

Research



**Cite this article:** Brennan JR, Martin AE. 2019

Phase synchronization varies systematically with linguistic structure composition. *Phil. Trans. R. Soc. B* **375**: 20190305.

<http://dx.doi.org/10.1098/rstb.2019.0305>

Accepted: 10 September 2019

One contribution of 16 to a theme issue 'Towards mechanistic models of meaning composition'.

**Subject Areas:**

cognition, neuroscience

**Keywords:**

compositionality, language, electroencephalography, neural oscillations, phase synchronization, naturalistic language processing

**Author for correspondence:**

Andrea E. Martin

e-mail: [andrea.martin@mpi.nl](mailto:andrea.martin@mpi.nl)

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.4723916>.

# Phase synchronization varies systematically with linguistic structure composition

Jonathan R. Brennan<sup>1</sup> and Andrea E. Martin<sup>2,3</sup>

<sup>1</sup>Department of Linguistics, University of Michigan, Ann Arbor, MI, USA

<sup>2</sup>Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands

<sup>3</sup>Donders Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition, and Behavior, Radboud University, Nijmegen, The Netherlands

JRB, 0000-0002-3639-350X; AEM, 0000-0002-3395-7234

Computation in neuronal assemblies is putatively reflected in the excitatory and inhibitory cycles of activation distributed throughout the brain. In speech and language processing, coordination of these cycles resulting in phase synchronization has been argued to reflect the integration of information on different timescales (e.g. segmenting acoustics signals to phonemic and syllabic representations; (Giraud and Poeppel 2012 *Nat. Neurosci.* **15**, 511 (doi:10.1038/nn.3063)). A natural extension of this claim is that phase synchronization functions similarly to support the inference of more abstract higher-level linguistic structures (Martin 2016 *Front. Psychol.* **7**, 120; Martin and Doumas 2017 *PLoS Biol.* **15**, e2000663 (doi:10.1371/journal.pbio.2000663); Martin and Doumas. 2019 *Curr. Opin. Behav. Sci.* **29**, 77–83 (doi:10.1016/j.cobeha.2019.04.008)). Hale *et al.* (Hale *et al.* 2018 Finding syntax in human encephalography with beam search. *arXiv* 1806.04127 (<http://arxiv.org/abs/1806.04127>)) showed that syntactically driven parsing decisions predict electroencephalography (EEG) responses in the time domain; here we ask whether phase synchronization in the form of either inter-trial phrase coherence or cross-frequency coupling (CFC) between high-frequency (i.e. gamma) bursts and lower-frequency carrier signals (i.e. delta, theta), changes as the linguistic structures of compositional meaning (*viz.*, bracket completions, as denoted by the onset of words that complete phrases) accrue. We use a naturalistic story-listening EEG dataset from Hale *et al.* to assess the relationship between linguistic structure and phase alignment. We observe increased phase synchronization as a function of phrase counts in the delta, theta, and gamma bands, especially for function words. A more complex pattern emerged for CFC as phrase count changed, possibly related to the lack of a one-to-one mapping between 'size' of linguistic structure and frequency band—an assumption that is tacit in recent frameworks. These results emphasize the important role that phase synchronization, desynchronization, and thus, inhibition, play in the construction of compositional meaning by distributed neural networks in the brain.

This article is part of the theme issue 'Towards mechanistic models of meaning composition'.

## 1. Introduction

A comprehensive account of linguistic compositionality must face up to difficult questions, namely, how such a powerfully expressive formal property can be achieved in a neurophysiological system like the human brain. Language is the canonical externally measurable human perception-action system that exhibits compositionality [1]. From a theorist's-eye view, a principal challenge is to consider how language comprehension might occur from speech or sign,

while both obeying the constraints of neurophysiology and staying faithful to the boundary conditions of language [2–4]. To meet this challenge, the brain must transform the sensory correlates of speech and sign into structured meaning. Gathering evidence for how the latter might occur is the focus of this paper. Specifically, we test here the hypothesis that the phase of neural oscillations, as measured with electroencephalography (EEG), is systematically related to composition in natural language, operationalized in terms of the number of phrases completed at a particular word.

The neurobiology of language has been concerned not only with the identification of network(s) of brain regions implicated in language processing, but also with inferring the computations that might take place across those networks even while the exact nature of those computations remains ill-defined [5].<sup>1</sup> Whatever computations are being carried out by the language network(s), it is likely that the coordination of activity across distributed populations plays a key role in bringing linguistic knowledge to bear on incoming speech- or sign-related sensory processing.<sup>2</sup> As such, our focus in this paper is to contribute to an emerging constellation of results that tries to understand how oscillatory patterns that reflect such coordinated neural activity might relate to the hypothesized states of the structured linguistic representations undergoing composition.

Here we examine the relationship between the oscillatory dynamics of induced neural activity and incremental composition, or parsing decisions [8–12]. While the majority of the literature has focused on the relationship between amplitude (in the evoked signal) or power (in the induced signal) and psycholinguistic manipulations; we turn our attention to the relationship between *phase* and a linguistic structure annotation that reflects compositionality on a word-by-word basis.

There are compelling reasons to believe that the phase of activation across distributed neural populations, not just amplitude or power, is critical for neural information coding, transmission and generation. Some of this evidence comes from physiology [7,13–18], but also some comes from the few formalized implementations of symbol-processing systems in artificial neural networks. See, for example, the notion of *phase set* in the model *Learning and Inference with Schemas and Analogies* (LISA; [19,20]) and *Discovery of Relations by Analogy* (DORA; [3,4,21–23]). In the latter model, for example, phase synchronization, desynchronization and yoked inhibitory ‘interneurons’ are mechanistically important for carrying and controlling binding information while also maintaining independence (see also [24]).

One way to explore the role of phase in the construction of higher-level linguistic representations from speech is to look at phase coherence; this measure tracks how consistent the phase of an oscillation is across repeated trials. Tracking this allows us to test whether phase in a particular frequency band becomes more or less consistent for more complex compositional linguistic representations. Another way is to examine cross-frequency coupling (CFC; [25,26]). Measures of CFC track the alignment or synchronization between oscillations at different frequencies. Neuronal assemblies putatively carry and exchange information through phase synchronization and CFC [7,13]. In models of cortical speech processing, for example, CFC links acoustic signals to phonemic and syllabic representations, which occur on different timescales [27–29].

We adapt from those models the leading idea that when more representations are inferred, phase between

assemblies will need to be more consistent, in terms of synchronization and desynchronization, in order to coordinate across populations. We do not directly address the lively debate of whether ongoing neural entrainment is causal to perception or neural computation (see endnote 2), a debate which is also related to the controversy about whether neural oscillations are the product of ‘true’ oscillators or, instead, reflect punctate local field potentials (LFPs). The literature connecting neural computations at this level of detail to linguistic composition is in its infancy, but we do not know any current tenable models for which linguistic structures<sup>3</sup> arise from any kind of true oscillators. Rather, Martin [2] describes a hybrid architecture based on an ongoing speech-envelope-driven oscillation that is transformed or biased in state space by punctate LFPs. These LFPs reflect perceptual inferences of abstract linguistic structures (*viz.*, phonetic features, phonemes, words, morphemes, syntactic structures, compositional units) cued by the oscillator. We note that the details of such a model are underspecified and the data in support of aspects of the theory are only beginning to emerge (e.g. [31,32]).

Against this nascent theoretical background, we address the basic question of how phase relationships at different timescales relate to compositional structure. We quantify the contribution of structure to compositional meaning as a function of the number of phrases that are completed at a given word. We ask how phase synchronization within and between high-frequency (i.e. gamma) bursts and lower-frequency carrier signals (i.e. delta, theta) correlates with compositional structure quantified in this way.

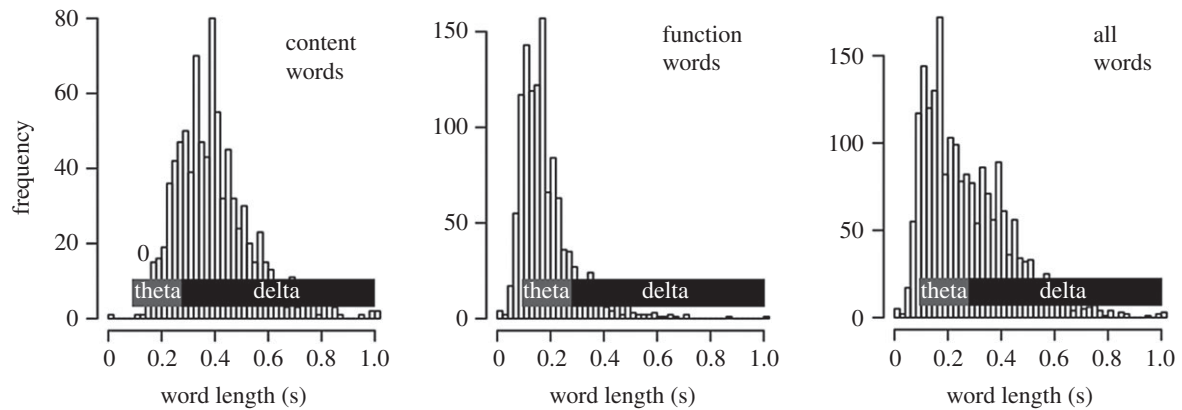
## 2. Methods

### (a) Stimulus and electroencephalography data

We use a publicly available set of 33 EEG datasets collected while adult participants passively listen to an audiobook of the first chapter of *Alice in Wonderland* [12,33].<sup>4</sup> This stimulus is 12.4 min long and comprises 2129 words. Participants listened to the story over in-ear headphones in an enclosed booth at a loudness of 45 dB above their hearing threshold. The EEG data were recorded at 500 Hz using 61 actively amplified electrodes with impedances kept below 25 kOhm referenced to the left mastoid electrode.

The data are processed using the FIELDTRIP toolbox in MATLAB in two stages. In the first stage, artefacts are identified and removed using previously published procedures [33]: The raw EEG data are (i) re-referenced to the average of two mastoid electrodes, (ii) high-pass filtered at 0.1 Hz, and (iii) divided into epochs spanning –300 to 1000 ms around each word onset. Ocular artefacts are isolated and removed using independent component analysis and remaining epochs or channels containing artefacts are removed based on visual inspection (an average of 13.5% of epochs are marked as containing artefacts across participants; see [33] for further details on data pre-processing).

In a second stage of data processing, we extract band-specific power and phase information for each word in the stimulus. First, each participant’s raw data is re-loaded from disc, re-referenced to the average of two mastoid electrodes, high-pass filtered at 0.1 Hz, and previously identified ocular independent components are subtracted. Second, data from missing or artefactual electrodes are reconstructed using surface-spline interpolation. Third, the continuous data are band-pass filtered in three bands: delta (1–4 Hz), theta (4–8 Hz) and gamma (30–50 Hz) (4th order, butterworth) and then converted to their respective analytic signal using the



**Figure 1.** Histograms of word length from the audiobook stimulus, divided into content words and function words. Annotations indicate predominant neural rhythms that have been associated with linguistic representations of different sizes (e.g. [27]).

**Table 1.** Number of function words and content words within each syntactic phrase bin in the audiobook stimulus.

	function words	content words
one	694	396
two	272	201
three+	64	181

Hilbert transformation. Finally, the instantaneous power and phase as measured at the vertex electrode (Cz) is extracted for each band and for each word at the time-point corresponding to that word's onset, excluding epochs marked as artefactual. We focus on word onset because our main focus includes relatively low frequencies (e.g. 1–4 Hz) where temporal resolution is the same order of magnitude as the rate of word presentation (2–4 words  $s^{-1}$ ).

### (b) Compositional annotation

Following previous work, we quantify composition by counting the number of phrases completed by a given word in the story ([9,34]; see also [8,11]). Phrase completion was derived from a bottom-up tree traversal that enumerates mothers before daughters based on the Penn Treebank annotation scheme [35]. As described in Brennan *et al.* [34], the story-book text was annotated by the Bikel implementation of the Collins parser trained on the Penn Treebank and the resulting phrase-structure annotation was manually reviewed. This annotation yields an estimate for each word in the story of the number of phrases completed by that word. Phase synchronicity must be assessed across sets of epochs. Accordingly, in the last annotation step we create three bins: words that complete a single phrase, words that complete two phrases, and words that complete three or more phrases. For each bin, we derive several phase synchronicity measures as described below.

Table 1 shows the count of words for each of the three phrase bins, divided into content words and function words. Several reasons motivate this division. Function words and content words probably affect online processing in distinct ways. This is evidenced, in part, by differential effects of unexpectedness that have been observed in both the eye-tracking record (e.g. [36]), and in EEG [33]. Separating these two word categories is further motivated in the present study by the different average length between categories. As shown in figure 1, function words average about 0.2 s long in our stimulus; thus a single function word occupies about one theta cycle (4–8 Hz). By contrast, longer content words average about 0.4 s in length, which is about one delta cycle (1–4 Hz). As previous work postulates a systematic link between the size of a linguistic unit and

associated oscillatory responses (e.g. [27,28]), we test the prediction that these two word classes may have distinct effects on the theta- and delta-band responses, respectively.

### (c) Phase synchronicity and cross-frequency coupling

We assess phase synchronicity within and between three frequency bands: delta (1–4 Hz), theta (4–8 Hz) and gamma (30–50 Hz).

Within-band phase synchronicity is assessed using inter-trial phase clustering (ITPC a.k.a. phase-locking value or phase coherence). This value quantifies the uniformity of phases across trials. This value is higher when different epochs show similar phases, as is the case, for example, if the phase is reset at each word. If phase synchronicity plays a role in syntactic composition, we expect increased ITPC when words are completing larger numbers of phrases. This follows if phase synchronization mediates the processing of additional compositional structure, as predicted by the perceptual inference account of Martin [2]. ITPC is calculated for a given electrode and time-point following Cohen [37, p. 244]:

$$\text{ITPC} = \left| n^{-1} \sum_{r=1}^n e^{i\phi_r} \right|,$$

where  $e^{i\phi_r}$  provides the polar representation of phase angle  $\phi$  at frequency  $f$  for epoch  $r$ .

We assess CFC using two commonly applied measures: power–power correlation (P-P) and phase-amplitude coupling (PAC). P-P is simply the Pearson's correlation between power at two frequency bands. We compute this pairwise measure for each combination of frequency bins in our analysis: delta–theta, delta–gamma and theta–gamma. PAC quantifies the degree to which the phase of a lower-frequency oscillation affects the amplitude of a higher-frequency oscillation. We calculate this following Cohen [37, p. 413]:

$$\text{PAC} = \left| n^{-1} \sum_{r=1}^n a_r e^{i\phi_r} \right|,$$

where  $a_r$  is the power at word onset for epoch  $r$  at the higher of two frequency bands, and  $e^{i\phi_r}$  is the polar representation of phase for the lower of two frequency bands. Similar to ITPC, if increased composition is mediated by CFC, for example, between delta and gamma bands (cf. [2]), we expect to find higher P-P and/or PAC values at words that complete a larger number of phrases.

These two measures of CFC have been linked with distinct neurobiological mechanisms [25,38,39]. P-P coupling is more likely to be detectable if there is a direct relationship between activation in one cell assembly with another, for example, if power in a network tracking the speech envelope is amplified by power in a network involved in structural inference as the sentence unfolds. PAC, on the other hand, requires that the phase of one network

is affected by the power of the other. This latter might hold, for example, when a lower-frequency signal is used to synchronize (perhaps by resetting) a higher-frequency cell assembly.

Several of these measures, especially ITPC and PAC, are sensitive to the number of epochs entered into the calculation. As this value varies across different phrase bins for our naturalistic stimulus (table 1), we normalize each measure to allow for comparison. To do this, we compute a ‘null’ variant of each measure by shifting the phrase bin assigned to each word by 100 epochs. This removes any potential relationship between phrase count and any of the phase measures. We then recompute the target measure within each phrase bin. These offer an estimate of what we would expect of each measure under the null hypothesis, taking into account the different numbers of epochs per phrase bin and word category. Finally, we compute a z-score by subtracting the mean null value from the target values and dividing the result by the standard deviation of the null variant.

We carry out two additional analyses to complement these assessments of phase. First, we quantify power across phrase bins by simply averaging the single-trial power estimates within each frequency band. Second, test the relationship between phrase completion and the evoked signal using single-trial regression in the following way (e.g. [33,40–42]). Starting with the same artefact-cleaned epochs, described above, the data are low-pass filtered at 40 Hz and then subject to a linear regression, by participant, testing the scalp voltage at each electrode and time-point (0–1 s after word onset) as a function of the number of completed phrases as well as a set of control variables: sound power at word onset, epoch order and word at the target word, the previous word, and following word frequency (HAL corpus, log transformed). A control regression is also carried out per participant in which the rows of the design matrix are randomly permuted. Single-subject regression coefficients are pooled at the group level using the cluster-based permutation test of Maris & Oostenveld [43]. This test returns clusters of electrodes and time-points where the test coefficient for phrase completions is reliably different than the matched term from the control regression. Following Pallier *et al.* [8], we conduct this analysis on the count of phrase completions and also the  $\log_{10}$  transformation of that count.

#### (d) Statistical analysis

The z-scored ITPC, P-P, PAC and power values are statistically analysed using Bayesian mixed-effects linear regression, implemented in the STAN programming language using the *brms* package in R. For each measure, we model the z-scored value as a function of the number of phrases (one, two or three+) frequency band (delta, theta or gamma; or frequency band-pair for CFC), and word category (content, function) and all higher-order interactions between these terms. We include a full by-participants random-effects structure that includes each of these terms and their interactions.

Our research question concerns whether phase relationships change systematically as a function of syntactic composition. But, the direction or magnitude of such changes are underspecified by current theories. We statistically test for any such relationship in two stages. First, we conduct a joint test for a main effect or any higher-order interaction involving the number of phrases. This is done with  $\chi^2$  tests implemented in the *emmeans* package. Given the exploratory nature of these tests, we set  $\alpha = 0.01$ . To unpack any higher-order interactions, we conduct *post-hoc* pairwise tests for the effect of phrases within each frequency band and word category. These are evaluated with Bayes factors (BF), calculated with the *bayestestR* package, that indicate the relative likelihood of the data under the target hypothesis—that there will be difference in phase synchronization—compared to the point null hypothesis that there is no such difference. We focus our attention on differences with a  $\text{BF} > 3$  (‘positive’

to ‘strong’ evidence for the target hypothesis; [44]). The estimated effect size and BF value for *every* pairwise comparison is given in appendix A.

## 3. Results

### (a) Power

We observe little difference in power as a function of syntactic phrases (figure 2a). The statistical analysis suggests a reliable interaction between word category and phrases ( $\chi^2_2 = 5.96$ ,  $p = 0.003$ ) and also a three-way interaction between those terms and frequency ( $\chi^2_4 = 3.98$ ,  $p = 0.003$ ). But, *post-hoc* tests for power differences as a function of phrases within each frequency band did not show any pairwise differences with positive or strong support (all  $\text{BF} < 2$ ). Accordingly, we do not see substantial statistical support for a relationship between phrase count and power in these three frequency bands. The full set of pairwise comparisons is given in appendix A, table 2.

### (b) Inter-trial phase clustering

ITPC increases with more phrases for function words but decreases for content words (figure 2b). This pattern is the strongest in the delta band and is supported by interactions between phrases and word category ( $\chi^2_2 = 39.3$ ,  $p < 0.001$ ) and phrases and frequency band ( $\chi^2_4 = 3.4$ ,  $p = 0.009$ ). Pairwise comparisons indicate strong support for a difference in ITPC for function words in the delta band (one-versus-three+ phrases,  $\text{BF} = 4513.2$ ; two-versus-three+ phrases,  $\text{BF} = 47.0$ ) and in the theta band (one-versus-three+ phrases,  $\text{BF} = 5836.6$ ; two-versus-three+ phrases,  $\text{BF} = 295.96$ ). Support is also found for differences for content words, where more phrases lead to lower ITPC (delta band, one-versus-two  $\text{BF} = 14.4$ ; one-versus-three+  $\text{BF} = 72.7$ ; gamma band, one-versus-three+  $\text{BF} = 10.01$ ). The full set of pairwise comparisons is given in appendix A, table 3.

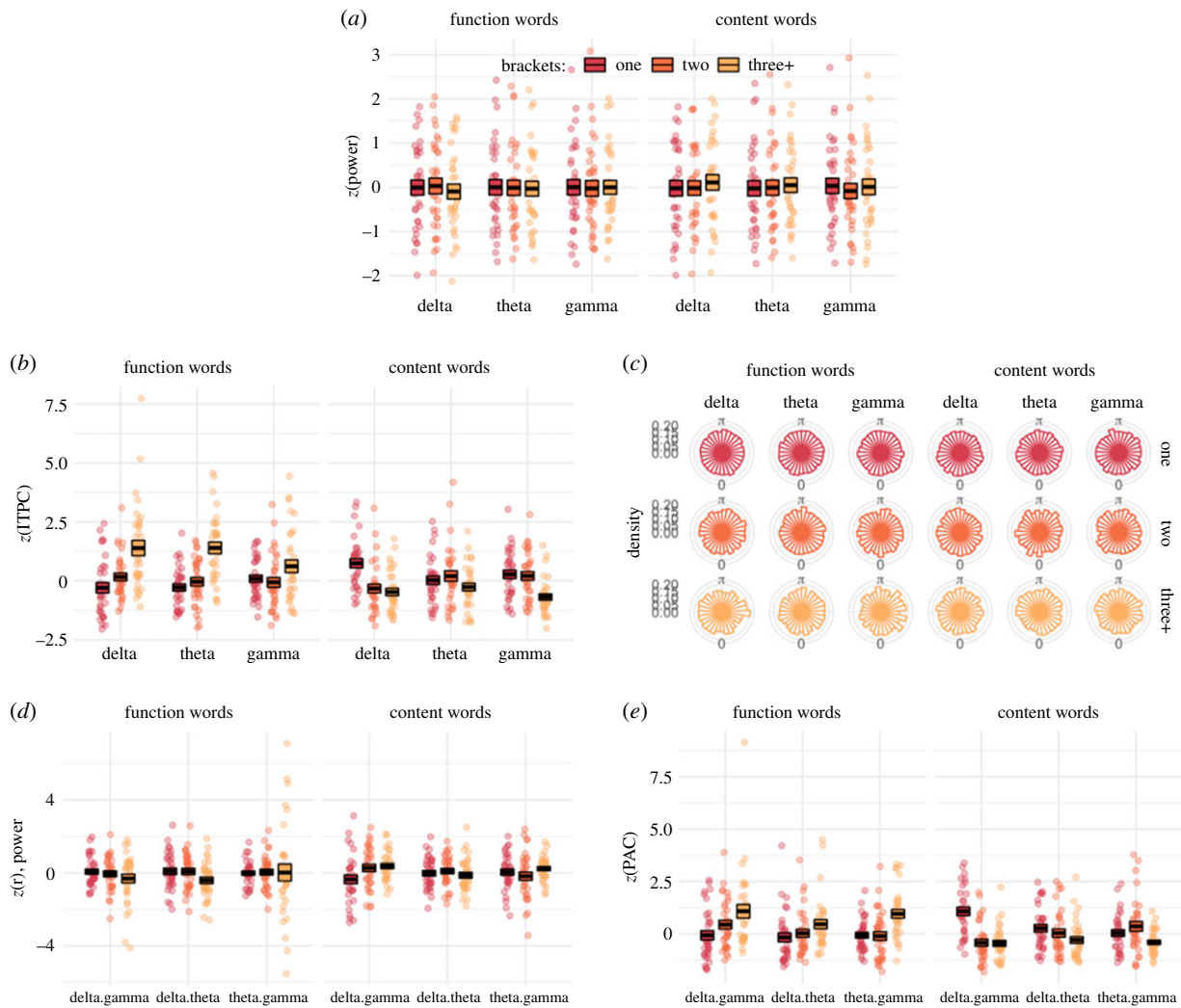
Polar histograms are shown in figure 2c that pool single-trial phase angles from all participants. These appear to show relative uniformity in the distribution of phase angles even for higher numbers of syntactic phrases. This appearance of uniformity seems to be at odds with the differences in ITPC indicated in figure 2b, but this summary at the group level masks individual differences. The electronic supplementary material, figure S1 shows that numerous individual participants show increased phase consistency for more phrases, but the peak phase differs across participants, which in turn leads to the more uniform distribution of phases at the group level.

### (c) Power–power coupling

There is a small trend for increased correlation between power from delta to gamma as the number of syntactic phrases increases for content words (figure 2d). But, this trend is not statistically reliable and there is no evidence for reliable pairwise differences in P-P coupling. The full set of pairwise comparisons is given in appendix A, table 4.

### (d) Phase-amplitude coupling

We observe a decrease in PAC at higher phrase counts for content words, as shown on the right-hand side in figure 2e. This decrease contrasts with an increase in PAC for function words with higher phrase counts. These observations are supported



**Figure 2.** (a) Power at word onset in three frequency bands, divided into function and content words. Points indicate individual averages per participant, and boxes indicate the group mean  $\pm$  1 s.e. (b) Inter-trial phase consistency (ITPC) across three frequency bands for function and content words. (c) Circular histograms of single-trial phase in each of three phrase bins (rows) for three frequency bands separated by function and content words (columns). (d–e) CFC as measured by P-P correlations (d) or PAC (e) across three frequency pairs.

by a statistical interaction between phrases and word category ( $\chi^2_2 = 27.8$ ,  $p < 0.001$ ) and a three-way interaction between phrases, word category and frequency ( $\chi^2_4 = 6.4$ ,  $p < 0.001$ ). Post-hoc pairwise tests for content words strongly support a decrease in PAC as a function of phrase between delta and gamma (one-versus-three + BF = 7355.2; one-versus-two BF = 1702.8). For function words, also between delta and gamma, there is positive support for increased PAC with higher phrase counts (one-versus-three + BF = 62.2); this also obtained between theta and gamma (function words one-versus-three + BF = 46.1; two-versus-three + BF = 23.9). The full set of pairwise comparisons is given in appendix A, table 5.

### (e) Evoked amplitude

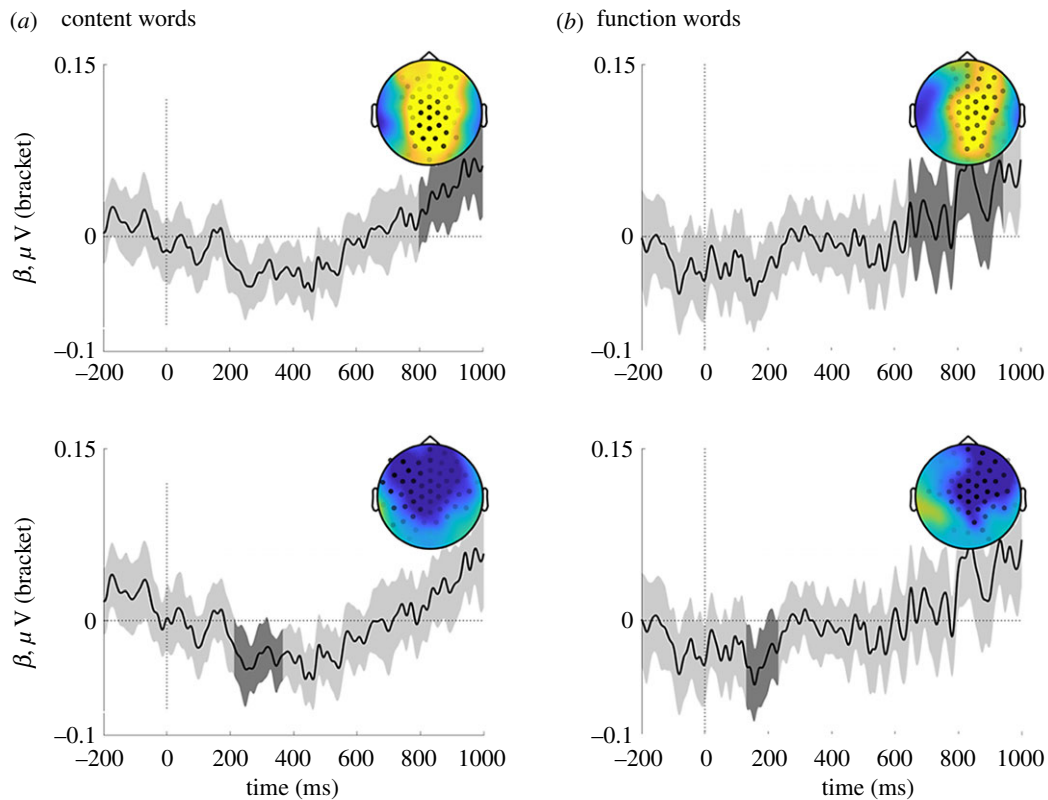
We complement our primary analyses of phase synchronization with an analysis of how the evoked signal changes as a function of phrase count. While we needed to bin the single-trial data for the phase-based analyses above, no such constraint holds for the evoked signal. Accordingly, we compute a regression between phrase counts and the evoked signal across all electrodes and time-points.

Two reliable effects are observed: a late positivity associated with more phrases found primarily on central electrodes

(figure 2, top), and an earlier negativity associated with more phrases found primarily on central-anterior electrodes (figure 2, bottom). Aside from slight changes in topography and precise timing, these results are the same for both content words and for function words. A similar pattern is found when phrase count is log transformed, following Pallier *et al.* [8] (electronic supplementary material, figure S2).

## 4. Discussion

Using EEG data from a naturalistic story-listening experiment, we explore the relationship between a measure of linguistic composition and phase synchronicity across three bands that have been implicated in sentence-level processing: delta, theta and gamma. Broadly, we see evidence for a systematic relationship between changes in phase synchronicity and CFC that vary as a function of the number of phrases completed by a word. But, the observed patterns neither follow a simple ‘more composition equals more synchronization’ account, nor is there a tight alignment between the ‘size’ of a particular linguistic unit and the frequency bands showing changes in synchronicity (e.g. between content words and slower delta waves). We unpack these observations below.



**Figure 3.** Regression event-related potential plots of spatio-temporal clusters that show a reliable linear effect of brackets on evoked EEG data from (a) content words or (b) function words. Two statistically reliable clusters are observed in both sets of data. In all panels, black traces and grey shading indicate the grand-averaged regression coefficient ( $\pm CI_{95\%}$ ) from electrodes contributing to that effect; insets show those electrodes along with a topographic plot of activity averaged across the cluster's time-window (dark-grey shading). The top plots for both content and function words show a reliable cluster of positive-going activity over central posterior-electrodes starting about 700–800 ms after word onset ( $p = 0.009$ ,  $p = 0.001$ , respectively); the bottom plot shows a reliable negative-going cluster over anterior electrodes starting around 200 ms ( $p = 0.013$ ,  $p = 0.025$ , respectively). (Online version in colour.)

Evidence for increased phase synchronicity with increased composition was observed for function words in several measures: ITPC in the delta band (with similar trends in theta and gamma), and also between delta and gamma bands as measured with PAC. But, both ITPC and PAC showed decreases in the same bands when considering content words. These patterns are not consistent with a simple word-by-word mapping between phrase completion and phase synchronization. Concerning content words, there is a wealth of psycholinguistic data which suggests that different mechanisms might be engaged at the end of large phrases and sentences than during other moments in structure building and sentence processing (e.g. [45–47]), and that these differences affect event-related brain potentials (e.g. [48]). Some of these theories link phrase boundaries with increased discourse-level integration, which may be producing the difference we observed here because content words that complete three or more phrases include a number of clause and sentence final items. A decrease in ITPC at sentence boundaries could also be consistent with the notion that activity builds up as a phrase or sentence is structured but does not persist in the same way across phrase and sentence boundaries, as reported for power in the gamma band in Nelson *et al.* [11]. More broadly, various sentence-related phenomena occur more often at clause boundaries. These include syntactic and semantic operations associated with identifying and interpreting phrases, but also other interpretive processes like the resolution of referential dependencies. Further, our annotations do not allow us to distinguish the perceptual cues that

might correlate with these syntactic and semantic properties, such as acceleration or deceleration of speech rate and segment lengthening [49,50].

The effects for phase synchronization indicated by changes in ITPC and PAC do not pattern, in a simple way, with the evoked response for phrase completion. Our control evoked analysis identified two distinct components that correlated with phrases: an early anterior negativity and a later central positivity. These components were observed on both content words and function words (figure 3), and were also qualitatively similar when treating phrase counts on a linear or log scale (see the electronic supplementary material, figure S2). Both of these evoked components align well with prior literature on event-related potentials associated with syntactic processing (e.g. the LAN, or the P600; see [51] for an overview). These complexities point towards a need in future work to develop annotations that tease apart these distinct phenomena.

One question that follows from the literature is whether the mechanism by which the pattern of ITPC arises is phase resetting as linguistic structure is being inferred or built [52]. Our answer from the present analysis is: not obviously. In figure 2c, we see a uniformity across bands that does not provide evidence for coordination around any one phase angle. It is still logically possible that phase resetting is occurring, but our results indicate that if it is, then participants must unsystematically vary in the phase angle expressing the change that the ITPC results reflect. Indeed, such individual differences in preferred phase is at least tentatively

supported by per-participant phase-angle distributions shown in the electronic supplementary material, figure S1. No account of sentence processing makes any statements about individual variation in neural oscillations. However, the finding that that phase consistency across trials within participants does not extend to consistency across individuals is largely in line with an analysis-by-synthesis view. In an architecture like that of Martin [2], there may be differences across individuals in the time course of perceptual inference or the invocation of a higher-level structure via the LFP. Such an account aligns with observations linking, for example, behavioural outcomes with differences in the peak alpha rhythm between individuals (e.g. [53]). Further, measurable differences between individuals could be driven by differences in brain morphology alone.

Another question from the literature concerns links between the size of linguistic units (phonemes, syllables, phrases) and the oscillatory responses that are associated with cognitive operations over those units [27–29]. There are at least two ways such a link, if it were to extend to higher levels of syntactic composition, might be born out in our analysis. First, we might expect to find that function words, which are shorter and thus occur primarily on the same timescale as theta oscillations, might associate primarily with theta activity. Likewise, content words, which occur more on the same timescale as delta oscillations, might thus associate primarily with delta activity. We do not see evidence for any such links between word category and specific frequency bands: both function words and content words show the most statistically reliable effects in the delta band, with trends observed in all three bands that we investigate here.

A related prediction that we examine is whether words that complete larger phrases are associated with activity at lower, longer timescale oscillations. This prediction comes from observations like those of Nelson *et al.* [11], already discussed above, where the build-up of high-gamma band activity is observed within phrases but not between phrases of different lengths. Our data are consistent with such a hypothesis in a limited way: there is indeed increased low-frequency delta phase synchronicity at function words that complete a larger number of phrases (see the left-hand side of figure 2*b*). However, our results again do not paint a simple picture: there is also increased phase synchronicity in the theta band for such words; and, more strikingly, increased coupling between the gamma and delta bands as a function of phrase count (figure 2*e*). This pattern of increased synchronicity does not align with a simplistic mapping between the size of a linguistic unit and oscillatory neural responses associated with that unit.

Another way to frame these observations is that the problem the brain faces when it must create linguistic structure from speech or sign is much more complex than the discrete frequency band analyses that has sometimes been assumed in the field (and, indeed, adopted in this study). Syllables might occur at a timescale that corresponds to the theta rate, but syllables can be morphemes or words, and can cue complex relations in information and discourse structure; words certainly span theta and delta (e.g. in our dataset, the correlation between delta for content words and theta for function words), as do phrases. At some times, a phrase is a single word lasting a few hundred milliseconds, and at other times, it is tens of words, and lasts a few seconds.

At the extremes, entire clauses may be elided with silence. The take-away message is that ‘size’ or length in time, and ergo frequency, are not directly predictive of linguistic structure and content. Consequently, it may not be fruitful to probe for functional data patterns which assume that fixed, structure-related computations can be associated with a given band or coupling pattern.

We are far from the first to point out that we have only a limited understanding of linking hypothesis between our measurements and the compositional processes we are trying to do inference about (e.g. [5,54]). Recognizing this challenge, we attempt to meet a low standard of rhetoric: we assert that, minimally, phase synchronization reflects some flow of information between distributed populations functioning as neuronal assemblies. This minimalistic hypothesis, from a cue-integration and perceptual inference perspective [2], is that as more linguistic representations and structures are inferred or transformed from sensory signals, the more synchronization (at certain time points) and desynchronization (at other time points) occurs. The ITPC and PAC patterns we observe are consistent with the synchronization of excitatory and inhibitory cycles across assemblies, synchronization which increases with the unfolding structure and concomitant accrual of compositional meaning. Larger and more distributed networks oscillate at lower frequencies [17], this is consistent with an interpretation where the formation of higher-level linguistic structures as indicated by phrase count draws upon distributed networks operating in phase. Speculatively, phase synchronization may reflect the transfer of information between ‘reader-integrator’ ensembles [13] that detect or infer [2,13] linguistic structure and meaning.

## 5. Conclusion

Our results indicate that neural oscillations show changes in phase synchronization within and between frequency bands in response to linguistic compositional demands. These results highlight the importance of neuronal synchronization and desynchronization to levels of linguistic structure. Current theoretical models hold that desynchronization and inhibition are key in the function of computational models that learn and represent a predicate calculus, (see DORA; [3,4,21,22]), and for maintaining variable-value independence (see [24] for discussion). In this way, our results suggest the importance of inhibition and cortical interneurons in controlling the flow of information in across assemblies to derive compositional structure and meaning.

**Data accessibility.** The raw data we analyse are openly available at <http://dx.doi.org/10.7302/Z29C6VNH>. We have provided the specific annotations used in this analysis and also the derived quantities (phase, power) as electronic supplementary material.

**Authors’ contributions.** A.E.M. developed the research question, and contributed to the data analysis plan and manuscript writing. J.R.B. developed and implemented the data analysis and contributed to the manuscript writing.

**Competing interests.** We declare we have no competing interests.

**Funding.** This study was funded, in part, by grant no. 1607251 from the National Science Foundation of the United States to J.R.B. A.E.M. was supported by the Max Planck Research Group ‘Language and Computation in Neural Systems’ and by the Netherlands Organization for Scientific Research (grant 016.Vidi.188.029).

**Table 2.** Pairwise estimates for differences in power as a function of phrase count along with 95% credibility intervals and Bayes factors.

word category	frequency band	comparison	estimate	lower bound	upper bound	Bayes factor
functional	delta	one–two	−0.038	−0.094	0.021	0.022
		one–three+	0.088	0.001	0.159	0.142
		two–three+	0.126	0.050	0.207	0.798
	theta	one–two	0.012	−0.045	0.070	0.006
		one–three+	0.031	−0.063	0.124	0.011
		two–three+	0.021	−0.082	0.112	0.006
	gamma	one–two	0.028	−0.043	0.103	0.009
		one–three+	0.003	−0.116	0.117	0.012
		two–three+	−0.025	−0.152	0.096	0.009
lexical	delta	one–two	−0.003	−0.065	0.056	0.006
		one–three+	−0.130	−0.205	−0.054	1.467
		two–three+	−0.127	−0.203	−0.054	0.972
	theta	one–two	−0.015	−0.077	0.045	0.004
		one–three+	−0.073	−0.156	0.010	0.024
		two–three+	−0.057	−0.146	0.027	0.009
	gamma	one–two	0.117	0.036	0.189	0.397
		one–three+	0.023	−0.075	0.134	0.007
		two–three+	−0.095	−0.204	0.022	0.020

**Table 3.** Pairwise comparisons for differences in ITPC as a function of phrase count along with 95% credibility intervals and Bayes factors. (Bayes factors with values greater than 3 are indicated in italics.)

word category	frequency band	comparison	estimate	lower bound	upper bound	Bayes factor
functional	delta	one–two	−0.371	−0.877	0.129	0.209
		one–three+	−1.577	−2.151	−1.056	<i>4513.206</i>
		two–three+	−1.205	−1.805	−0.637	<i>47.038</i>
	theta	one–two	−0.258	−0.761	0.295	0.090
		one–three+	−1.680	−2.229	−1.102	<i>5836.647</i>
		two–three+	−1.422	−2.017	−0.840	<i>295.966</i>
	gamma	one–two	0.119	−0.421	0.624	0.055
		one–three+	−0.549	−1.149	−0.008	0.368
		two–three+	−0.671	−1.260	−0.048	0.400
lexical	delta	one–two	0.987	0.500	1.545	<i>14.448</i>
		one–three+	1.146	0.606	1.736	<i>72.657</i>
		two–three+	0.146	−0.412	0.741	0.040
	theta	one–two	−0.156	−0.744	0.405	0.041
		one–three+	0.307	−0.281	0.866	0.057
		two–three+	0.463	−0.148	1.115	0.076
	gamma	one–two	0.096	−0.426	0.672	0.038
		one–three+	1.000	0.446	1.595	<i>10.010</i>
		two–three+	0.904	0.296	1.547	1.218



**Table 4.** Pairwise comparisons for differences in P-P as a function of phrase count along with 95% credibility intervals and Bayes factors.

word category	frequency band	comparison	estimate	lower bound	upper bound	Bayes factor
functional	delta–gamma	one–two	0.082	–0.359	0.506	0.077
		one–three+	0.343	–0.098	0.763	0.216
		two–three+	0.265	–0.203	0.687	0.084
	delta–theta	one–two	0.011	–0.442	0.458	0.045
		one–three+	0.501	0.062	0.969	0.532
		two–three+	0.489	–0.004	0.954	0.227
	theta–gamma	one–two	–0.048	–0.479	0.421	0.044
		one–three+	–0.011	–0.825	0.789	0.078
		two–three+	0.035	–0.812	0.803	0.050
lexical	delta–gamma	one–two	–0.606	–1.096	–0.104	0.844
		one–three+	–0.698	–1.158	–0.211	2.743
		two–three+	–0.096	–0.613	0.391	0.031
	delta–theta	one–two	–0.122	–0.640	0.386	0.033
		one–three+	0.100	–0.354	0.597	0.033
		two–three+	0.229	–0.368	0.756	0.032
	theta–gamma	one–two	0.213	–0.335	0.777	0.047
		one–three+	–0.202	–0.801	0.444	0.046
		two–three+	–0.411	–1.114	0.279	0.057

**Table 5.** Pairwise comparisons for differences in PAC as a function of phrase count along with 95% credibility intervals and Bayes factors. (Bayes factors with values greater than 3 are indicated in italics.)

word category	frequency band	comparison	estimate	lower bound	upper bound	Bayes factor
functional	delta–gamma	one–two	–0.426	–0.933	0.069	0.300
		one–three+	–1.058	–1.532	–0.482	<i>58.450</i>
		two–three+	–0.635	–1.204	–0.066	0.593
	delta–theta	one–two	–0.219	–0.728	0.312	0.073
		one–three+	–0.657	–1.149	–0.085	0.802
		two–three+	–0.425	–0.969	0.168	0.102
	theta–gamma	one–two	0.013	–0.505	0.482	0.049
		one–three+	–1.049	–1.555	–0.510	<i>42.104</i>
		two–three+	–1.053	–1.628	–0.513	<i>22.727</i>
lexical	delta–gamma	one–two	1.435	0.944	1.968	<i>1703.729</i>
		one–three+	1.462	0.963	2.011	<i>6501.473</i>
		two–three+	0.022	–0.543	0.543	0.033
	delta–theta	one–two	0.245	–0.293	0.777	0.050
		one–three+	0.578	0.063	1.130	0.336
		two–three+	0.330	–0.219	0.925	0.043
	theta–gamma	one–two	–0.287	–0.894	0.331	0.058
		one–three+	0.461	–0.046	1.032	0.141
		two–three+	0.752	0.113	1.385	0.398

## Endnotes

<sup>1</sup>However see Martin [2] for an argument that they must be based on, composed of, or at least bounded by, summation and divisive normalization, two operations neurons have been shown to do [6].

<sup>2</sup>Our study does not speak directly to the debate about whether neural oscillations are causal or epiphenomenal to neural computation. However, implicit in the theoretical background for this

study (e.g. [2,4]) is the idea that cycles of excitation and inhibition/refraction are what carry out the computations in question [7].

<sup>3</sup>Which, interestingly, further could be said to descend to below phonemic and phonetic levels given recent results suggesting that auditory cortex might represent a generative model of articulator dynamics for constructing linguistic representations [30].

<sup>4</sup>Stimuli and data are available for download at: <https://dx.doi.org/10.7302/Z29C6VNH>.

## References

- Partee B. 1975 Montague grammar and transformational grammar. *Linguistic Inquiry* **6**, 203–300.
- Martin AE. 2016 Language processing as cue integration: grounding the psychology of language in perception and neurophysiology. *Front. Psychol.* **7**, 120. (doi:10.3389/fpsyg.2016.00120)
- Martin AE, Doumas LA. 2017 A mechanism for the cortical computation of hierarchical linguistic structure. *PLoS Biol.* **15**, e2000663. (doi:10.1371/journal.pbio.2000663)
- Martin AE, Doumas LA. 2019 Predicate learning in neural systems: using oscillations to discover latent structure. *Curr. Opin. Behav. Sci.* **29**, 77–83. (doi:10.1016/j.cobeha.2019.04.008)
- Poeppel D. 2012 The maps problem and the mapping problem: two challenges for a cognitive neuroscience of speech and language. *Cogn. Neuropsychol.* **29**, 34–55. (doi:10.1080/02643294.2012.710600)
- Carandini M, Heeger DJ. 2012 Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* **13**, 51–62. (doi:10.1038/nrn3136)
- Buzsáki G. 2006 *Rhythms of the brain*. Oxford, UK: Oxford University Press.
- Pallier C, Devauchelle A-D, Dehaene S. 2011 Cortical representation of the constituent structure of sentences. *Proc. Natl Acad. Sci. USA* **108**, 2522–2527.
- Brennan J, Nir Y, Hasson U, Malach R, Heeger DJ, Pylkkänen L. 2012 Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain Lang.* **120**, 163–173. (doi:10.1016/j.bandl.2010.04.002)
- Brennan JR, Pylkkänen L. 2017 MEG evidence for incremental sentence composition in the anterior temporal lobe. *Cogn. Sci.* **41**, 1515–1531.
- Nelson MJ *et al.* 2017 Neurophysiological dynamics of phrase-structure building during sentence processing. *Proc. Natl Acad. Sci. USA* **114**, E3669–E3678. (doi:10.1073/pnas.1701590114)
- Hale J, Dyer C, Kuncoro A, Brennan JR. 2018 Finding syntax in human encephalography with beam search. arXiv 1806.04127. (<http://arxiv.org/abs/1806.04127>)
- Buzsáki G. 2010 Neural syntax: cell assemblies, synapsembles, and readers. *Neuron* **68**, 362–385. (doi:10.1016/j.neuron.2010.09.023)
- Fries P. 2015 Rhythms for cognition: communication through coherence. *Neuron* **88**, 220–235. (doi:10.1016/j.neuron.2015.09.034)
- Gray CM, König P, Engel AK, Singer W. 1989 Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* **338**, 334–337. (doi:10.1038/338334a0)
- Klimesch W, Sauseng P, Hanslmayr S, Gruber W, Freunberger R. 2007 Event-related phase reorganization may explain evoked neural dynamics. *Neurosci. Biobehav. Rev.* **31**, 1003–1016. (doi:10.1016/j.neubiorev.2007.03.005)
- Sauseng P, Klimesch W. 2008 What does phase information of oscillatory brain activity tell us about cognitive processes? *Neurosci. Biobehav. Rev.* **32**, 1001–1013. (doi:10.1016/j.neubiorev.2008.03.014)
- von der Malsburg C, Buhmann J. 1992 Sensory segmentation with coupled neural oscillators. *Biol. Cybern.* **67**, 233–242. (doi:10.1007/BF00204396)
- Hummel JE, Holyoak KJ. 1997 Distributed representations of structure: a theory of analogical access and mapping. *Psychol. Rev.* **104**, 427–466. (doi:10.1037/0033-295X.104.3.427)
- Hummel JE, Holyoak KJ. 2003 A symbolic-connectionist theory of relational inference and generalization. *Psychol. Rev.* **110**, 220–264. (doi:10.1037/0033-295X.110.2.220)
- Doumas LA, Hummel JE, Sandhofer CM. 2008 A theory of the discovery and predication of relational concepts. *Psychol. Rev.* **115**, 1–43. (doi:10.1037/0033-295X.115.1.1)
- Doumas LA, Martin AE. 2018 Learning structured representations from experience. *Psychol. Learn. Motiv.* **69**, 165–203. (doi:10.1016/bs.plm.2018.10.002)
- Hummel JE. 2011 Getting symbols out of a neural architecture. *Connect. Sci.* **23**, 109–118. (doi:10.1080/09540091.2011.569880)
- Martin AE, Doumas LAA. 2019 Tensors and compositionality in neural systems. *Phil. Trans. R. Soc. B* **375**, 20190306. (doi:10.1098/rstb.2019.0306)
- Canolty RT, Knight RT. 2010 The functional role of cross-frequency coupling. *Trends Cogn. Sci.* **14**, 506–515. (doi:10.1016/j.tics.2010.09.001)
- Lisman JE, Jensen O. 2013 The theta-gamma neural code. *Neuron* **77**, 1002–1016. (doi:10.1016/j.neuron.2013.03.007)
- Arnal LH, Poeppel D, Giraud AL. 2015 Temporal coding in the auditory cortex. In *Handbook of clinical neurology*, vol. 129 (eds GG Celesia, G Hickok), pp. 85–98. Amsterdam, The Netherlands: Elsevier.
- 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. (doi:10.1038/nn.4186)
- Giraud AL, Poeppel D. 2012 Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* **15**, 511–517. (doi:10.1038/nn.3063)
- Anumanchipalli GK, Chartier J, Chang EF. 2019 Speech synthesis from neural decoding of spoken sentences. *Nature* **568**, 493–498. (doi:10.1038/s41586-019-1119-1)
- Kaufeld G, Ravensschlag A, Meyer AS, Martin AE, Bosker HR. In press. Knowledge-based and signal-based cues are weighted flexibly during spoken language comprehension. *J. Exp. Psychol. Learn. Mem. Cogn.* (doi:10.1037/xlm0000744)
- Martin AE. 2018 Cue integration during sentence comprehension: electrophysiological evidence from ellipsis. *PLoS ONE* **13**, e0206616. (doi:10.1371/journal.pone.0206616)
- Brennan JR, Hale JT. 2019 Hierarchical structure guides rapid linguistic predictions during naturalistic listening. *PLoS ONE* **14**, e0207741.
- Brennan JR, Stabler EP, Van Wagenen SE, Luh W-M, Hale JT. 2016 Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain Lang.* **157–158**, 81–94.
- Marcus M, Kim G, Marcinkiewicz M, MacIntyre R, Bies A Ferguson M, Katz K, Schasberger B. 1994 The Penn Treebank: annotating predicate argument structure. ARPA human language technology workshop, pp. 114–119. Philadelphia, PA: Department of Computer and Information Science, University of Pennsylvania.
- Roark B, Bachrach A, Cardenas C, Pallier C. 2009 Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. In *Proc. of the Conf. on Empirical Methods in Natural Language Processing (EMNLP)*, Singapore, pp. 324–333. Stroudsburg, PA: Association for Computational Linguistics.
- Cohen MX. 2014 *Analyzing neural time series data*. Cambridge, MA: MIT Press.
- Helfrich RF, Herrmann CS, Engel AK, Schneider TR. 2016 Different coupling modes mediate cortical cross-frequency interactions. *Neuroimage* **140**, 76–82. (doi:10.1016/j.neuroimage.2015.11.035)
- Watrous AJ, Fell J, Ekstrom AD, Axmacher N. 2015 More than spikes: common oscillatory mechanisms

- for content specific neural representations during perception and memory. *Curr. Opin Neurobiol.* **31**, 33–39. (doi:10.1016/j.conb.2014.07.024)
40. Hauk O, Davis MH, Ford M, Pulvermuller F, Marslen-Wilson WD. 2006 The time course of visual word recognition as revealed by linear regression analysis of ERP data. *NeuroIm-age* **30**, 1383–1400.
41. Solomyak O, Marantz A. 2009 Lexical access in early stages of visual word processing: a single-trial correlational MEG study of heteronym recognition. *Brain Lang.* **108**, 191–196. (doi:10.1016/j.bandl.2008.09.004)
42. Lewis G, Poeppel D. 2014 The role of visual representations during the lexical access of spoken words. *Brain Lang.* **134**, 1–10. (doi:10.1016/j.bandl.2014.03.008)
43. Maris E, Oostenveld R. 2007 Nonparametric statistical testing of EEG- and MEG- data. *J. Neurosci. Meth.* **164**, 177–190.
44. Kass RE, Raftery AE. 1995 Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795. (doi:10.1080/01621459.1995.10476572)
45. Hirotani M, Frazier L, Rayner K. 2006 Punctuation and intonation effects on clause and sentence wrap-up: evidence from eye movements. *J. Mem. Lang.* **54**, 425–443. (doi:10.1016/j.jml.2005.12.001)
46. Rayner K, Kambe G, Duffy SA. 2000 The effect of clause wrap-up on eye movements during reading. *Quart. J. Exp. Psychol. Sect. A* **53**, 1061–1080. (doi:10.1080/713755934)
47. Stowe LA, Kaan E, Sabourin L, Taylor RC. 2018 The sentence wrap-up dogma. *Cognition* **176**, 232–247. (doi:10.1016/j.cognition.2018.03.011)
48. Hagoort P. 2003 Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *J. Cogn. Neurosci.* **15**, 883–899. (doi:10.1162/089892903322370807)
49. Turk AE, Shattuck-Hufnagel S. 2007 Multiple targets of phrase-final lengthening in American English words. *J. Phonet.* **35**, 445–472. (doi:10.1016/j.wocn.2006.12.001)
50. Wightman CW, Shattuck-Hufnagel S, Ostendorf M, Price PJ. 1992 Segmental durations in the vicinity of prosodic phrase boundaries. *J. Acoust. Soc. Am.* **91**, 1707–1717. (doi:10.1121/1.402450)
51. Swaab TY, Ledoux K, Camblin CC, Boudewyn MA. 2012 Language-related ERP components. In *The Oxford handbook of event-related potential components* (eds SJ Luck, ES Kappenman), pp. 397–439. Oxford, UK: Oxford University Press.
52. 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. (doi:10.1111/ejn.13748)
53. Gulbinaite R, van Viegen T, Wieling M, Cohen MX, VanRullen R. 2017 Individual alpha peak frequency predicts 10 Hz flicker effects on selective attention. *J. Neurosci.* **37**, 10 173–10 184. (doi:10.1523/JNEUROSCI.1163-17.2017)
54. Krakauer JW, Ghazanfar AA, Gomez-Marín A, MacIver MA, Poeppel D. 2017 Neuroscience needs behavior: correcting a reductionist bias. *Neuron* **93**, 480–490. (doi:10.1016/j.neuron.2016.12.041)