UNIVERSITY OF BIRMINGHAM
University of Birmingham Research at Birmingham

Language prediction Is reflected by coupling between frontal gamma and posterior alpha oscillations

Wang, Ling; Hagoort, Peter; Jensen, Ole

DOI:
10.1162/jocn_a_01190

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Wang, L, Hagoort, P & Jensen, O 2018, 'Language prediction Is reflected by coupling between frontal gamma and posterior alpha oscillations', *Journal of Cognitive Neuroscience*, vol. 30, no. 3, a01190, pp. 432-447. https://doi.org/10.1162/jocn_a_01190

Link to publication on Research at Birmingham portal

Publisher Rights Statement:
Checked for eligibility:30/04/2018
Language Prediction Is Reflected by Coupling between Frontal Gamma and Posterior Alpha Oscillations
Lin Wang, Peter Hagoort, and Ole Jensen
Journal of Cognitive Neuroscience 2018 30:3, 432-447
https://www.mitpressjournals.org/loi/jocn

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

•Users may freely distribute the URL that is used to identify this publication.
•Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
•User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
•Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 19. Apr. 2024

Language Prediction Is Reflected by Coupling between Frontal Gamma and Posterior Alpha Oscillations

Lin Wang[1,2,3], Peter Hagoort[1], and Ole Jensen[1,4]

Abstract

■ Readers and listeners actively predict upcoming words during language processing. These predictions might serve to support the unification of incoming words into sentence context and thus rely on interactions between areas in the language network. In the current magnetoencephalography study, participants read sentences that varied in contextual constraints so that the predictability of the sentence-final words was either high or low. Before the sentence-final words, we observed stronger alpha power suppression for the highly compared with low constraining sentences in the left inferior frontal cortex, left posterior temporal region, and visual word form area. Importantly, the temporal and visual word form area alpha power correlated negatively with left frontal gamma power for the highly constraining sentences. We suggest that the correlation between alpha power decrease in temporal language areas and left prefrontal gamma power reflects the initiation of an anticipatory unification process in the language network. ■

INTRODUCTION

Language comprehension requires the coordination of brain activity associated with perception on a very fast timescale. This might be facilitated by activating predicted input, thus giving language processing a head start (Kuperberg & Jaeger, 2016; Van Petten & Luka, 2012; but see Huettig & Mani, 2016, for counter arguments). Such a process would serve to support the unification of incoming words into sentence context (Hagoort, 2005, 2013). Although prediction long has been considered to play an important role for cognition (Clark, 2013; Bar, 2009; Friston & Kiebel, 2009; Rao & Ballard, 1999), little is known about the neural mechanisms supporting predictions in language processing.

Violation paradigms have been applied to study linguistic preactivation by measuring brain responses to unexpected words. For instance, context (e.g., "The day was breezy so the boy went out to fly…") was used to make a particular word (e.g., "kite") highly predictable (DeLong, Urbach, & Kutas, 2005). Presenting the incorrect article "an" instead of "a" elicited a larger N400 ERP component. Because the magnitude of the N400 relates to the violation of an expectation, this finding suggests that the sentence-final word was preactivated, then allowing for the article to be predicted. ERP effects reflecting predictions were also observed when syntactic (Szewczyk & Schriefers, 2013; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Moreno, & Kutas, 2004) and orthographic predictions were violated (Brothers, Swaab, & Traxler, 2015; Kim & Lai, 2012; Laszlo & Federmeier, 2011). Although these studies provide evidence for prediction during language processing, they do not speak to neuronal dynamics associated with the preactivation of anticipated words.

Predictions require that incoming sensory information is integrated with recent sentence context, but also knowledge in general. Therefore, the unification processing relies on the integration of memory and sensory information distributed in multiple brain regions. Neuronal oscillations have been shown to play an important role in coordinating distributed brain regions. Different frequency bands (theta, alpha, beta, and gamma) have been associated with various cognitive functions, such as working memory and long-term memory, attention, as well as different aspects of language processing (for a review, see Bastiaansen et al., 2012). Several studies have demonstrated anticipatory effects as reflected by alpha power modulation (Mayer, Schwiedrzik, Wibral, Singer, & Melloni, 2015; Spaak, Fonken, Jensen, & de Lange, 2015; Wöstmann, Herrmann, Wilsch, & Obleser, 2015; Strauß, Wöstmann, & Obleser, 2014; Foxe & Snyder, 2011; Rohenkohl & Nobre, 2011). These findings are consistent with the view that alpha suppression reflects engagement of task-relevant brain regions (Payne & Sekuler, 2014; Jensen & Mazaheri, 2010; Klimesch, Sauseng, & Hanslmayr, 2007). Beta oscillations have been proposed to relate to the top–down propagation of predictions to hierarchically lower processing levels (Friston, Bastos, Pinotsis, & Litvak, 2015; Lewis & Bastiaansen, 2015), albeit the empirical evidence is scarce

[1]Radboud University, [2]Institute of Psychology, Chinese Academy of Sciences, Beijing, China, [3]Harvard Medical School, [4]University of Birmingham

© 2017 Massachusetts Institute of Technology

Journal of Cognitive Neuroscience 30:3, pp. 432–447
doi:10.1162/jocn_a_01190

(Weiss & Mueller, 2012). In addition, low and middle gamma has been related to matching of prediction and bottom–up input, whereas high gamma has been associated with prediction error or integration effort in general (Lewis & Bastiaansen, 2015). Moreover, it has been shown that interactions between neuronal oscillations in different frequency bands serve a functional role to integrate information across different spatial and temporal scales (Fries, 2005; Engel, Fries, & Singer, 2001).

It remains unclear which neural mechanisms support language prediction. Only a few studies have directly investigated the neuronal dynamics associated with language preactivation. In a picture–noun matching task, Dikker and Pylkkänen (2013) found increased theta band activity in left temporal, ventromedial prefrontal, and visual cortex before a target word primed by highly predictive pictures. Two recent studies asked participants to name pictures in highly or low constraining sentence contexts. Before the naming, they found suppression in the alpha and beta bands over the left inferior frontal, left temporal, and bilateral ventral premotor cortex in the highly constraining contexts, which might reflect memory retrieval and motor preparation of spoken word production (Piai, Roelofs, Rommers, & Maris, 2015; Piai, Roelofs, & Maris, 2014). Also alpha power modulation was observed during sentence comprehension for strongly constraining contexts, indicating that the brain recruits domain-general preparation mechanisms in language prediction (Rommers, Dickson, Norton, Wlotko, & Federmeier, 2016). Recently, Park, Ince, Schyns, Thut, and Gross (2015) found that top–down predictions supported by frontal cortex modulate speech–brain coupling in auditory cortex during intelligible speech perception. Overall, the research on oscillations involved in language prediction is sparse, and thus, oscillatory dynamics in different frequency bands subserving language prediction remains poorly understood.

In the present magnetoencephalography (MEG) study, participants read sentences in which the sentence-final words were predictable or not depending on the sentence context. In addition, the sentence-final words were either congruent or incongruent relative the sentence contexts. Different outcomes can be expected. First, the preactivation of linguistic information before the sentence-final word might engage language-related brain regions, resulting in greater alpha power suppression (Payne & Sekuler, 2014; Jensen, Bonnefond, & VanRullen, 2012). In addition, stronger beta activity might be expected for the highly constraining condition (Friston et al., 2015; Lewis & Bastiaansen, 2015). Moreover, there might be stronger high gamma power in response to the unexpected than the expected sentence-final words if the prediction error is reflected in the high gamma band. As we will demonstrate, we observed robust modulations in the gamma and alpha bands. This motivated investigating the cross-frequency interregional interactions between these bands to identify the dynamics reflecting functional connectivity.

METHODS

Participants

Thirty-four university students (mean age = 24 years old, range 20–35 years; 13 men) served as paid volunteers. They were all right-handed native Dutch speakers with normal or corrected-to-normal vision. None of them had dyslexia or any neurological impairment. They signed a written consent form according to the Declaration of Helsinki. The data of two participants (one woman) were excluded because of severe metal-related artifacts from dental work. The final set of participants therefore consisted of 32 participants (mean age = 24 years old, range 20–35 years; 12 men).

Stimulus

We manipulated the semantic constraint of sentence contexts as well as the semantic congruency of sentence-final words of 240 sentences (see Table 1 for two examples). As for the semantic constraint manipulation, each sentence pair differed only in one word (i.e., the critical word), which was at least two words preceding the target word. Following the sentence context, the target word could be predicted in the highly constraining (HC) context, whereas it was not the case in the low constraining (LC) context. The semantic constraint of contexts was measured in two rounds of cloze probability

Table 1. Examples of Two Items in Four Conditions

1. High/low constraining (HC/LC), congruent/incongruent (C/IC)
HC-C/IC: Hij gaf haar een **ketting** voor haar verjaardag/borstel.
(He gave her a **necklace** for her birthday/brush.)
LC-C/IC: Hij gaf haar een **ticket** voor haar verjaardag/borstel.
(He gave her a **ticket** for her birthday/brush.)
2. High/low constraining (HC/LC), congruent/incongruent (C/IC)
HC-C/IC: Om de **cellen** te kunnen zien gebruikte hij een microscoop/kathedraal.
(To see the **cells** he used a microscope/cathedral.)
LC-C/IC: Om de **objecten** te kunnen zien gebruikte hij een microscoop/kathedraal.
(To see the **objects** he used a microscope/cathedral.)
Statement: Hij gebruikte een apparaat om iets te kunnen zien.
(He used a device to see something.)

The examples were originally in Dutch, with the sentence-final words underlined. The critical words that create different contextual constraints are in **bold**. The target words are underlined. The English translations are given in brackets below the original Dutch materials. An example of the statement (which required a YES answer) was provided for Example 2.

test (34 participants in each round). The participants were asked to complete the sentences missing the last word with the first word that came to mind. The semantic constraints of the contexts were quantified by the percentage of the participants who filled in the same word for each sentence in each condition. In the end, the HC sentences have higher contextual constraints than the LC sentences: mean (*SE*) = 86% (1.9%) and 28% (1.8%), respectively; $t(478) = 62.27$, $p < .001$ (item-based analysis). As for the semantic congruency manipulation, the congruent sentence-final words were replaced with words that made the sentences incongruent (IC) in both the HC and LC conditions. The semantic congruency was measured in a sentence plausibility test by a different group of 32 participants. They were asked to rate the plausibility of each sentence on a scale from 1 (*highly implausible*) to 7 (*highly plausible*). Then, we averaged the plausibility ratings across sentences within each condition and each participant, which entered ANOVA that contained two within-subject factors (Contextual Constraint × Congruency). The congruent (C) sentences were rated to be more plausible than the IC sentences (main effect of Congruency: $F(1, 31) = 7011.81$, $p < .001$). The plausibility difference between the IC and C sentences was larger for the HC sentences than for the LC sentences, as supported by the interaction between Context and Congruency ($F(1, 31) = 150.82$, $p < .001$). The mean and *SE* of the ratings in the four conditions were as follows: HC/C: 6.49 (0.09), HC/IC: 1.59 (0.10), LC/C: 5.79 (0.12), LC/IC: 1.94 (0.12). Moreover, the IC and C words were matched on word category, animacy, word frequency ($F(1, 239) = 1.710$, $p = .192$; the mean (*SE*) was 38.662 (14.106) and 57.156 (37.367), respectively, based on the Dutch SUBTLEX-NL database; Keuleers, Brysbaert, & New, 2010), as well as word length ($F(1, 239) = 2.352$, $p = .126$; the mean (*SE*) was 5.796 (0.345) and 5.962 (0.361), respectively).

Procedure

Participants were tested individually in a dimly illuminated, magnetically shielded room. They were seated in a comfortable chair under the MEG helmet, facing a projected screen at approximately 80 cm distance. The stimuli were presented in gray color on a black background on the screen, with a font size of 36 for the words and of 30 for the probe statements. A trial started with a blank screen (duration 1600 msec), followed by a sentence that was presented word by word. Each word was presented for 200 msec, with an ISI of 800 msec (Figure 1). The last word ended with a period. To ensure that participants read for comprehension, they were required to judge the correctness of a statement following the sentence by pressing one of two buttons in 20% of the trials. The statement referred to the semantic content of the sentence, but not the semantic violations. In the remaining trials, a Dutch word "VOLGENDE" (meaning "NEXT") appeared on the screen, and participants were instructed to press a third button. The probe question and the "NEXT" signal followed 1600 msec after the presentation of the last word. All responses were required to be delivered within 5000 msec. After a response, the next trial began. The adding of the comprehension task might have encouraged the participants to make predictions and thus allowed us to better observe the prediction effect. Participants were asked not to move or blink when individual words appeared, but they were encouraged to blink during the presentation of the questions.

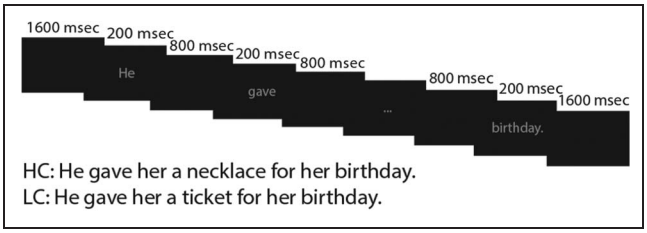


Figure 1. Illustration of the procedure and an example of the stimuli. A trial started with a blank screen (duration 1600 msec). Then the sentences were presented word-by-word (200 msec/word + 800 msec blank). The sentence was either followed by a probe question (20% of the trials) or a "NEXT" signal 1600 msec after the presentation of the last word. All responses were required to be delivered within 5000 msec.

Participants read 240 sentences in a pseudorandom order. No more than three sentences of the same condition were presented in succession. The 240 sentences in one list were divided into 12 blocks (24 trials per block), with each block lasting about 5 min. Between each block, there was a small break, after which participants could start the next block by informing the experimenter. The whole experiment took about 1.5 hr, including participants' preparation, instructions, and a short practice session consisting of 12 sentences.

Data Acquisition

MEG signals were recorded with 275 axial gradiometers (CTF Omega system, CTF Systems Inc., Port Coquitlam, Canada). In addition, the horizontal and vertical EOG as well as electrocardiography were recorded to later discard trials contaminated by eye movements, blinks, and heart beats. The ongoing MEG and EOG signals were low-pass filtered at 300 Hz, digitized at 1200 Hz, and stored for offline analysis. To measure the head position with respect to the axial gradiometers, three coils were placed at anatomical landmarks of the head (nasion, left and right ear canal). Head position was monitored in real time (Stolk, Todorovic, Schoffelen, & Oostenveld, 2013). MRIs of 30 participants were obtained with a 1.5-T or 3.0-T Siemens system (Berlin, Germany), with markers attached in the same position as the head coils. The MRIs were aligned to the MEG coordinate system according to the anatomical landmarks.

Data Preprocessing

Data were analyzed using the Fieldtrip software package, an open-source Matlab toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011). We analyzed the time window preceding and following the target words to examine the anticipatory and integration effects, respectively. First, the MEG data were segmented into trials starting 2 sec before and ending 2 sec after the onset of the target words. A third-order synthetic gradiometer correction was applied to remove noise from the environment. Trials contaminated with muscle or MEG jump artifacts were identified and removed using a semiautomatic routine. After that, we performed independent component analysis (ICA; Jung et al., 2000; Bell & Sejnowski, 1995) to the data and removed ICA components associated with the eye movement and cardiac-related activities from the MEG signals. These ICA components were identified by comparing them with the EOG and electrocardiography recordings. Finally, we inspected the data visually and removed any remaining artifacts. In the end, on average 96% of trials were kept, with equal numbers of trials for the four conditions (all $ps > .19$).

Time–Frequency Representations of Power

The time–frequency representations (TFRs) of the single trials were calculated in two different, partially overlapping frequency ranges. In the low-frequency range (2–30 Hz), a 500-msec Hanning window was applied in frequency steps of 2 Hz and time steps of 10 msec. In the high-frequency range (25–100 Hz), a multitaper approach was used (Mitra & Pesaran, 1999). Power estimates were computed with a 200-msec time-smoothing and a 10-Hz frequency-smoothing window in 5-Hz frequency steps and 50-msec time steps. The TFRs were calculated at each sensor location for the vertical and horizontal planar gradient and then averaged, as planar gradient maxima are strongest above neural sources (Bastiaansen & Knösche, 2000). Then we averaged the planar gradient TFRs of the trials separately for different conditions and for each participant (i.e., HC and LC conditions during the prediction period, IC and C conditions during the integration period). The TFRs were log10-transformed, and the difference between conditions was obtained by subtraction ("log ratio"). Because of temporal smearing, any given time point in the resulting TFR is a weighted average of the time window of ±250 msec. To avoid any power leakage from the evoked response to the target words (presented at time 0) and their preceding words (presented at time −1 sec and lasted for 0.2 sec), we constrained the TFR analysis to the time window from −550 to −250 msec relative to the target words onset to examine the prediction effect (HC vs. LC). After the target words were presented, we analyzed the whole time window (0–1000 msec) to examine the congruency effect (IC vs. C). To rule out the possibility that the anticipatory effect was caused by the different words presented in the context (i.e., the critical words that created different contextual constraints), we also analyzed the critical words with a similar approach.

Event-related Field Analysis

The event-related fields (ERFs) of the four conditions were obtained by averaging the trials separately for each condition with a −200 to 0 msec baseline correction. The ERFs were calculated for both axial and planar gradient data. On the basis of previous literature (Wang, Zhu, & Bastiaansen, 2012; Halgren et al., 2002) and visual inspection, we constrained our analysis of the N400m to the 300–600 msec time window. The averaged values of the planar gradient within this time interval entered the statistical analyses.

Source Analysis

To estimate the sources of the observed TFR effects (HC vs. LC, IC vs. C), we used a beamforming approach, Dynamic Imaging of Coherent Sources (DICS; Gross et al., 2001). The DICS algorithm computes a spatial filter from a lead field matrix and the cross-spectral density matrix (CSD) of the data from the axial gradiometers. To obtain the lead field for each participant, we first spatially coregistered the individual anatomical MRI to sensor space MEG data by identifying fiducials in the nasion and the two ears. Then, a realistically shaped single-shell head model was constructed based on the segmented anatomical MRI for each participant (Nolte, 2003). After that, each brain volume was divided into a grid spaced 10 mm apart and warped to the template Montreal Neurological Institute (MNI) brain (Montreal, Quebec, Canada), after which the lead field was calculated for each grid point (Nolte, 2003). The MNI template brain was used for two participants who did not come back for the MRI scan. On the basis of the sensor-level results of the target words (see the Alpha and Beta Effects Preceding the Target Words section), the CSD was computed for two frequency bands (centering at 10 Hz, with ±2 Hz spectral smoothing for alpha band; centering at 70–80 Hz with ±10 Hz spectral smoothing for gamma band) within the time windows (relative to target word onset) that showed significant effects (−600 to −200 msec and 450–1000 msec for alpha band, −600 to −200 msec for beta band, 150–650 msec for gamma band). Note that the width of the time window in the prediction period (−600 to −200 msec) was selected to obtain a ±2 Hz spectral resolution. A common spatial filter was constructed by combining the CSDs of the two conditions. The power at each grid point was estimated by applying the common filter to the Fourier transformed data of the contrasting conditions (HC and LC, IC and C) separately. The estimated power in "source space" was averaged over trials and then log10-transformed. The power difference between conditions was estimated by

subtracting the log10-transformed power ("log ratio"). For visualization purposes, the grand-averaged grid was interpolated onto the MNI template brain (Figures 3C, E, 6C, and 7C).

Cross-frequency Connectivity

On the basis of the source analysis, we defined three ROIs by selecting the grid points that showed the largest alpha power modulation: left inferior frontal, left middle temporal region, as well as left fusiform (see Alpha and Beta Effects Preceding the Target Words section). We examined cross-frequency connectivity by correlating the alpha and gamma power from the three ROIs over trials separately for the HC and LC conditions. We first calculated the alpha power (8–12 Hz) of the three ROIs for each trial using the DICS approach (multitaper). Likewise, we calculated the gamma power between 60 and 90 Hz (multitaper, 70–80 Hz with ±10 Hz spectral smoothing). After that, the correlation between alpha power and gamma power was calculated across trials for the HC and LC conditions separately (i.e., power–power correlation as indicated in Jensen & Colgin, 2007, as well as Mazaheri, Nieuwenhuis, van Dijk, & Jensen, 2009). The alpha and gamma power was obtained in the prediction time window (i.e., −600 to −200 msec) preceding the target words. Following the presentation of the target words, the alpha and gamma power was estimated respectively in the time windows of 150–650 msec and 450–1000 msec. We examined both the prediction and the integration periods, because we expected the connectivity to be modulated by predictability in both intervals. Then, we quantified the significance of the correlations by performing a cluster-based permutation test between the whole-brain correlation values against zero.

Cluster-based Permutation Statistics

We performed cluster-based permutation tests (Maris & Oostenveld, 2007) across participants for the TFR and the source results to control for multiple comparisons over time and space On the basis of previous literature (Friston et al., 2015; Lewis & Bastiaansen, 2015; Spaak et al., 2015), we statistically quantified the alpha (8–12 Hz), beta (16–20 Hz), and gamma (60–90 Hz) power differences between the HC and LC conditions as well as between the IC and C conditions across subjects. For each "voxel" of the observed data (i.e., sensor or Sensor × Time for sensor-level TFR analysis; *x*/*y*/*z* grid point for source–space analysis), we computed the mean difference between conditions. The mean difference was computed both for the observed data and for 1000 permutations obtained for relabeled conditions. Based on the per-voxel permutation distribution of the descriptives thus obtained, we thresholded the observed values with the 95th percentile of this distribution to obtain cluster candidates. For each permutation, the cluster candidate with the highest sum of voxel-level descriptives was added to the permutation distribution of cluster statistics. The sum of descriptives for each observed cluster candidate was compared with this permutation distribution to assess significance for each cluster. Clusters falling in the highest or lowest 2.5th percentile were considered significant.

RESULTS

Participants were asked to read sentences that were visually presented one word at a time (Figure 1). The sentence context was either highly constraining (HC) or low constraining (LC) with respect to the sentence-final word. The sentence-final word was either congruent (C) or incongruent (IC) relative to the preceding context. Therefore, a full factorial design comprising context (HC, LC) and congruency (C, IC) was used, with 60 trials in each condition. Participants were asked to judge the correctness of statements in 20% of the sentences. They made slightly more accurate responses in the C condition than in the IC condition (main effect of congruency: $F(1, 31) = 4.394$, $p = .044$, $\eta^2 = .124$; mean (*SE*) = 98.7% (0.2%), 98.3% (0.3%), 98.3% (0.3%), and 97.6% (0.3%) respectively for the HC/C, HC/IC, LC/C, and LC/IC conditions). The overall high accuracy suggests that the participants carefully read the sentences for comprehension. In addition, no difference was found in the RT: mean (*SE*) = 1318 msec (96 msec), 1330 msec (102 msec), 1298 msec (96 msec), and 1334 msec (102 msec) respectively for the HC/C, HC/IC, LC/C, and LC/IC sentences; all p values $> .150$.

Alpha and Beta Effects Preceding the Target Words

Although the sentences were presented, the ongoing MEG was recorded in 32 participants. Figure 2 presents the raw TFRs of power averaged across all conditions for a representative left posterior sensor (shown in the topography). Note that the power is calculated for the synthetic planar gradients. The presentation of each word induced both alpha and gamma power modulations. To quantify the consequences of prediction, we compared the TFRs of the alpha (8–12 Hz), beta (16–20 Hz), and gamma (60–90 Hz) band activity for the HC and LC conditions in the interval just before the presentation of the final word (−550 to −250 msec). As seen in six representative sensors (see Figure 3B for location), the alpha power in the HC condition was lower than in the LC condition. The effect was clearly left-lateralized (see Figure 3B for the topographic distribution). A cluster-based permutation test revealed significant left ($p < .002$) and right clusters ($p = .038$). In addition, the beta power (16–20 Hz) was lower in the HC condition than in the LC condition over left frontal and temporal regions ($p = .004$; Figure 3D). No significant differences were observed in the gamma band.

We then applied a beamformer approach to identify the sources. Two significant clusters were identified for

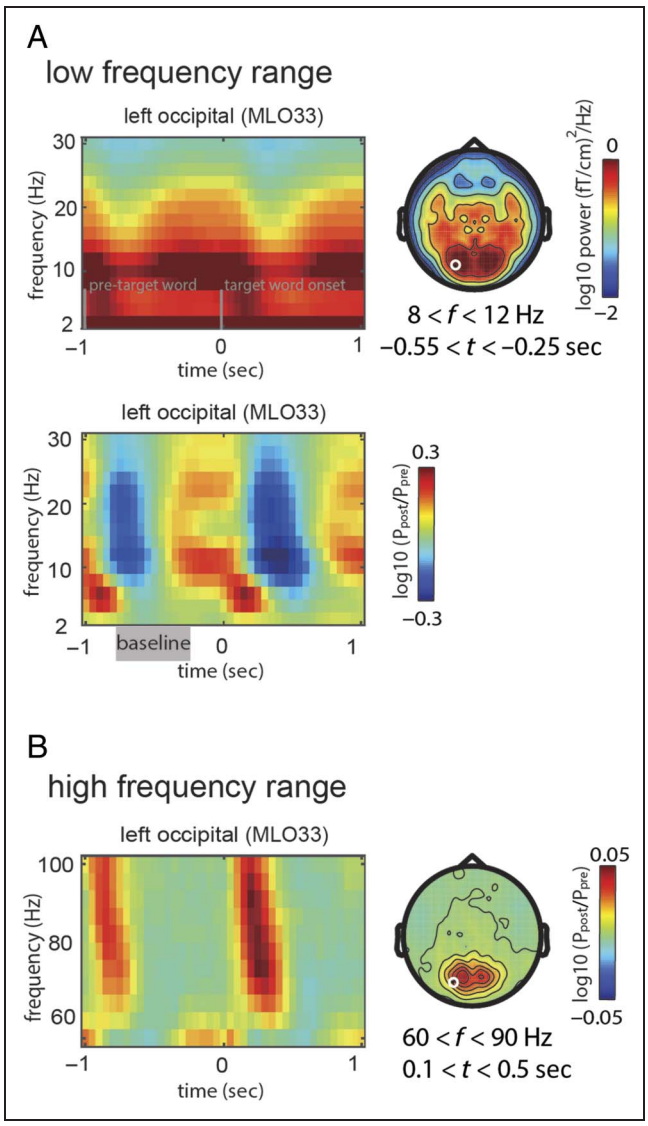


Figure 2. TFRs of power collapsed over conditions at one left posterior sensor (MLO33, indicated by circles on topographic plots). The target word started at 0 sec. The presentation of words (−1 to −0.8 sec and 0–0.2 sec) induced initial alpha power suppression and gamma power increase. Note that the effect of the second to the last word is also present. (A) TFR in the low-frequency band without baseline correction (top) and with a relative (−0.75 to −0.25 sec) baseline correction (bottom). (B) TFR in the high-frequency band with relative power change compared with the baseline period (−0.75 to −0.25 sec).

the alpha effect (Figure 3C). The first cluster shows stronger alpha power suppression for the HC condition than for the LC condition in the left inferior frontal cortex (LIFC), extending to ventromedial pFC ($p < .002$). The second cluster shows alpha modulation in the left posterior temporal region, which extends to the visual word form area (VWFA), the left hippocampus (but the maximum activation was more toward the surface of the temporal cortex), and the right cerebellum ($p = .032$). The left-hemisphere modulations seen in the sources and the bilateral modulation at the sensor level might be explained by reduced sensitivity of the cluster-based permutation test at the source level. The sources of the beta depression (16–20 Hz) were estimated to left posterior temporal regions, extending to VWFA and the left angular region (Figure 3E; $p = .012$). The source localizations of the beta depression partly overlapped with the source of the alpha depression. Given a possible harmonic relationship between the alpha (8–12 Hz) and beta (16–20 Hz) frequency bands, the beta suppression might partly reflect similar cognitive processes as the alpha suppression.

To exclude the possibility that the observed effects preceding the target words were merely carried over from the critical words that created different semantic constraints, we also tested the TFR difference induced by the critical words (i.e., the words that defined different constraints in the contexts). We found increases in theta (2–6 Hz) and low beta power (16–20 Hz) after the critical word onset in the HC condition compared with the LC condition in the 400–1000 msec time window (Figure 4). The different effects between the critical words and the target words suggest that the prediction effect observed preceding the target words was not a prolonged effect following the critical words.

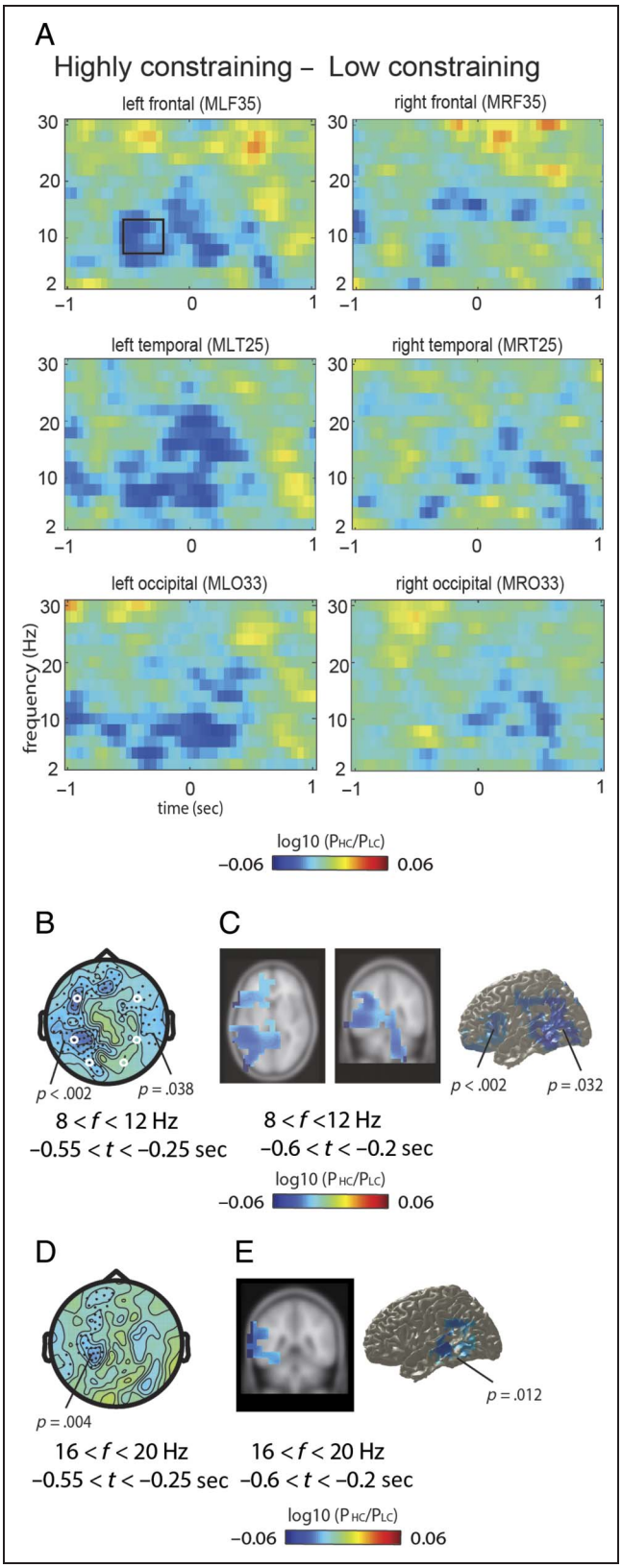
Evoked and Induced Effects following Target Word Onsets

We also calculated the ERFs of the four conditions. Figure 5A shows the planar gradient of the ERFs time-locked to the onset of the sentence-final words. The IC words elicited a larger N400m than the C words (main effect of Congruence: $p < .002$). The violation effect was stronger in the HC condition (HC/IC vs. HC/C: $p < .002$) than in the LC condition (HC/IC vs. HC/C: $p < .002$), as supported by a significant interaction between Congruency and Context ($p = .01$). In addition, the congruent words in the HC condition elicited a significantly smaller N400m compared with the LC condition (HC/C vs. LC/C: $p = .01$), whereas no significant difference was found between the incongruent words in the HC and LC conditions ($p = .124$). Our data are compatible with previously reported ERP effects (Thornhill & Van Petten, 2012; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007). Figure 5B shows the topographical distribution of the semantic violation effect (the N400m effect) in the 300–600 msec time window of the axial and planar gradient. The N400m effect of the axial gradient showed a strong dipolar pattern over the left hemisphere and a weaker dipolar pattern over the right hemisphere, which confirmed the left-hemisphere dominance of the N400m effect.

Next, we quantified the modulation of oscillatory activity with respect to congruency of the sentence-final words. Figures 6A and 7A show the contrast between the IC and C words after the presentation of the sentence-final word for six representative sensors. The alpha and gamma power differences between the IC and C conditions were statistically tested in 0–1000 msec interval

(not averaged over time) using the cluster randomization approach. After identifying the time windows that showed significant effects, we averaged the data in this time interval. The IC words resulted in a stronger decrease in alpha power (8–12 Hz) compared with the C words in the 450–1000 msec interval over left temporal and bilateral visual cortex. The topography of this effect was highly robust (Figure 6B; main effect of congruency; $p < .002$). In addition, the modulation in the alpha band in response to the congruency was more prominent in the HC condition than in the LC condition, as supported by a significant interaction between congruency and context ($p < .002$). The source localization of the alpha band modulation (IC vs. C) revealed regions in left inferior temporal cortex, bilateral occipital cortex, as well as bilateral cerebellum ($p < .002$; Figure 6C).

Subsequently, we investigated the congruency effect in the higher-frequency bands. We found that the IC compared with C words induced stronger gamma power (60–90 Hz) in the 150–650 msec interval (Figure 7A) in sensors over the left temporal and frontal cortices (Figure 7B; $p = .018$). The sources of the gamma modulation (IC vs. C) were located in the left frontal and temporal cortices as well as in the right middle temporal cortex (Figure 7C). Note that, albeit the effect was significant at sensor level, it was not significant at source level. Note that we need to keep in mind that the source-level statistical results mainly inform us how likely the two conditions differ in the ROIs. We further compared the gamma power across four conditions. As shown in Figure 7D, the induced gamma power in the HC/IC condition was not stronger compared with the LC/IC condition, even though the prediction error in the HC/IC condition was the greatest.



Relationship between Alpha and Gamma Power

To characterize the functional connectivity between brain regions, we conducted power–power correlation analysis

Figure 3. The TFRs of power contrast between the highly (HC) and low constraining (LC) contexts preceding the presentation of the sentence-final words at six selected sensors (as indicated in B). (A) The HC context included stronger alpha power suppression than the LC context particular in the sensors over the left hemisphere. (B) Topographic distributions of the alpha effect in the time window of −0.55 to −0.25 sec (see box in MLF 35). The clusters of sensors that showed a significant difference are highlighted with black dots. (C) Source localization results of the alpha effect (HC minus LC) in the prediction period. The results are masked by statistically significant clusters. The results are shown both on the transverse plane (Talairach coordinates: $z = 2.5$ mm) and the coronal plane (Talairach coordinates: $y = -41.5$ mm) as well as projected to the cortical surface. The alpha suppression for the HC contexts was estimated to the LIFC, left posterior temporal region, VWFA, left hippocampus, as well as right cerebellum (not shown in the cortical surface plot). (D) Topographic distributions of the beta power modulations (16–20 Hz) in the −0.55 to −0.25 sec time window. The clusters of sensors that showed significant difference are highlighted with black dots. (E) Source localization results of the beta effect (HC minus LC) in the prediction period. The results are masked by statistically significant clusters. They are shown both on the coronal plane (Talairach coordinates: $y = -41.5$ mm) as well as projected to the cortical surface. The sources of the low beta modulation were estimated to left posterior temporal regions, extending to VWFA and left angular region.

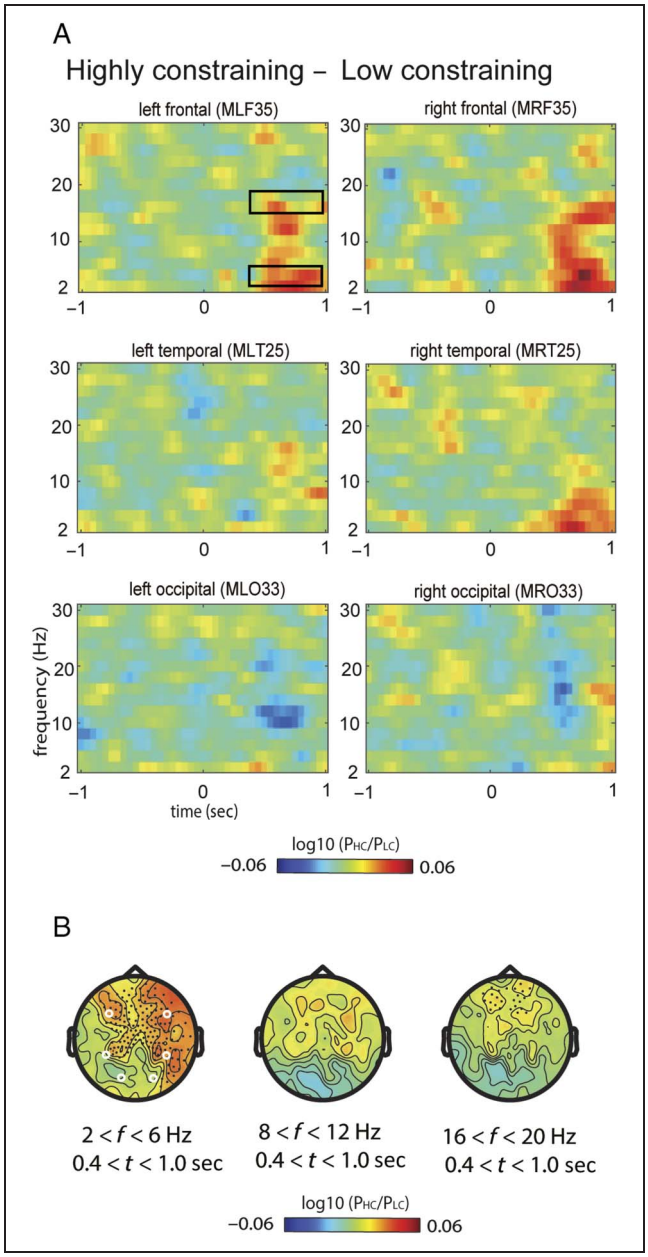


Figure 4. The contrast between the highly (HC) and lowly constraining (LC) conditions after the presentation of the critical words that create different contextual constraints. (A) Difference in TFRs of power at six selected sensors (as indicated in B). The words for the HC context included stronger theta and low beta power than the words in the LC context particular in MLF 35 and MRF 35. (B) Topographic distributions of the theta and beta effects in the time window of 0.4–1.0 sec (see boxes in MLF 35). The clusters of sensors that showed significant difference are highlighted with black dots. The stronger power increases for HC than LC were dominated by the frontal region and right hemisphere.

across alpha and gamma frequency bands. We first asked which activity correlated with the alpha power depressed in the VWFA. This was done by calculating the correlation coefficient over trials between alpha power in the reference regions (i.e., VWFA; identified in Figure 3C) with the gamma power at all other locations. Before the sentence-final word (−600 to −200 msec), alpha power in the VWFA correlated negatively over trials with gamma power in left inferior and middle frontal cortex in the HC condition (p = .014; left) but not in the LC condition (right). This effect was statistically evaluated by a cluster randomization approach (see Methods) controlling for multiple

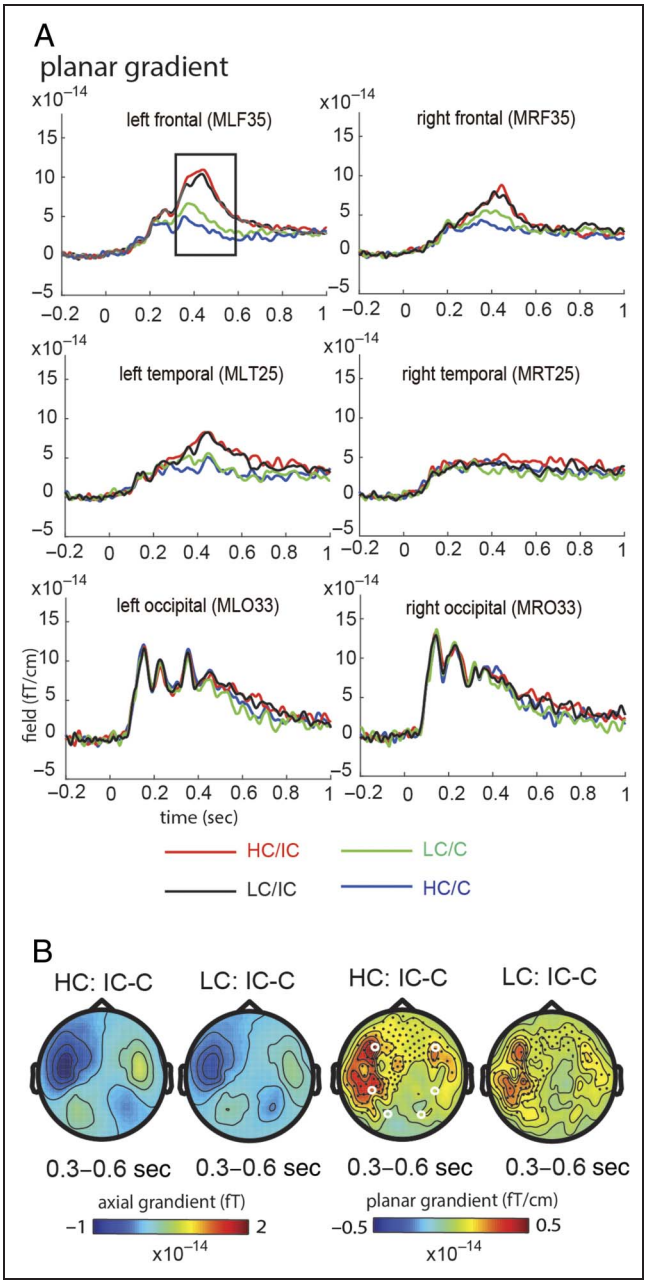


Figure 5. The ERFs elicited by the target words. (A) ERFs (planar gradient) elicited by the target words in the four conditions (congruence in highly constraining cortex: HC/C; incongruent in highly constraining: HC/IC; congruence in low constraining: LC/C; incongruent in low constraining: LC/IC) at six selected sensors (as indicated in B). The IC words elicited a larger N400m than the C words in the time window of 0.3–0.6 sec, with a greater congruency effect in the HC context than in the LC context. (B) Topographic distributions of the observed N400m effect for both axial and planar gradient data. The congruency effect was distributed over bilateral frontal/temporal sensors, with left hemisphere dominance.

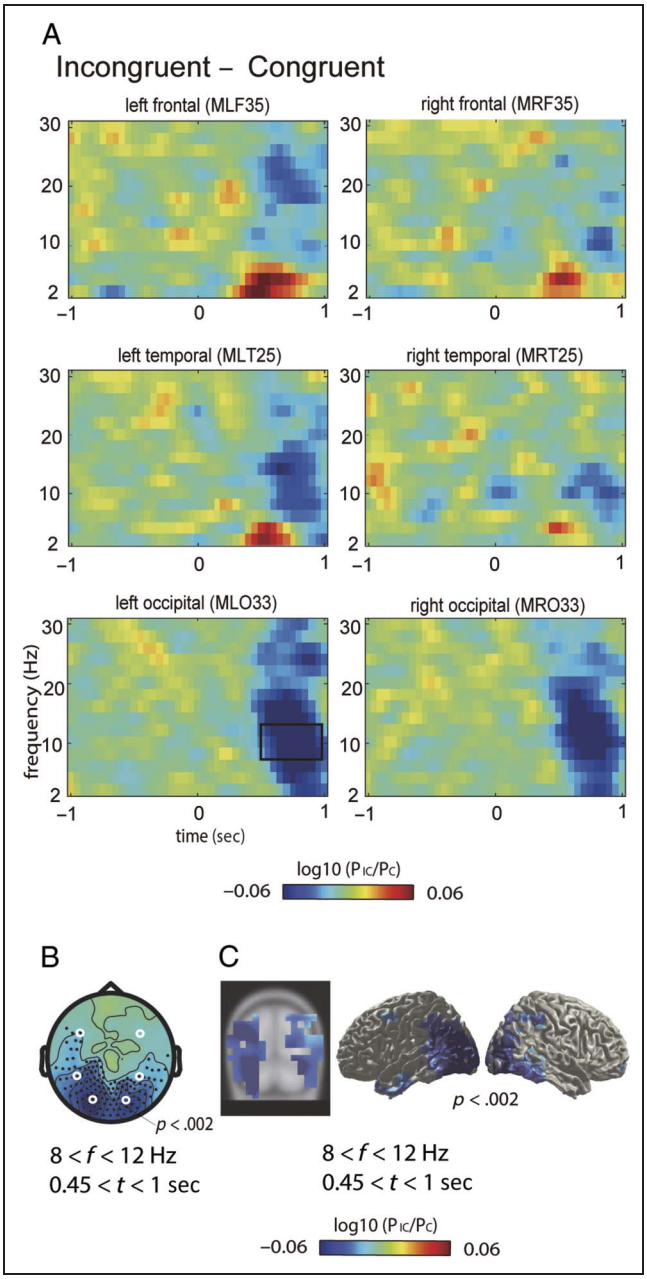


Figure 6. The TFRs of power contrast between the incongruent (IC) and congruent (C) words after the presentation of the sentence-final words in the low-frequency band at six selected sensors (as indicated in B). (A) The IC words induced stronger alpha power suppression than the C words over temporal and posterior sensors, which were dominated by the left hemisphere. (B) Topographic distributions of the alpha effect in the time windows that showed significant differences (see box in MLO 33). The clusters of sensors that showed significant difference are highlighted with black dots. (C) Source localization results of the alpha effect (IC minus C) after the presentation of the sentence-final words. The results are masked by statistically significant clusters. The results are shown on the coronal plane (Talairach coordinates: $y = -67.5$ mm) as well as projected to the cortical surface. The alpha suppression for the IC words was estimated to left temporal cortex, bilateral occipital cortex as well as bilateral cerebellum (not shown in the cortical surface plot).

comparisons of correlation values (tested against zero) over all the grid points in brain volume.

After the target words were presented, the left VWFA alpha power continued to be negatively correlated with the gamma power over the bilateral prefrontal and right posterior temporal regions in the HC contexts ($p < .002$; cluster-based randomization approach) but not in the LC condition (Figure 8B). This demonstrates that the VWFA and the left pFC were functionally connected, as revealed by anticorrelation in the alpha–gamma band when a given word can be anticipated. This effect was sustained also after word presentation.

Next, we used the left temporal cortex (identified in Figure 3C) as reference regions (Figure 9). The left temporal alpha power correlated negatively over trials with the gamma power over the superior medial frontal and left prefrontal regions in the HC condition ($p = .022$). In the LC condition, there was a negative correlation with gamma power in the superior medial frontal region, but not in the left prefrontal region ($p = .024$; Figure 9A). After the target words were presented, the alpha power in the left temporal region negatively correlated with the gamma power over the left prefrontal, right superior parietal, and occipital regions only in the HC contexts ($p < .002$), whereas no correlation was found in the LC condition (Figure 9B).

Finally, when alpha power in the LIFC (identified in Figure 3C) served as reference (Figure 10), the HC condition did not produce a significant alpha–gamma power correlation, whereas the LC contexts resulted in a significant negative alpha–gamma correlation over the right parietal and occipital cortex ($p = .006$; Figure 10A). After the target words were presented, the alpha power in the left inferior frontal region negatively correlated with the gamma power over right parietal, temporal, and visual cortex only in the HC contexts ($p < .002$; Figure 10B).

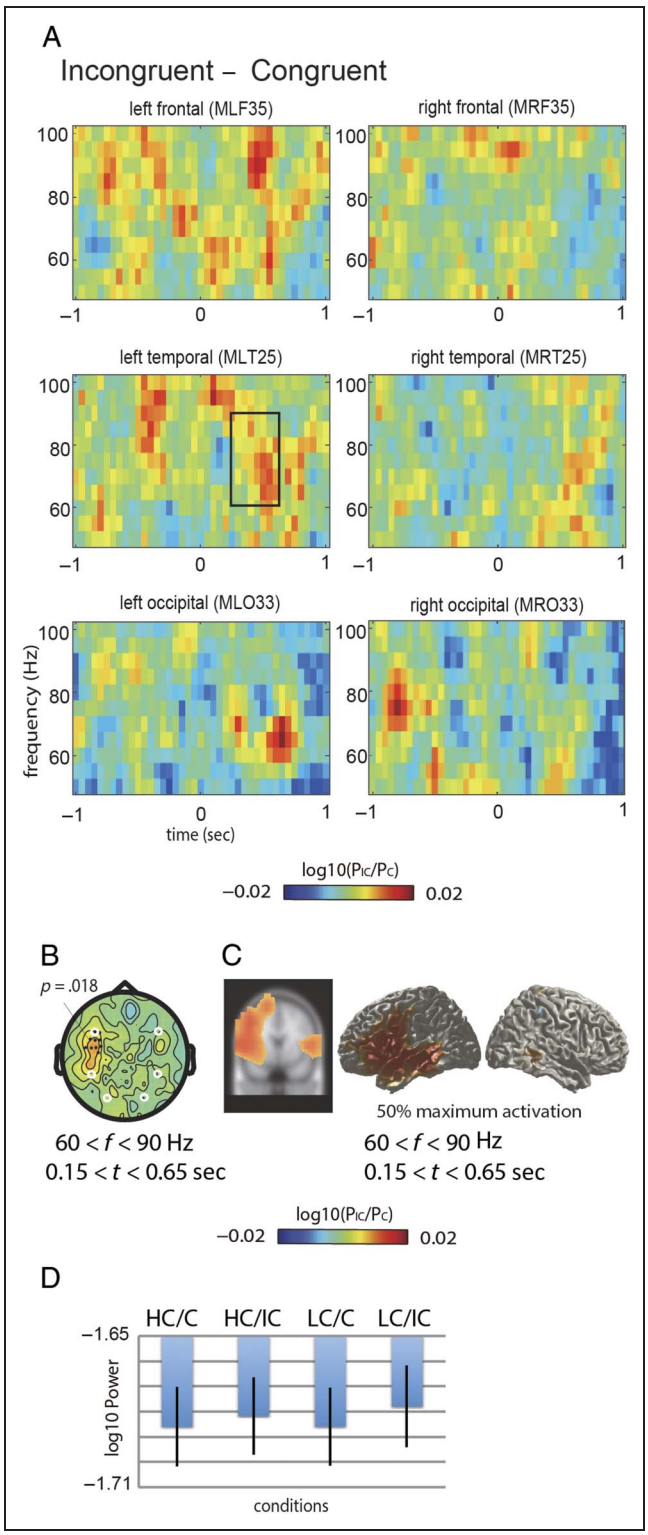
To substantiate the connectivity results, we also calculated the power–power correlation over trials between the left inferior frontal gamma and the alpha power. Before the target words were presented, the HC condition induced a marginally significant negative gamma–alpha correlation in the left inferior temporal and right visual cortex ($p = .080$), whereas the LC condition induced a significant negative gamma–alpha correlation in the right inferior frontal and temporal regions ($p = .018$; Figure 11A). After the target words were presented, the HC condition induced a significant negative gamma–alpha correlation in bilateral visual cortex ($p = .006$), whereas the LC context induced a significant negative gamma–alpha correlation in the right posterior temporal region ($p = .008$; Figure 11B). The results further demonstrate the negative correlation between the left frontal gamma power and the temporal and visual alpha power in the HC rather than in the LC condition.

In summary, when the sentence final words could be anticipated, the left pFC became functionally connected to

the VWFA and the left temporal cortex. This was revealed as anticorrelation between gamma and alpha power.

DISCUSSION

We have identified neural mechanisms involved in language prediction. In the time interval when the last word could be anticipated, we observed stronger alpha and beta power suppressions for highly (HC) compared with low constraining (LC) contexts. The sources of the alpha suppression were localized to the language network including the LIFC, the left posterior temporal region, and the VWFA. We then further identified the functional connectivity between the nodes. We found gamma power in the left prefrontal region negatively correlated with the alpha power over the left temporal and visual regions in the HC condition, both before and after the presentation of the sentence-final words.



Alpha Depression Reflects Engagement of the Language Network during Prediction

What explains the stronger anticipatory alpha power suppressions for the HC context in the left language system? Alpha oscillations have been initially regarded to reflect "cortical idling" as alpha power increased when eyes are closed compared with when eyes are opened (Pfurtscheller, Stancak, & Neuper, 1996). This view has been replaced by the notion that alpha power is involved in the allocation of computational resources by actively inhibiting task-irrelevant regions (Payne & Sekuler, 2014; Foxe & Snyder, 2011; Jensen & Mazaheri, 2010). This also implies that an alpha power decrease reflects the engagement of task-relevant regions. As such, modulations in alpha power reflect the active allocation of resources in the working brain (Boudewyn et al., 2017). A strong case has been made for sensory regions, such as the visual (nonlinguistic stimuli: Spaak et al., 2015; van Dijk, Schoffelen, Oostenveld, & Jensen, 2008; Hanslmayr et al., 2007; linguistic stimuli: Magazzini, Ruhnau, & Weisz, 2016), auditory (Leske et al., 2014, 2015), and somatosensory system (Haegens, Nácher, Luna, Romo, & Jensen, 2011; Haegens, Osipova, Oostenveld, & Jensen, 2010). We here suggest that the HC context engages the language network as reflected by reduced alpha power. Importantly, these results extend the functional role of the alpha activity from

Figure 7. The TFRs of power contrast between the incongruent (IC) and congruent (C) words after the presentation of the sentence-final words in the high-frequency band at six selected sensors (as indicated in B). (A) The IC words induced stronger gamma power increase than the C words over left frontal and temporal sensors. (B) Topographic distributions of the gamma effect in the time windows that showed significant differences (see box in MLT 25). The clusters of sensors that showed significant difference are highlighted with black dots. (C) Source localization results of the gamma effect (IC minus C) after the presentation of the sentence-final words. The results are masked by 50% maximum difference as no statistically significant cluster was found between the IC and C conditions at the source level. The results are shown on the coronal plane (Talairach coordinates: y = 10.5 mm) as well as projected to the cortical surface. The gamma power increase was estimated to left inferior frontal and left temporal cortex. (D) Gamma power (60–90 Hz) between 0.15 and 0.65 sec of sensors that showed significant congruency effect (as indicated in B) for four conditions. HC = highly constraining; LC = low constraining; C = semantically congruent; IC = semantically incongruent.

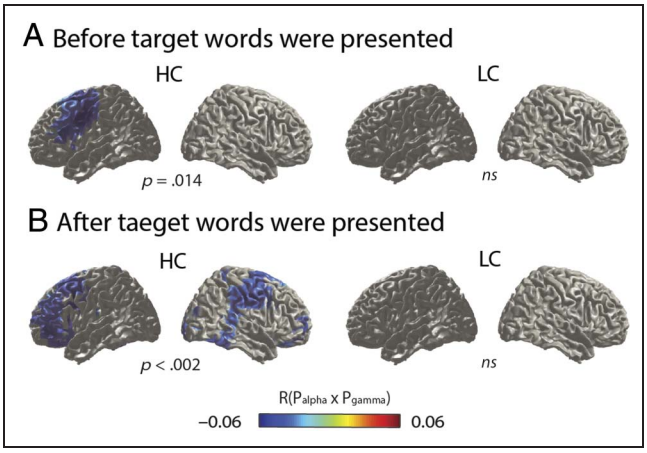


Figure 8. Cross-frequency connectivity between VWFA alpha power and gamma power in the source space (A) before and (B) after the target words were presented. The power–power correlation values were projected to the cortical surface (masked by the statistically significant cluster tested against zero). (A) Before the target words were presented, the HC context induced a significant negative alpha–gamma correlation over the left inferior and middle frontal cortex, whereas no significant negative correlation was observed for the LC context. (B) After the target words were presented (shown in bottom row), the HC context induced a significant negative alpha–gamma correlation over bilateral prefrontal and right posterior temporal regions, whereas no significant negative correlation was observed for the LC context.

primarily operating in sensory regions to the extended language network.

The sources of the alpha suppression were found in LIFC extending to ventromedial pFC, left posterior tem-

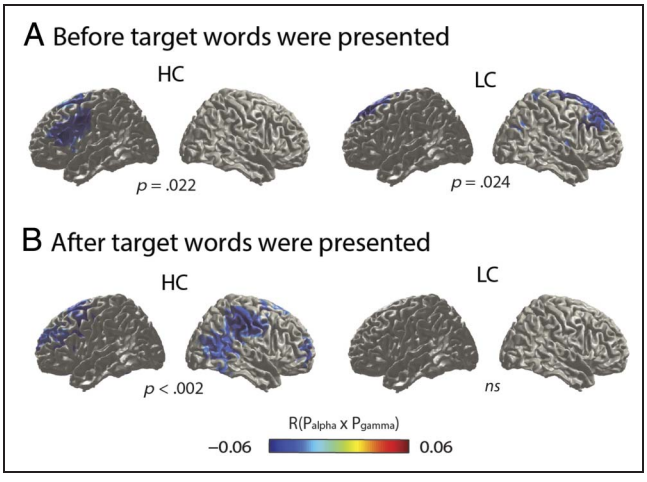


Figure 9. Cross-frequency connectivity between left temporal alpha power and gamma power in the source space (A) before and (B) after the target words were presented. (A) Before the target words were presented, the HC context induced a significant negative alpha–gamma correlation over the superior medial frontal and left prefrontal regions, whereas the LC context induced a significant negative alpha–gamma correlation only over the superior medial frontal region. (B) After the target words were presented, the HC context induced a significant negative alpha–gamma correlation over the left prefrontal, right superior parietal, and occipital regions. No significant correlation was found in the LC context.

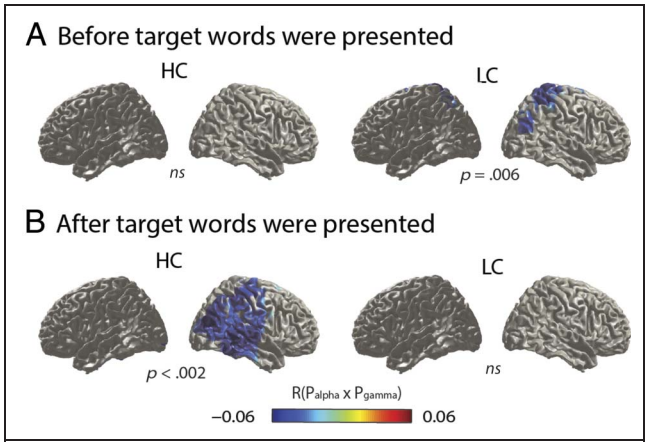


Figure 10. Cross-frequency connectivity between left inferior frontal alpha power and gamma power in the source space (A) before and (B) after the target words were presented. (A) Before the target words were presented, the low constraining (LC) context induced a significant negative alpha–gamma correlation over the right parietal and occipital cortex. No significant correlation was found in the highly constraining (HC) context. (B) After the target words were presented, the HC context induced a significant negative alpha–gamma correlation over right parietal, temporal, and visual cortex, whereas no significant correlation was found in the LC context.

poral region, VWFA, left hippocampus, and right cerebellum. The VWFA supports processing of orthographic information (Dehaene & Cohen, 2011), which might also be influenced by top–down mechanisms (Price & Devlin,

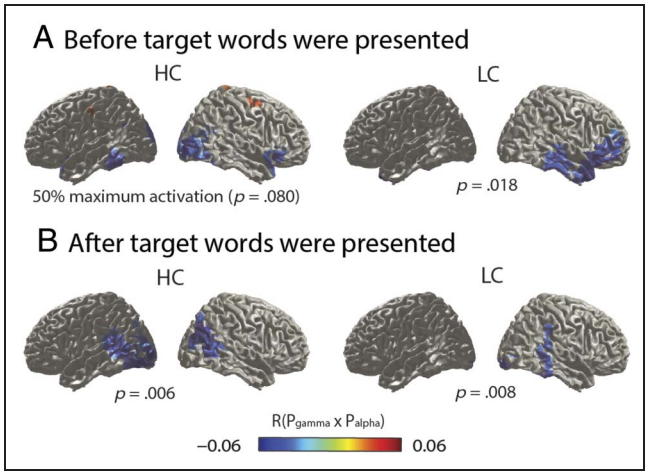


Figure 11. Cross-frequency connectivity between left inferior frontal gamma power and the alpha power before (A) and after (B) the presentation of the target words. (A) Before the target words were presented, the highly constraining (HC) context induced a marginally significant negative gamma–alpha correlation in the left inferior temporal and right visual cortex (power–power correlation values were masked by 50% maximum difference), whereas the low constraining (LC) context induced a significant negative gamma–alpha correlation in the right inferior frontal and temporal regions. (B) After the target words were presented, the HC context induced a significant negative gamma–alpha correlation in bilateral visual cortex, whereas the LC context induced a significant negative gamma–alpha correlation in right posterior temporal region.

2011). Consistently, Levy et al. (2013) found that alpha-band suppression in the VWFA mediates conscious word form perception. In addition, Strauß, Kotz, Scharinger, and Obleser (2014) found decreased alpha power over left posterior temporal areas with increased activation of semantic features. The anticipatory alpha effect in the VWFA suggests preactivation of the predicted orthographic representation. The posterior temporal cortex was found to be involved in long-term storage and retrieval of lexical representations, including phonological word forms, morphological information, word meanings, and syntactic templates (Hagoort, 2013, 2016; Lau & Nguyen, 2015; Lau, Phillips, & Poeppel, 2008). We therefore suggest that the preactivation of the VWFA is supported by the lexical access in the left posterior temporal cortex.

Lau et al. (2008) suggested that the LIFC mediates the retrieval and selection of lexical representations. Hagoort (2005, 2013) proposed that LIFC serves to unify multiple sources of information, as has been supported in previous fMRI studies (e.g., Hagoort & Indefrey, 2014; Obleser & Kotz, 2010; Snijders et al., 2008; Baumgaertner, Weiller, & Büchel, 2002). The LIFC alpha suppression might reflect predictions based on unified lexical representation from the preceding context in association with lexical retrieval from the temporal cortex. The LIFC engagement extended to ventromedial pFC, which plays an important role in generating predictions in relation to context (Bar, 2007). In addition, the LIFC and ventromedial pFC have numerous reciprocal anatomical connections with temporal and occipital brain regions (Xiang, Fonteijn, Norris, & Hagoort, 2010; Rilling et al., 2008; Bar et al., 2006). These anatomical connections might provide the infrastructure to support the top–down control on lower-level processes. Previously, Dikker and Pylkkänen (2013) used prelearned pictures to create predictions for a specific word. They found stronger theta (4–7 Hz) activities in left temporal, ventromedial prefrontal and visual cortex for highly predictive compared with less predictive pictures. In natural language comprehension, prediction generally arises from complex interactions between semantic associations and high-level syntactic structure. Our findings suggest that prediction engages several nodes from the language network as reflected by a decrease in alpha band activity.

We also identified alpha power depression in the hippocampus and cerebellum. Both regions have been related to sequential processing and prediction. Some fMRI studies found that hippocampal activation was modulated by the degree to which upcoming events could be predicted (Schiffer, Ahlheim, Wurm, & Schubotz, 2012; Harrison, Duggins, & Friston, 2006; Strange, Duggins, Penny, Dolan, & Friston, 2005) and that hippocampal pattern completion is related to action-based mnemonic expectation (Hindy, Ng, & Turk-Browne, 2016). Given the sequential and relational nature of language, the hippocampus might also be involved in generating language predictions. Recently, Piai et al. (2016) recorded brain activities from the hippocampal structure during sentence completion. They found theta power increase for the HC relative to LC contexts, providing direct evidence for the involvement of hippocampus during the generation of semantic meanings on the basis of prediction. The cerebellum was found to be important for representing the temporal relationships between events (Pisotta & Molinari, 2014) and for generating anticipations during language comprehension (Moberget, Gullesen, Andersson, Ivry, & Endestad, 2014; Leggio et al., 2008). The cerebellum also contributed to speech timing, phonological aspects of lexical access, and articulatory control (Hertrich, Mathiak, & Ackermann, 2016; Marvel & Desmond, 2016; Ziegler, 2016), so the right cerebellum activation might indicate the preactivation of the phonological properties of the highly expected words. A recent fMRI study measured BOLD signals before participants made anticipatory eye movements toward the spatial location that was associated with expected target word category (Bonhage, Mueller, Friederici, & Fiebach, 2015). Both language network (e.g., left inferior frontal gyrus and left superior temporal areas) as well as hippocampus and cerebellum were activated. In summary, consistent with literature from other domains, we have shown that the language network and subcortical regions are engaged before the onset of a highly predicted word.

In addition to the modulation in the alpha band, stronger beta power suppression was induced by the highly relative to the low constraining context. Increased beta synchronization has been proposed to reflect top–down predictions to hierarchically lower processing levels (Friston et al., 2015; Lewis & Bastiaansen, 2015). However, we observed greater beta power suppression associated with stronger predictions. Our findings are consistent with a study by Magyari, Bastiaansen, de Ruiter, and Levinson (2014) reporting stronger beta power suppression (11–18.5 Hz) when there was a strong prediction of the termination of a conversation. In addition, Piai et al. (2014, 2015) found decreased power in the alpha and beta ranges (8–30 Hz) before naming pictures in highly relative to low predictive context. In short, our finding provides converging evidence that beta power suppression is associated with prediction.

After the presentation of the target words, the incongruent (IC) words induced greater alpha power suppression compared with the congruent (C) words in the left temporal, bilateral occipital cortex, as well as bilateral cerebellum. These regions were also more strongly activated in the HC context than the LC context in the prediction period. In addition, the alpha modulation in response to the semantic congruency (IC vs. C) was stronger in the HC condition than the LC condition. Therefore, the alpha suppression might reflect further engagement of the preactivated brain regions to integrate the words into the contexts. Because the violation of expectancy is greatest in the HC/IC condition, the engagement of brain areas

for integration of word meaning into context was found to be strongest in this condition.

Induced Gamma Band Activity Does Not Reflect Prediction Error

After presentation of the target words, the IC words induced stronger gamma power than the C words in left prefrontal and temporal regions in both the HC and LC conditions. The comparable gamma effect between the HC and LC conditions argues against the claim that gamma activity relates to prediction error (Friston et al., 2015; Lewis & Bastiaansen, 2015), as the prediction error in the HC condition is stronger than that in the LC condition. The absence of a significant gamma effect between the HC and LC words was unlikely due to low statistical power, as in fact the LC/IC words seemed to induce stronger gamma power than the HC/IC words (see Figure 7D). The left frontal and temporal regions have been related to semantic unification and retrieval (Hagoort, 2005, 2013), so the increased gamma activity might reflect increased unification load and semantic retrieval effort of the IC words relative to the C words (Hagoort, Hald, Bastiaansen, & Petersson, 2004). Some previous studies have reported greater gamma activity in response to highly expected compared with unexpected words (e.g., Monsalve, Pérez, & Molinaro, 2014; Molinaro, Barraza, & Carreiras, 2013; Wang et al., 2012; Penolazzi, Angrilli, & Job, 2009). The discrepancy might be related to the composition of the stimuli set in a particular experimental setting, such as the level of attention and prediction (as discussed in Lewis & Bastiaansen, 2015).

Functional Connectivity between Left Inferior Frontal and Temporal Areas as Reflected by Cross-frequency Correlations between Alpha and Gamma Activity

The negative trial-by trial correlation between alpha and gamma power suggests that the stronger alpha suppression is associated with increased gamma power. This alpha–gamma coupling was found between alpha band activity in the left temporal and VWFA and gamma activity in the left prefrontal area. It was only present in the HC condition. Moreover, the alpha–gamma coupling was present both before and after the presentation of the sentence-final words. Given that the temporal and VWFA activities reflect the preactivation of the lexical representation of highly predicted words, the coupling between the alpha in these regions and the gamma in the left prefrontal region is likely to support the exchange and integration of information between frontal and posterior areas during language comprehension (Baggio & Hagoort, 2011). Because the coupling was only found for HC contexts, it might further facilitate the unification of upcoming words. Such frontal–posterior coupling is in line with the study by Park et al. (2015). They found that low-frequency brain oscillations in frontal brain region modulate the brain activity over the left auditory cortex during continuous speech perception. Therefore, our study for the first time demonstrates that the frontal gamma and posterior alpha plays an important role in supporting predictive top–down control over the processing of lexical information during reading.

In conclusion, stronger alpha power suppression was found for the highly compared with low constraining sentences just before predicted words. The sources of this effect were localized to the LIFC, left posterior temporal region, VWFA, extending to left hippocampus and right cerebellum. In addition, the incongruent words induced a stronger gamma power increase over left frontal and temporal regions as well as stronger alpha power decrease over left temporal, bilateral visual context, and cerebellum compared with the congruent words. Importantly, the left temporal and VWFA alpha power was negatively correlated with the frontal gamma power for the highly constraining sentences in both the prediction and integration periods of the sentence-final words. These results suggest that the involved areas functionally interact by cross-frequency coupling between alpha and gamma activity. Our study extends previous research on the function of alpha oscillations by demonstrating that decreased alpha power reflects the engagement of higher-level language areas and that language processing might be supported by the coupling between the alpha and gamma activities. In the future, it would be of great interest to conduct studies in which prediction is associated with preactivation in a representational-specific sense and relate those to alpha suppression.

Acknowledgments

This work was supported by the China Scholarship Council (CSC) and the Natural Science Foundation of China (grant 31540079). LW was supported by the Natural Science Foundation of China (31200849); PH was supported by the NWO Spinoza Prize, the Academy Professorship Award of the Netherlands Academy of Arts and Sciences, and the NWO Language in Interaction grant; OJ was supported by James S. McDonnell Foundation Understanding Human Cognition Collaborative Award (220020448) and the Royal Society Wolfson Research Merit Award. We thank Jim Herring for his assistance with the data analysis and Mathilde Bonnefond for helpful discussions.

Reprint requests should be sent to Lin Wang, Building 149, 13th Street, Charlestown, MA 02129, or via e-mail: lwang48@mgh.harvard.edu.

REFERENCES

Baggio, G., & Hagoort, P. (2011). The balance between memory and unification in semantics: A dynamic account of the N400. *Language and Cognitive Processes, 26,* 1338–1367.

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences, 11,* 280–289.

Bar, M. (2009). Predictions: A universal principle in the operation of the human brain. *Philosophical Transactions of the Royal Society, Series B, Biological Sciences, 364,* 1181–1182.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top–down facilitation of visual recognition. *Proceedings of the National Academy of Sciences, U.S.A., 103,* 449–454.

Bastiaansen, M. C. M., & Knösche, T. R. (2000). Tangential derivative mapping of axial MEG applied to event-related desynchronization research. *Clinical Neurophysiology, 111,* 1300–1305.

Bastiaansen, M. Mazaheri, A., & Jensen, O. (2012). Beyond ERPs: oscillatory neuronal dynamics. In *The Oxford handbook of event-related potential components* (pp. 31–50): Oxford University Press.

Baumgaertner, A., Weiller, C., & Büchel, C. (2002). Event-related fMRI reveals cortical sites involved in contextual sentence integration. *Neuroimage, 16,* 736–745.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation, 7,* 1129–1159.

Bonhage, C. E., Mueller, J. L., Friederici, A. D., & Fiebach, C. J. (2015). Combined eye tracking and fMRI reveals neural basis of linguistic predictions during sentence comprehension. *Cortex, 68,* 33–47.

Boudewyn, M. A., Carter, C., Long, D. L., Traxler, M. J., Lesh, T. A., Mangun, G. R., et al. (2017). Language context processing deficits in schizophrenia: The role of attentional engagement. *Neuropsychologia, 96,* 262–273.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition, 136,* 135–149.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36,* 181–204.

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences, 15,* 254–262.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8,* 1117–1121.

Dikker, S., & Pylkkänen, L. (2013). Predicting language: MEG evidence for lexical preactivation. *Brain and Language, 127,* 55–64.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top–down processing. *Nature Reviews Neuroscience, 2,* 704–716.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research, 1146,* 75–84.

Foxe, J. J., & Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in Psychology, 2,* 154.

Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences, 9,* 474–480.

Friston, K. J., Bastos, A. M., Pinotsis, D., & Litvak, V. (2015). LFP and oscillations—What do they tell us? *Current Opinion in Neurobiology, 31,* 1–6.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 364,* 1211–1221.

Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 694–699.

Haegens, S., Nácher, V., Luna, R., Romo, R., & Jensen, O. (2011). α-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. *Proceedings of the National Academy of Sciences, U.S.A., 108,* 19377–19382.

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. *Human Brain Mapping, 31,* 26–35.

Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences, 9,* 416–423.

Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology, 4,* 416.

Hagoort, P. (2016). MUC (memory, unification, control): A model on the neurobiology of language beyond single word processing. In G. Hickok & S. Small (Eds.), *Neurobiology of language* (pp. 339–347). Amsterdam: Elsever.

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

Hagoort, P., & Indefrey, P. (2014). The neurobiology of language beyond single words. *Annual Review of Neuroscience, 37,* 347–362.

Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., et al. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *Neuroimage, 17,* 1101–1116.

Hanslmayr, S., Aslan, A., Staudigl, T., Klimesch, W., Herrmann, C. S., & Bäuml, K.-H. (2007). Prestimulus oscillations predict visual perception performance between and within subjects. *Neuroimage, 37,* 1465–1473.

Harrison, L. M., Duggins, A., & Friston, K. J. (2006). Encoding uncertainty in the hippocampus. *Neural Networks, 19,* 535–546.

Hertrich, I., Mathiak, K., & Ackermann, H. (2016). Chapter 2—The role of the cerebellum in speech perception and language comprehension A2. In P. Mariën & M. Manto (Eds.), *The linguistic cerebellum* (pp. 33–50). San Diego, CA: Academic Press.

Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nature Neuroscience, 19,* 665–667.

Huettig, F., & Mani, N. (2016). Is prediction necessary to understand language? Probably not. *Language, Cognition and Neuroscience, 31,* 19–31.

Jensen, O., Bonnefond, M., & VanRullen, R. (2012). An oscillatory mechanism for prioritizing salient unattended stimuli. *Trends in Cognitive Sciences, 16,* 200–206.

Jensen, O., & Colgin, L. L. (2007). Cross-frequency coupling between neuronal oscillations. *Trends in Cognitive Sciences, 11,* 267–269.

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience, 4*, 1–8.

Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKeown, M. J., Iragui, V., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37,* 163–178.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods, 42,* 643–650.

Kim, A., & Lai, V. (2012). Rapid interactions between lexical semantic and word form analysis during word recognition in context: Evidence from ERPs. *Journal of Cognitive Neuroscience, 24,* 1104–1112.

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition–timing hypothesis. *Brain Research Reviews, 53,* 63–88.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31,* 32–59.

Laszlo, S., & Federmeier, K. D. (2011). The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology, 48,* 176–186.
Lau, E. F., & Nguyen, E. (2015). The role of temporal predictability in semantic expectation: An MEG investigation. *Cortex, 68,* 8–19.
Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Review Neuroscience, 9,* 920–933.
Leggio, M. G., Tedesco, A. M., Chiricozzi, F. R., Clausi, S., Orsini, A., & Molinari, M. (2008). Cognitive sequencing impairment in patients with focal or atrophic cerebellar damage. *Brain, 131,* 1332–1343.
Leske, S., Ruhnau, P., Frey, J., Lithari, C., Müller, N., Hartmann, T., et al. (2015). Prestimulus network integration of auditory cortex predisposes near-threshold perception independently of local excitability. *Cerebral Cortex, 25,* 4898–4907.
Leske, S., Tse, A., Oosterhof, N. N., Hartmann, T., Müller, N., Keil, J., et al. (2014). The strength of alpha and beta oscillations parametrically scale with the strength of an illusory auditory percept. *Neuroimage, 88,* 69–78.
Levy, J., Vidal, J. R., Oostenveld, R., FitzPatrick, I., Démonet, J.-F., & Fries, P. (2013). Alpha-band suppression in the visual word form area as a functional bottleneck to consciousness. *Neuroimage, 78,* 33–45.
Lewis, A. G., & Bastiaansen, M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex, 68,* 155–168.
Magazzini, L., Ruhnau, P., & Weisz, N. (2016). Alpha suppression and connectivity modulations in left temporal and parietal cortices index partial awareness of words. *Neuroimage, 133,* 279–287.
Magyari, L., Bastiaansen, M. C. M., de Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience, 26,* 2530–2539.
Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164,* 177–190.
Marvel, C. L., & Desmond, J. E. (2016). Chapter 3—The cerebellum and verbal working memory A2. In P. Mariën & M. Manto (Eds.), *The linguistic cerebellum* (pp. 51–62). San Diego, CA: Academic Press.
Mayer, A., Schwiedrzik, C. M., Wibral, M., Singer, W., & Melloni, L. (2015). Expecting to see a letter: Alpha oscillations as carriers of top–down sensory predictions. *Cerebral Cortex*, *26*, 3146–3160.
Mazaheri, A., Nieuwenhuis, I. L. C., van Dijk, H., & Jensen, O. (2009). Prestimulus alpha and mu activity predicts failure to inhibit motor responses. *Human Brain Mapping, 30,* 1791–1800.
Mitra, P. P., & Pesaran, B. (1999). Analysis of dynamic brain imaging data. *Biophysical Journal, 76,* 691–708.
Moberget, T., Gullesen, E. H., Andersson, S., Ivry, R. B., & Endestad, T. (2014). Generalized role for the cerebellum in encoding internal models: Evidence from semantic processing. *Journal of Neuroscience, 34,* 2871–2878.
Molinaro, N., Barraza, P., & Carreiras, M. (2013). Long-range neural synchronization supports fast and efficient reading: EEG correlates of processing expected words in sentences. *Neuroimage, 72,* 120–132.
Monsalve, I. F., Pérez, A., & Molinaro, N. (2014). Item parameters dissociate between expectation formats: A regression analysis of time–frequency decomposed EEG data. *Frontiers in Psychology, 5*, 1–12.
Nolte, G. (2003). The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Physics in Medicine and Biology, 48,* 3637.
Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex, 20,* 633–640.
Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011,* 156869.
Park, H., Ince, R. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top–down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology, 25,* 1649–1653.
Payne, L., & Sekuler, R. (2014). The importance of ignoring: Alpha oscillations protect selectivity. *Current Directions in Psychological Science, 23,* 171–177.
Penolazzi, B., Angrilli, A., & Job, R. (2009). Gamma EEG activity induced by semantic violation during sentence reading. *Neuroscience Letters, 465,* 74–78.
Pfurtscheller, G., Stancak, A., & Neuper, C. (1996). Event-related synchronization (ERS) in the alpha band—An electrophysiological correlate of cortical idling: A review. *International Journal of Psychophysiology, 24,* 39–46.
Piai, V., Anderson, K. L., Lin, J. J., Dewar, C., Parvizi, J., Dronkers, N. F., et al. (2016). Direct brain recordings reveal hippocampal rhythm underpinnings of language processing. *Proceedings of the National Academy of Sciences, U.S.A.*, *113*, 11366–11371.
Piai, V., Roelofs, A., & Maris, E. (2014). Oscillatory brain responses in spoken word production reflect lexical frequency and sentential constraint. *Neuropsychologia, 53,* 146–156.
Piai, V., Roelofs, A., Rommers, J., & Maris, E. (2015). Beta oscillations reflect memory and motor aspects of spoken word production. *Human Brain Mapping, 36,* 2767–2780.
Pisotta, I., & Molinari, M. (2014). Cerebellar contribution to feedforward control of locomotion. *Frontiers in Human Neuroscience, 8,* 475.
Price, C. J., & Devlin, J. T. (2011). The interactive account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences, 15,* 246–253.
Rao, R., & Ballard, D. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2,* 79–87.
Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., et al. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature Neuroscience, 11,* 426–428.
Rohenkohl, G., & Nobre, A. C. (2011). Alpha oscillations related to anticipatory attention follow temporal expectations. *Journal of Neuroscience, 31,* 14076–14084.
Rommers, J., Dickson, D. S., Norton, J. J. S., Wlotko, E. W., & Federmeier, K. D. (2017). Alpha and theta band dynamics related to sentential constraint and word expectancy. *Language, Cognition and Neuroscience, 32*, 576–589.
Schiffer, A.-M., Ahlheim, C., Wurm, M. F., & Schubotz, R. I. (2012). Surprised at all the entropy: Hippocampal, caudate and midbrain contributions to learning from prediction errors. *PLoS One, 7,* e36445.
Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J., Petersson, K. M., & Hagoort, P. (2008). Retrieval and unification of syntactic structure in sentence comprehension: An fMRI study using word-category ambiguity. *Cerebral Cortex, 19,* 1493–1503.
Spaak, E., Fonken, Y., Jensen, O., & de Lange, F. P. (2016). The neural mechanisms of prediction in visual search. *Cerebral Cortex*, *26,* 4327–4336.

Stolk, A., Todorovic, A., Schoffelen, J.-M., & Oostenveld, R. (2013). Online and offline tools for head movement compensation in MEG. *Neuroimage, 68,* 39–48.
Strange, B. A., Duggins, A., Penny, W., Dolan, R. J., & Friston, K. J. (2005). Information theory, novelty and hippocampal responses: Unpredicted or unpredictable? *Neural Networks, 18,* 225–230.
Strauß, A., Kotz, S. A., Scharinger, M., & Obleser, J. (2014). Alpha and theta brain oscillations index dissociable processes in spoken word recognition. *Neuroimage, 97,* 387–395.
Strauß, A., Wöstmann, M., & Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Frontiers in Human Neuroscience, 8*, 1–7.
Szewczyk, J. M., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language, 68,* 297–314.
Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology, 83,* 382–392.
Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 443–467.
van Dijk, H., Schoffelen, J.-M., Oostenveld, R., & Jensen, O. (2008). Prestimulus oscillatory activity in the alpha band predicts visual discrimination ability. *Journal of Neuroscience, 28,* 1816–1823.
Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83,* 176–190.
Wang, L., Zhu, Z., & Bastiaansen, M. (2012). Integration or predictability? A further specification of the functional role of gamma oscillations in language comprehension. *Frontiers in Psychology, 3,* 187.
Weiss, S., & Mueller, H. M. (2012). "Too many betas do not spoil the broth": The role of beta brain oscillations in language processing. *Frontiers in Psychology, 3,* 201.
Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16,* 1272–1288.
Wöstmann, M., Herrmann, B., Wilsch, A., & Obleser, J. (2015). Neural alpha dynamics in younger and older listeners reflect acoustic challenges and predictive benefits. *Journal of Neuroscience, 35,* 1458–1467.
Xiang, H.-D., Fonteijn, H. M., Norris, D. G., & Hagoort, P. (2010). Topographical functional connectivity pattern in the perisylvian language networks. *Cerebral Cortex, 20,* 549–560.
Ziegler, W. (2016). Chapter 1—The phonetic cerebellum: Cerebellar involvement in speech sound production A2. In P. Mariën & M. Manto (Eds.), *The linguistic cerebellum* (pp. 1–32). San Diego, CA: Academic Press.