations: the inhibition–timing hypothesis. Brain Res Rev 53:63–88.

Kojima K, Oganian Y, Cai C, Findlay A, Chang E, Nagarajan S (2020) Low-frequency neural tracking of speech envelope reflects evoked responses to acoustic edges. BioRxiv 022616. https://doi.org/10.1101/2020.04.02.022616

Kösem A, Basirat A, Azizi L, van Wassenhove V (2016) High-frequency neural activity predicts word parsing in ambiguous speech streams. J Neurophysiol 116:2497–2512.

Lewis AG, Bastiaansen M (2015) A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. Cortex 68:155–168.

Lewis A, Wang L, Bastiaansen M (2015) Fast oscillatory dynamics during language comprehension: unification versus maintenance and prediction? Brain Lang 148:51–63.

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

Martin A (2020) A compositional neural architecture for language. J Cogn Neurosci 32:1407–1427.

Martin AE (2016) Language processing as cue integration: grounding the psychology of language in perception and neurophysiology. Front Psychol 7:120.

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

Martin A, McElree B (2008) A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. J Mem Lang 58:879–906.

Matchin W, Hammerly C, Lau E (2017) The role of the IFG and pSTS in syntactic prediction: evidence from a parametric study of hierarchical structure in fMRI. Cortex 88:106–123.

Matchin W, Hickok G (2020) The cortical organization of syntax. Cereb Cortex 30:1481–1498.

McElree B, Foraker S, Dyer L (2003) Memory structures that subserve sentence comprehension. J Mem Lang 48:67–91.

Mel'cuk I (1988) Dependency syntax: theory and practice. https://books.google.com/books?hl=en&lr=&id=diq29vrjAa4C&oi=fnd&pg=PR13&ots=ZgyIxqyG8h&sig=dOGJJrGS-cb3p2cKbkTW3YyM2zg.

Meyer L (2018) The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms. Eur J Neurosci 48:2609–2621.

Meyer L, Obleser J, Friederici A (2013) Left parietal alpha enhancement during working memory-intensive sentence processing. Cortex 49:711–721.

Meyer L, Sun Y, Martin A (2020) 'Entraining' to speech, generating language? Lang Cogn Neurosci 35:1138–1148.

Meyer M, Alter K, Friederici AD, Lohmann G, Von Cramon DY (2002) FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Hum Brain Mapp 17:73–88.

Morillon B, Schroeder C, Wyart V (2014) Motor contributions to the temporal precision of auditory attention. Nat Commun 5:1–9.

Morillon B, Baillet S (2017) Motor origin of temporal predictions in auditory attention. Proc Natl Acad Sci USA 114:E8913–E8921.

Murphy E (2015) The brain dynamics of linguistic computation. Front Psychol 6:1515.

New B, Pallier C, Brysbaert M, Ferrand L (2004) Lexique 2: a new French lexical database. Behav Res Methods Instrum Comput 36:516–524.

Nicol J, Swinney D (1989) The role of structure in coreference assignment during sentence comprehension. J Psycholinguist Res 18:5–19.

Nivre J, De Marneffe M, Ginter F, Goldberg Y, Hajic J, Manning C, McDonald R, Petrov S, Pyysalo S, Silveira N, Tsarfaty R (2016) Universal dependencies v1: a multilingual treebank collection. Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 1659–1666.

Nolte G (2003) The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. Phys Med Biol 48:3637–3652.

Obleser J, Weisz N (2012) Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. Cereb Cortex 22:2466–2477.

Oostenveld R, Fries P, Maris E, Schoffelen J (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intel Neurosci 2011:1–9.

den Ouden DB, Saur D, Mader W, Schelter B, Lukic S, Wali E, Timmer J, Thompson CK (2012) Network modulation during complex syntactic processing. Neuroimage 59:815–823.

Palva S, Palva J (2007) New vistas for α-frequency band oscillations. Trends Neurosci 30:150–158.

Pan Y, Tan Z, Gao Z, Li Y, Wang L (2018) Neural activity is dynamically modulated by memory load during the maintenance of spatial objects. Front Psychol 9:1071.

Park H, Ince R, Schyns P, Thut G, Gross J (2015) Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. Curr Biol 25:1649–1653.

Park H, Thut G, Gross J (2018) Predictive entrainment of natural speech through two fronto-motor top-down channels. Lang Cogn Neurosci 35:739–751.

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

Qi P, Zhang Y, Zhang Y, Bolton J, Manning C (2020) Stanza: a python natural language processing toolkit for many human languages. ArXiv 2003.07082.

Rimmele J, Morillon B, Poeppel D, Arnal L (2018) Proactive sensing of periodic and aperiodic auditory patterns. Trends Cogn Sci 22:870–882.

Rodd J, Davis M, Johnsrude I (2005) The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. Cereb Cortex 15:1261–1269.

Röhm D, Klimesch W, Haider H, Doppelmayr M (2001) The role of theta and alpha oscillations for language comprehension in the human electroencephalogram. Neuroscience Letters 310:137–140.

Rose NS, LaRocque JJ, Riggall AC, Gosseries O, Starrett MJ, Meyering EE, Postle BR (2016) Reactivation of latent working memories with transcranial magnetic stimulation. Science 354:1136–1139.

Sauseng P, Klimesch W, Gerloff C, Hummel FC (2009) Spontaneous locally restricted EEG alpha activity determines cortical excitability in the motor cortex. Neuropsychologia 47:284–288.

Scharinger C, Kammerer Y, Gerjets P (2015) Pupil dilation and EEG alpha frequency band power reveal load on executive functions for link-selection processes during text reading. PloS One 10:e0130608.

Scheeringa R, Petersson K, Oostenveld R, Norris D, Hagoort P, Bastiaansen M (2009) Trial-by-trial coupling between EEG and BOLD identifies networks related to alpha and theta EEG power increases during working memory maintenance. Neuroimage 44:1224–1238.

Schoffelen JM, Hultén A, Lam N, Marquand AF, Uddén J, Hagoort P (2017) Frequency-specific directed interactions in the human brain network for language. Proc Natl Acad Sci USA 114:8083–8088.

Schubert R, Haufe S, Blankenburg F, Villringer A, Curio G (2009) Now you'll feel it, now you won't: EEG rhythms predict the effectiveness of perceptual masking. J Cogn Neurosci 21:2407–2419.

Segaert K, Mazaheri A, Hagoort P (2018) Binding language: structuring sentences through precisely timed oscillatory mechanisms. Eur J Neurosci 48:2651–2662.

Spitzer B, Blankenburg F (2011) Stimulus-dependent EEG activity reflects internal updating of tactile working memory in humans. Proc Natl Acad Sci USA 108:8444–8449.

Spitzer B, Haegens S (2017) Beyond the status quo: a role for beta oscillations in endogenous content (re)activation. eNeuro 4:ENEURO.0170-17.2017.

Spitzer B, Gloel M, Schmidt T, Blankenburg F (2014) Working memory coding of analog stimulus properties in the human prefrontal cortex. Cereb Cortex 24:2229–2236.

ten Oever ST, Martin A (2021) An oscillating computational model can track pseudo-rhythmic speech by using linguistic predictions. Elife 10:e68066.

ten Oever ST, Carta S, Kaufeld G, Martin A (2022a) Task-dependent and automatic tracking of hierarchical linguistic structure. BioRxiv 479571. https://doi.org/10.1101/2022.02.08.479571

ten Oever ST, Kaushik K, Martin AE (2022b) Inferring the nature of linguistic computations in the brain. PLoS Comput Biol 18:e1010269.

Terporten R, Schoffelen J, Dai B, Hagoort P, Kösem A (2019) The relation between alpha/beta oscillations and the encoding of sentence induced contextual information. Sci Rep 9:1–12.

Tesnière L (2015) Elements of structural syntax. https://library.oapen.org/handle/20.500.12657/30722.

Tuladhar AM, Ter Huurne N, Schoffelen JM, Maris E, Oostenveld R, Jensen O (2007) Parieto-occipital sources account for the increase in alpha activity with working memory load. Hum Brain Mapp 28:785–792.

van Bree S, Sohoglu E, Davis MH, Zoefel B (2021) Sustained neural rhythms reveal endogenous oscillations supporting speech perception. PLoS Biol 19:e3001142.

Varela F, Lachaux JP, Rodriguez E, Martinerie J (2001) The brainweb: phase synchronization and large-scale integration. Nat Rev Neurosci 2:229–239.

Veen BV, Van Drongelen W, Yuchtman M, Suzuki A (1997) Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans Biomed Eng 44:867–880.

Vos S, Gunter T, Schriefers H, Friederici A (2001) Syntactic parsing and working memory: the effects of syntactic complexity, reading span, and concurrent load. Lang Cogn Process 16:65–103.

Weissbart H, Kandylaki K, Reichenbach T (2020) Cortical tracking of surprisal during continuous speech comprehension. J Cogn Neurosci 32:155–166.

Westerlund M, Pylkkänen L (2014) The role of the left anterior temporal lobe in semantic composition vs. semantic memory. Neuropsychologia 57:59–70.

Wilson B, Spierings M, Ravignani A, Mueller JL, Mintz TH, Wijnen F, van der Kant A, Smith K, Rey A (2020) Non-adjacent dependency learning in humans and other animals. Top Cogn Sci 12:843–858.

Wimmer K, Ramon M, Pasternak T, Compte A (2016) Transitions between multiband oscillatory patterns characterize memory-guided perceptual decisions in prefrontal circuits. J Neurosci 36:489–505.

Worden MS, Foxe JJ, Wang N, Simpson GV (2000) Anticipatory biasing of visuospatial attention indexed by retinotopically specific α-bank electroencephalography increases over occipital cortex. J Neurosci 20:RC63.

Zaccarella E, Meyer L, Makuuchi M, Friederici A (2017) Building by syntax: the neural basis of minimal linguistic structures. Cereb Cortex 27:411–421.

Zoefel B, VanRullen R (2015) The role of high-level processes for oscillatory phase entrainment to speech sound. Front Hum Neurosci 9:651.