Cortical tracking of surprisal during continuous speech comprehension

Hugo Weissbart[1], Katerina D. Kandylaki[1,2], and Tobias Reichenbach[1,*]

[1]Department of Bioengineering and Centre for Neurotechnology, Imperial College London, South Kensington Campus, SW7 2AZ, London, U.K.

[2]Faculty of Psychology and Neuroscience, Department of Neuropsychology and Psychopharmacology, Maastricht University, 6229 ER Maastricht, The Netherlands.

*To whom correspondence should be addressed (email: reichenbach@imperial.ac.uk)

Acknowledgements

Abstract

Speech comprehension requires rapid online processing of a continuous acoustic signal to extract structure and meaning. Previous studies on sentence comprehension have found neural correlates of the predictability of a word given its context, as well as a of the precision of such a prediction. However, they have focussed on single sentences and on particular words in those sentences. Moreover, they compared neural responses to words with low and high predictability, as well as with low and high precision. However, in speech comprehension a listener hears many successive words whose predictability and precision vary over a large range. Here we show that cortical activity in different frequency bands tracks word surprisal in continuous natural speech, and that this tracking is modulated by precision. We obtain these results through quantifying surprisal and precision from naturalistic speech using a deep neural network, and through relating these speech features to electroencephalographic (EEG) responses of human volunteers acquired during auditory story comprehension. We find significant cortical tracking of surprisal at low frequencies including the delta band as well as in the higher-frequency beta and gamma bands, and observe that the tracking is modulated by the precision. Our results pave the way to further investigate the neurobiology of natural speech comprehension.

Keywords

predictive processing, deep neural networks, language processing, entropy, surprisal, cortical tracking

Introduction

To understand spoken language, a listener must rapidly process information that unfolds over several timescales, including the duration of syllables at around 150 ms, words of about 300 ms, and phrases of 1 s (Giraud & Poeppel, 2012). Recent studies have shown that cortical activity in the delta, theta and gamma frequency bands tracks acoustic features of speech such as the speech envelope as well as phonemic features (Di Liberto, O'Sullivan, & Lalor, 2015; Ding et al., 2018; Ding & Simon, 2014; Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007; Zion Golumbic et al., 2013). This cortical tracking of speech features has accordingly been proposed to reflect neural mechanisms of speech processing, for instance an online segmentation of speech into acoustic speech tokens such as phonemes that occur on the time scale of a few hundreds of milliseconds (Giraud & Poeppel, 2012; Hyafil, Fontolan, Kabdebon, Gutkin, & Giraud, 2015).

The processing of higher-level linguistic information in speech may employ cortical tracking as well. Recent findings showed that cortical activity in the delta and theta frequency bands synchronized to sequential cues such as the rhythm of phrases and sentences in continuous speech (Ding, Melloni, Zhang, Tian, & Poeppel, 2016; Keitel, Gross, & Kayser, 2018), to hierarchical cues such as context-free grammar structure (Brennan & Hale, 2019), as well as to the semantic dissimilarity between successive words (Broderick, Anderson, Di Liberto, Crosse, & Lalor, 2018).

An important property of word sequences is that they can allow the prediction of an upcoming word, resulting in a word expectation. The degree to which a word can be predicted is referred to as precision and reflects the certainty with which a neural population generates its prediction. Predictions and precision are both closely related to putative implementations of predictive processing (Feldman & Friston, 2010; Heilbron & Chait, 2017; Kanai, Komura, Shipp, & Friston, 2015). Behavioral studies have indeed corroborated that the brain makes predictions about upcoming speech segments: words can be better distinguished from noise when transition probabilities between words are high rather than low (Miller, Heise, & Lichten, 1951), and a highly-expected word can be perceived as heard even when obscured by noise (Miller & Isard, 1963).

63   Neurophysiological research on event-related potentials elicited by a word in a sentence has

64   shown that the brain response to a word reflects the word expectancy through modulation of the N400

65   response (Kutas & Hillyard, 1984). Although this response has not been found to be further modulated

66   by the precision of the prediction (Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007), precision

67   can influence the neural power in the alpha and theta band (Rommers, Dickson, Norton, Wlotko, &

68   Federmeier, 2017). The power in the beta frequency band has been found to be reduced by semantic

69   and syntactic violations, and may therefore relate to word expectation as well (Bastiaansen, Magyari,

70   & Hagoort, 2010; Davidson & Indefrey, 2007; Kielar, Meltzer, Moreno, Alain, & Bialystok, 2014).

71   Gamma power has been observed to increase when a word is highly predictable but not when its

72   predictability is low (Molinaro, Barraza, & Carreiras, 2013; Wang, Zhu, & Bastiaansen, 2012).

73   However, these prior studies on neural correlates of word expectancy and precision have

74   focused on specific words in single sentences, contrasting words with high and low expectancy as well

75   as with high and low precision. But natural speech often consists of many sentences, and the expectancy

76   and the corresponding precision of successive words take a range of values that do not fall in only two

77   classes of 'high' and 'low'. It therefore remains unclear how neural responses to word expectancy and

78   precision correlate with this graded variability.

79   Furthermore, assessing the cortical responses to the linguistic features of successive words in

80   naturalistic stories allows to quantify the cortical tracking of these features. A recent investigation on

81   word predictability and hierarchical structure in naturalistic speech used such an approach to show

82   cortical tracking of word surprisal, but did not investigate an influence of precision and did not

83   investigate power modulation in higher frequency bands (Brennan & Hale, 2019; Frank & Willems,

84   2017).

85   Here we therefore set out to investigate cortical tracking, including through power modulation

86   in higher frequency bands, of word surprisal and the precision of word prediction in naturalistic stories.

87   The surprisal of a word denotes the log-transformed conditional probability of a word based on the

88   preceding context. The surprisal has been argued to relate to processing load (Levy, 2008) and predicts

89   reading time (Frank, Otten, Galli, & Vigliocco, 2015; Smith & Levy, 2013). Precision is the inverse of

the entropy of the conditional probability distribution over a close vocabulary set. We quantified word

surprisal and precision from naturalistic stories using language modelling as estimated by a recurrent

deep neural network, and then related the obtained word features to electroencephalographic (EEG)

responses of volunteers who listened to the stories.

Materials and Methods

Participants. 13 subjects (aged 25 ± 3 years, 6 females) participated in the experiment. The volunteers

were all right-handed native English speakers. They had no history of hearing or neurological

impairment. All participants provided written informed consent. The experimental procedures were

approved by the Imperial College Research Ethics Committee.

Experimental Design. We employed naturalistic speech narratives in the subjects' native language

(English). The experiment consisted of one session in which we measured electroencephalographic

(EEG) responses to the short stories *'Gilray's flower pot'* and *'My brother Henry'* by J.M. Barrie as well

as *'An undergraduate's aunt'* by F. Anstey (Patten, 1910). The stimuli were sourced from the public

domain 'librivox.org' and were spoken by a male voice. The corresponding text was obtained from the

project Gutenberg (http://www.gutenberg.org/ebooks/32846). The audio material was presented in 15

parts, each of which were 2.6 ± 0.43 min long. The total length of the stories was 40 min. After each

part of a story, participants answered comprehension questions about what they just heard. These

questions were presented as multiple-choice questions on a monitor. Participants were asked 30

questions in total.

Language modeling. We used computational linguistics methods to quantify linguistic features in the

employed stories. Specifically, we employed statistical language modelling to compute word frequency,

entropy and suprisal from the text of the stories.

Word frequency is a property of each individual word out of context, which was computed from

Google *N*-grams by using only the unigram values. This word feature is an estimate of the unconditional

115   probability of the occurrence of a word w, $P(w)$. We use the negative logarithm of this probability such

116   that all our information-theoretic word features are expressed in the same unit.

117   Both entropy and surprisal follow from conditional probabilities of a particular word given the

118   preceding words. We denote by $P(w_m|w_1, \ldots, w_{m-1})$ the conditional probability of the mth word in the

119   sequence, w_m, given the previous $m-1$ words $w_1, w_2, \ldots, w_{m-1}$. Taking the negative logarithm of this

120   probability yields the *surprisal* value $S(w_m)$ for that word:

121
$$S(w_m) = -\log\big(P(w_m|w_1, \ldots, w_{m-1})\big) \qquad (1)$$

122   The surprisal, also referred to as self-information or information content, quantifies the information

123   gain that an upcoming word generates with respect to the prior sequence of words. It can be related to

124   how unexpected a word is given the previous words in the sentence. Inasmuch as surprisal informs

125   about expected words, *precision* relates to the confidence about the predictions made (Koelsch, Vuust,

126   & Friston, 2018). A high precision translates in a high confidence about a word expectation, meaning

127   that the word is predictable.

128   The entropy $E(m)$ of the prediction of the mth word w_m, that is, the uncertainty for predicting

129   the word w_m from the context $(w_1, \ldots, w_{m-1})$, is given by the sum of the conditional probabilities for

130   each possible word w_k, weighted by the logarithm of this probability. In other words, the entropy is the

131   expected surprisal

132
$$E(m) = \sum_{w_k} P(w_k|w_1, \ldots, w_{m-1}) \ln \log\left[P(w_k|w_1, \ldots, w_{m-1})\right] \qquad (2)$$

133   The precision of the mth word w_m follows as the inverse of entropy $1/E(m)$. We note that the

134   precision of the mth word is not a function of that word itself, but of the probability distribution

135   of the words at that position.

136   The conditional probabilities for the different words in the sequence, given the preceding

137   words, were computed through a recurrent neural network language model (Bengio, Ducharme,

138   Vincent, & Jauvin, 2003; Graves, 2013). The network had a hidden layer with recurrent connections to

139   encode previous input. Such networks are particularly useful for processing sequences and have

140 previously been successfully applied to language modelling (Bengio et al., 2003; Graves, 2013). In

141 particular, a recurrent neural network can capture long-term dependencies, of variable length, by

142 encoding preceding words through its recurrent connection into the state of the hidden neurons. This is

143 enabled by a careful balance between short- and long-term memory and means that there is in principle

144 no limit on the number of preceding words that such a network can take into account (Pascanu, Mikolov,

145 & Bengio, 2013). This contrasts with *N*-gram language models, for instance, that are limited to a context

146 window of *N*-1 words (Brown, Desouza, Mercer, Pietra, & Lai, 1992).

147   The network was implemented using the feature-augmented recurrent neural network language

148 modelling toolkit (Mikolov, Kombrink, Burget, Černocký, & Khudanpur, 2011). To decrease the

149 computational time required for training, this toolbox assigns words to classes and factorizes the output

150 layer into a part that describes the probability of each class given the previous words, as well as another

151 part that describes the probability of each word within a class given the previous words. This

152 factorization yields a significant decrease in training time at a small cost to accuracy; importantly, the

153 network still computes the probability of individual words following the previous words (Mikolov et

154 al., 2011). We employed 300 classes. As an embedding layer we used the pre-trained global vectors for

155 word representation trained on the Wikipedia 2014 and the Gigaword 5 datasets (Pennington, Socher,

156 & Manning, 2014). The recurrent layer encompassed 350 hidden units. The source code was customized

157 to compute the entropy of each word, a feat that the original code did not allow. The neural network

158 was then trained on the *text8* dataset that consists of 100 MB of data from Wikipedia (Mahoney, 2011),

159 using back propagation through time, truncated to five words with a starting learning rate of 0.1. The

160 data was cleaned to remove punctuation, html tags, capitalisation and numbers before training. Since

161 the network can only train well on words that appear frequently enough in the training data to allow

162 meaningful training, we limited the vocabulary to the 35,000 most common words in the training

163 dataset. The remaining words were mapped to an 'unknown' token. Infrequent words in the stories, such

164 as compound nouns used for style, that appeared repeatedly throughout the stories did therefore not

165 obscure the results.

166   The output of the recurrent neural network was obtained from a softmax function, and could

167   therefore be interpreted as the probability distribution for an upcoming word given the preceding words

168   in the input sequence. The network was therefore trained to predict the next word, that is, to compute

169   an output that was as close as possible to a probability distribution that was one for the actual upcoming

170   word and zero for all remaining ones. The trained network was then run on the stories that the

171   participants heard. Precision and surprisal of each word were determined from the network's computed

172   probability distribution at the corresponding word through Equations (1) and (2).
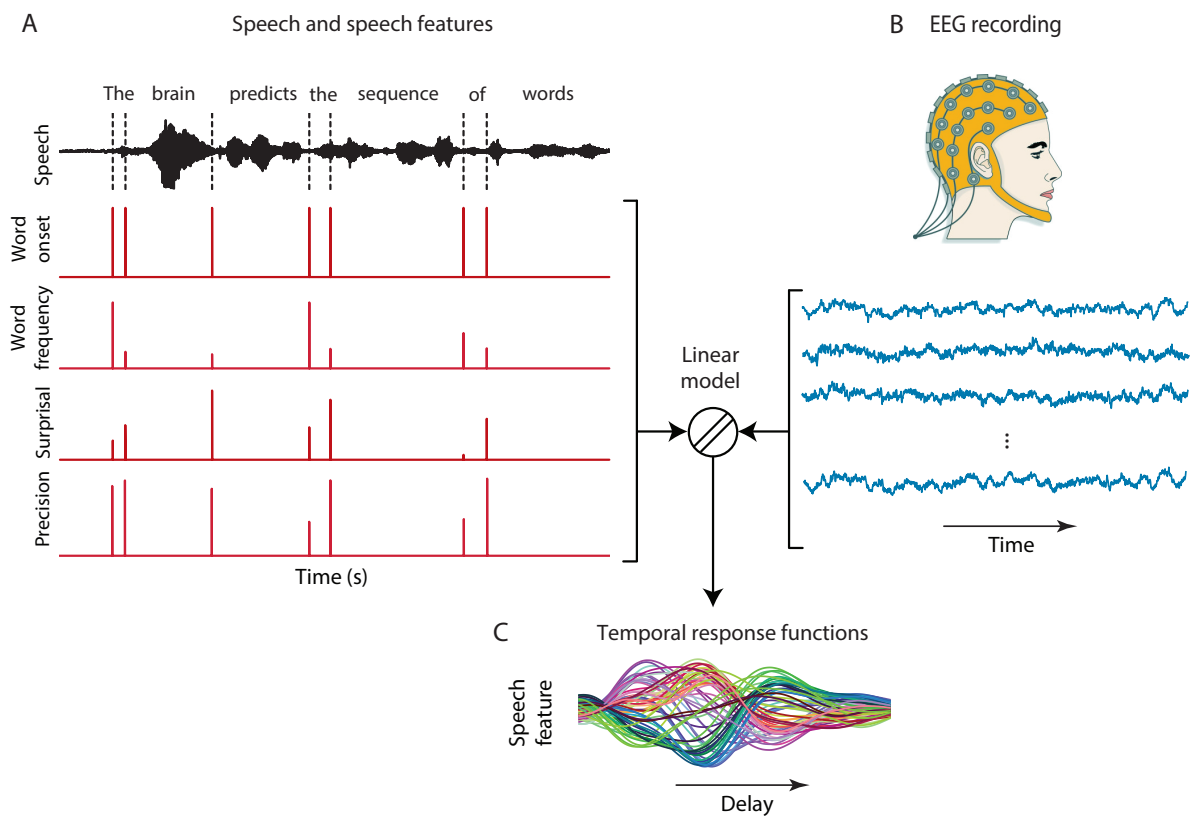
173

174   **Figure 1: Experimental overview.** (**A**), We employ continuous speech narratives and utilize speech

175   processing as well as language modelling to extract acoustic and linguistic features, namely word

176   onset, word frequency, precision and surprisal. (**B**), The participant's neural activity is recorded

177   through EEG while they listen to the stories. (**C**), We extract temporal response functions for each of

178   the four speech features through computing a linear model that estimates the EEG recordings from the

179   speech features.

180

181   **Speech features.** To relate surprisal and entropy to the EEG data, we constructed a time series

182   for each linguistic feature. We first aligned each word of the speech to the acoustic signal through forced

183   alignment using the Prosodylab-Aligner software (Gorman, Howell, & Wagner, 2011). We thereby

184   obtained the time at which each word began. To construct features for surprisal and for precision that

185   were aligned with the speech stimuli, we assigned each of the time points where a new word started a

186   spike of a magnitude that corresponded to the surprisal respectively precision of that word (Figure 1A).

187   A similar procedure has been employed recently for assessing neural responses to the semantic

188   dissimilarity of consecutive words (Broderick et al., 2018).

189   Because surprisal and precision are high-level linguistic features of speech, we sought to

190   ascertain that any putative cortical tracking of them could not be explained by lower-level features. To

191   this end we added three low-level speech features. First, cortical activity can track the onset of words,

192   which can partly be based on changes in the acoustics at word boundaries and partly result from the

193   brain's parsing of the acoustic signal to form discrete linguistic units (Brodbeck, Presacco, & Simon,

194   2018; Ding & Simon, 2014). To account for this onset response, we constructed a word onset feature

195   as a series of spikes, each of which had unit amplitude and was located at the onset of a word. Second,

196   we computed the word position within a sentence. The latter can be correlated with precision, as the

197   entropy tends to decrease across words within the sentence. The word position feature therefore served

198   as a control to ensure that the neural response to precision is distinct from any incremental processing

199   occurring throughout a sentence. Third, the frequency of a word in a given language, outside its context,

200   is a linguistic feature that acts as a prior probability for computing the probability of a word in a

201   sequence (Brodbeck et al., 2018). Word frequency can also interfere with surprisal: less frequent words

202   may indeed often be more surprising. To capture the share of the neural response that could be explained

203   away by word frequency, we included the latter as a third linguistic feature. This feature was computed

204   by scaling the amplitude of the spike at each word onset by the negative logarithm of the frequency of

205   the corresponding word. The logarithm was used such that word frequency and surprisal were expressed

206   in the same units.

207   Finally, to investigate a possible modulating effect that precision may have on surprisal, we

208   added an interaction term "Surprisal * Precision". This was computed by multiplying precision values

209   with surprisal such that the interaction feature effectively stands as a confidence-weighted version of

210   surprisal.

211   In summary, we computed five speech features: one acoustic feature, word onset, and four

212   linguistic features, word position in its sentence, word frequency, precision, and surprisal. To those, we

213   added the interaction term between surprisal and precision. Each feature was a time series of spikes,

214   which each spike being located at the onset of a word. The amplitude of the spike was constant for the

215   word onset feature. For each other feature it was scaled the corresponding value for each respective

216   linguistic feature. All values of the different linguistic features were standardized to have unit variance

217   and zero mean.

218   **EEG acquisition and pre-processing.** We recorded brain activity using 64 active electrodes (actiCAP,

219   BrainProducts, Germany) and a multi-channel EEG amplifier (actiCHamp, BrainProducts, Germany).

220   The presented sound was recorded simultaneously through an acoustic adapter (Acoustical Stimulator

221   Adapter and StimTrak, BrainProducts, Germany) and was used for aligning the EEG recordings to the

222   audio signals. Both the EEG and the audio data were acquired at a sampling rate of 1 kHz. The left ear

223   lobe was used as a reference for the EEG.

224   The EEG data was processed by first applying an anti-aliasing filter (Kaiser window, FIR filter,

225   cutoff -6 dB at 125 Hz, transition bandwidth 50 Hz, order 130) and by downsampling the data to 250

226   Hz to reduce the computation time of subsequent operations. A high-pass filter (Hanning window, sinc

227   type I linear phase FIR filter, cutoff -6 dB at 0.3Hz, transition bandwidth 0.15 Hz, order 5168), was

228   then applied to every channel to remove non-stationary trends such as slow drifts and offsets. Bad

229   channels were identified using the procedure 'clean_rawdata' from the EEGLAB plugin ASR (Artefact

230   Subspace Reconstruction); they were then removed and interpolated with spherical interpolation. All

231   channels were then referenced to the channel average. We subsequently ran an ICA decomposition and

232   removed artifacts from eye blink, eyes movement as well as muscle motion by visual inspection of the

233   ICA components. The cleaned data were low-pass filtered (Hamming window, linear phase FIR filter,

234   cutoff -6 dB at 62 Hz, transition bandwidth 10 Hz, order 138) and further down-sampled to 125 Hz.

235   The filtered EEG data therefore contained the broad frequency range from 0.3 Hz to 62 Hz.

236   We computed temporal response functions (TRFs) from EEG data in several frequency bands.

237   The TRFs followed from a linear forward model that expressed the EEG signal at each electrode as a

238   linear combination of the speech features shifted by different latencies (Broderick et al., 2018; Ding &

239   Simon, 2012). We used FIR type I filters, designed with the synced windowed method, and employing

240   a hamming window. We filtered the EEG data in several frequency bands of interest: delta band (low-

241   pass filter, cutoff at 4.5 Hz, filter order 132), theta band (band-pass filter, cutoff frequencies at 4 Hz and

242   8 Hz, order 206), alpha band (band-pass filter, cutoff frequencies at 8 and 12 Hz, order 206), beta band

243   (band-pass filter, cutoff 20 Hz and 30 Hz, order 82) and gamma band (cutoff at 30 and 60 Hz, order

244   164). For every frequency band other than delta, we computed the power modulation by taking the

245   absolute value of the Hilbert transform of the band passed data and further band-pass filtered it between

246   0.5 Hz and 20 Hz (filter order 824) to remove the DC offset and higher frequencies that do not occur in

247   the speech features.

248   **EEG data analysis.** To relate the speech features to the EEG data, we used a linear spatio-temporal

249   forward model that reconstructed the EEG recordings from the acoustic feature and the three linguistic

250   features, shifted by different delays (Figure 1). Such an approach has recently been used successfully

251   for assessing the cortical tracking of the speech envelope, phonemic information as well as semantic

252   dissimilarity of words in speech (Broderick et al., 2018; Di Liberto et al., 2015; Ding & Simon, 2012).

253   The coefficients resulting from this regression constitute the TRFs that inform on the brain's response

254   to each feature at different latencies.

255   In particular, the forward model sought to express the pre-processed EEG recordings

256   $\{x_i(t_n)\}_{i=1}^N$ of the $N = 64$ channels at each time instance t_n through the time series $\left\{y_j(t_n - \tau_k)\right\}_{j=1}^F$ of

257   the $F = 6$ speech features word onset, word frequency, word position, word precision, word surprisal,

258   and the product of surprisal and entropy, shifted by T different delays $\{\tau_k\}_{k=1}^T$:

259

$$\hat{x}_i(t_n) = \sum_{j=1}^{6} \sum_{k=1}^{T} \beta_{ij}(\tau_k) y_j(t_n - \tau_k) \tag{3}$$

260

261

262   We hereby considered equally spaced delays $\{\tau_k\}_{k=1}^{T}$ that ranged from -400 ms to 1,100 ms. At the

263   sampling rate of 125 Hz this yielded a number of $T = 188$ lags. The obtained estimate for the EEG

264   channel i is denoted by $\hat{x}_i$. The coefficient $\beta_{ij}(\tau_k)$ is the TRF for the ith EEG channel and speech feature

265   j at the latency τ_k. The preprocessed EEG recording $\{x_i(t_n)\}_{i=1}^{N}$ was either the EEG signal in the delta

266   band, or the power of the EEG signal in the higher frequency bands. We computed the TRFs for each

267   participant separately, leading to a set of TRFs on which we could apply group-level statistical analysis

268   as described below. We then also computed the population average of the subject-specific TRFs; the

269   population averages are shown in the figures.

270   The different speech features that we employed were partly correlated. The largest correlation

271   emerged between surprisal and the interaction term "surprisal * precision", at a value of 0.61. We

272   wondered if these correlations would hinder the EEG analysis, and in particular if they would obscure

273   the neural responses to the individual speech features through the linear regression analysis, an issue

274   known as multicollinearity (Chatterjee & Hadi, 2015; Kumar, 1975). A high multicollinearity between

275   features could result in higher variance or leakage between the coefficient $\beta_{ij}(\tau_k)$. However, the

276   Frisch–Waugh–Lovell theorem from econometrics states that linear regression based on correlated

277   features yields the same results as when the features are first orthogonalized, that is, decorrelated (Frisch

278   & Waugh, 1933; Lovell, 2008). In addition, in our implementation of the multiple linear regression we

279   used a singular value decomposition of the design matrix of time-lagged features, resulting in

280   transformed features that were mutually uncorrelated (Klema & Laub, 1980). The correlation of the

281   features was therefore not problematic. The only issue that multicolinearity can cause is significantly

282   increased variance for each $\beta_{ij}(\tau_k)$ estimate, which typically emerges when the variance inflation factor

283   (VIF) is above 5. For our speech features we obtained VIFs between 1.22-2.25, indicating that increased

284   noise due to correlated features is not an issue.

285   As an additional control that our TRFs did not contain leakage from responses to different

286   features, we developed a null model that was employed to assess the statistical significance of the actual

287   TRFs (see below). The null model was constructed such that a potential leakage between features would

288   appear similarly both in the actual model and in the null model, and therefore would not result in

289   statistically significant results. It follows that any statistically-significant part in the TRFs that we

290   obtained did not result from leakage between the features.

291   **Statistical significance.** In order to determine the statistical significance of the estimated TRFs, we

292   determined chance-level TRFs as a null model. The chance-level TRFs were computed by constructing

293   unrelated speech features, and by relating these to the EEG recordings in the same way as for the

294   computation of the actual TRFs. To establish chance-level linguistic TRFs, only the linguistic

295   information of interest contained in the spike amplitude of the speech features but not the acoustic

296   information in the spike timing needed to be unrelated to the EEG. We therefore constructed unrelated

297   speech features by keeping the timing of the spikes identical to those in the true model. The speech

298   feature that described word onsets was therefore not altered. However, we changed the amplitude of the

299   spikes for the other linguistic speech features by taking their values from an unrelated story, that is, a

300   story that was not aligned with the EEG data. To obtain a large number of null models, we also

301   considered permutations of our 15 story parts. Through permutating entire story parts, and not the order

302   of individual words, the statistical relationship between the linguistic features of successive words was

303   conserved. Because we kept the timing of the spikes in the null model as in the actual stories, the

304   obtained null model could only be used to determine the significance of the neural responses to the

305   linguistic features, but not for those to the acoustic word onset.

306   The actual TRFs were then analyzed for statistical significance through comparison to 1,000

307   null models. The comparison was obtained from a permutation test together with cluster-based

308   correction for multiple comparison (Oostenveld, Fries, Maris, & Schoffelen, 2011), where only clusters

309   of at least four electrodes were kept. Specifically, we used the function *spatio_temporal_cluster_test*

310   from the MNE python library. The statistic for each model coefficient, at each electrode and each lag,

311   were computed using the empirical distribution formed by values from the null models, setting the

312 threshold at the 99th percentile of the null distribution. The cluster-level *p*-values were computed and

313 we considered only clusters with a *p*-value greater than 0.05/10. We hereby used the Bonferroni

314 correction to account for the ten different tests that reflected the different frequency bands and the

315 different linguistic features.

316 **Data availability.** The EEG data from all subjects together with the corresponding speech features are

317 available on figshare.com (10.6084/m9.figshare.9033983.v1). An exemplary script for computing TRFs

318 can be obtained from figshare as well (10.6084/m9.figshare.9034481.v1).

319

Results

321 **Behavioural assessment.** We first assessed to which degree the participants understood the stories

322 through asking them comprehension questions. These questions were answered with an average of 96%

323 accuracy, evidencing that the volunteers consistently understood the speech and paid attention.

324 **Cortical tracking of acoustic and linguistic speech features.** The cortical tracking of the speech

325 features can be found in different frequency bands. First, because all four features relate to words, the

326 frequency range of the features is similar to the rate of words in speech. The latter is about $1 - 4$ Hz and

327 corresponds to the delta frequency range. Cortical activity at low frequencies, including the delta

328 frequency band, can therefore be evoked by or entrain to the rhythm set by the acoustic and linguistic

329 word features. Second, the amplitude of the neural activity in higher frequency bands can be modulated

330 by the speech features. This may in particular occur for the theta band (4 - 8 Hz), the alpha band ($8 -$

331 12 Hz), the beta frequency band ($20 - 30$ Hz, and the gamma frequency band ($30 - 100$ Hz), the power

332 of which can be modulated by prediction in sentence comprehensio (Bastiaansen & Hagoort, 2006;

333 Bastiaansen et al., 2010; Wang, Jensen, et al., 2012; Weiss & Mueller, 2012).

334 We started by quantifying the neural tracking of the word features at low frequencies. We found

335 neural responses to word frequency between delays of 300 - 610 ms (Figure 2). The topographic plots

336 of the responses show large differences between the temporal scalp areas on the one hand and the

337 parietal and occipital areas on the other hand.

338
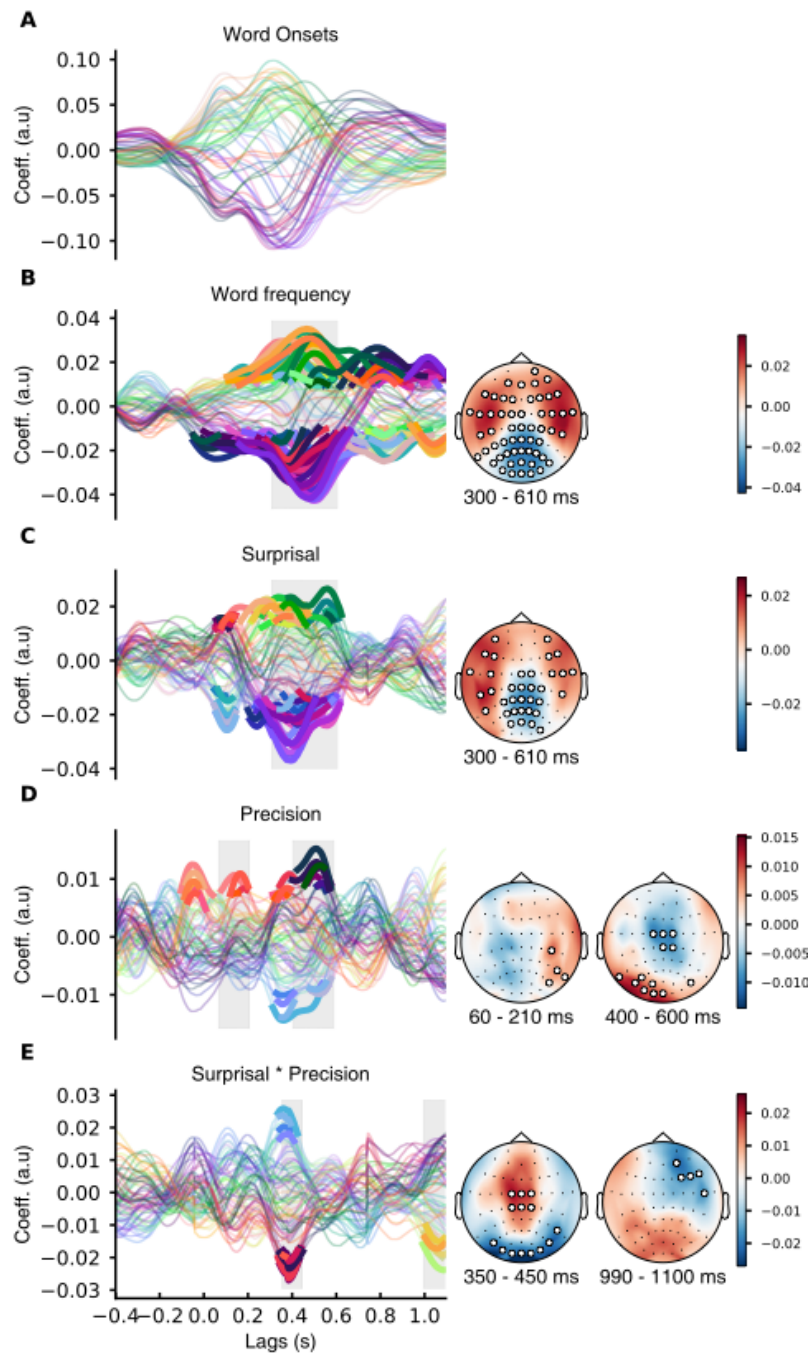
Figure 2: Temporal response functions (TRFs) for acoustic and linguistic speech features. The temporal response functions for each electrode are shown in bold at time instances where they are significant compared to a null model that is based on shuffled data. EEG channels that yield a significant response within a particular range of delays, highlighted in grey, are indicated in white in the topographic plots. (**A**), The responses to the word onset appear as insignificant due to the construction of the null model. (**B,C**), We obtain significant neural responses to word frequency as

346   well as surprisal for delays around 400 ms (**D**), Significant neural responses to precision arise around

347   delays of 100 ms as well as around 500 ms. (**E**), The interaction between surprisal and precision leads

348   to a neural response at a delay of 400 ms as well as at a long delay of 1,000 ms.

349
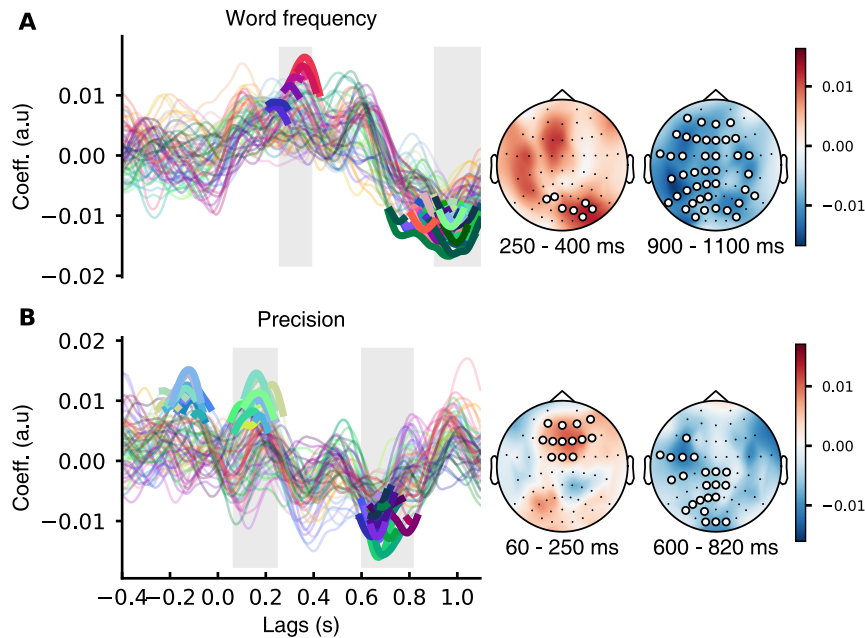

350

351   **Figure 3: Neural responses in the theta frequency band.** (**A**), Word frequency is positively

352   correlated to theta power at a delay of 300 ms, and is negatively correlated at a delay of 1,000 ms. (**B**),

353   Words that can be predicted with higher precision lead to an increased theta power at 150 ms and a

354   decreased theta power at a latency of 700 ms.

355

356   Importantly, we found significant responses to the word surprisal around a delay of 450 ms

357   (Figure 2). These responses emerged predominantly in the EEG channels on the temporal and occipital

358   scalp areas and were lateralized on the left hemisphere. Precision was tracked by cortical activity at

359   delays of around 100 ms and around 500 ms. Moreover, we observed a significant neural response to

360   the interaction of surprisal and precision, at an earlier latency of around 400 ms and at a longer latency

361   of around 1,000 ms.

362   We also computed the modulation of the power in the theta band, the alpha band, the lower and

363   higher beta band as well as in the gamma band by the acoustic and linguistic features (Figures 4, 5).

364   While the power in the alpha band and in the lower beta band was not significantly related to the

365 linguistic features, the power in the theta band was shaped by word frequency at delays of around 300

366 ms and around 1,000 ms (Figure 3). Furthermore, the power in the theta band was significantly

367 decreased by precision at delays of about 700 ms.
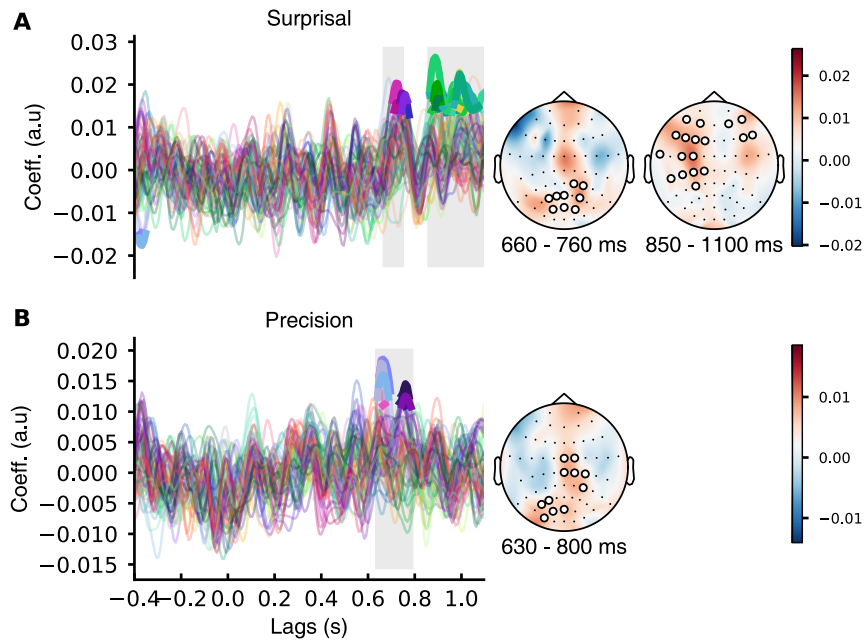
368

369 **Figure 4: Neural responses in the higher beta frequency band.** (**A**), There are significant neural

370 responses to surprisal, emerging at delays of 700 ms and 1,000 ms. (**B**), Precision causes an increased

371 power in the higher beta band activity around a delay of 700 ms.

372

373 The power in the higher beta band correlated positively with surprisal at delays of around 700

374 ms and 1,000 ms (Figure 4). At the latter delay, the influence of surprisal was strongest at the left

375 temporal channels. Moreover, the power in the higher beta band was modulated by precision at a delay

376 of about 700ms, with the main contributions coming from the occipital channels.

377 The power in the gamma band was increased by words with higher surprisal at the long latency

378 of around 1,000 ms, mainly for the left temporal channels (Figure 5). The interaction of surprisal and

379 precision shaped the gamma power as well, at the early delay of about 0 ms.
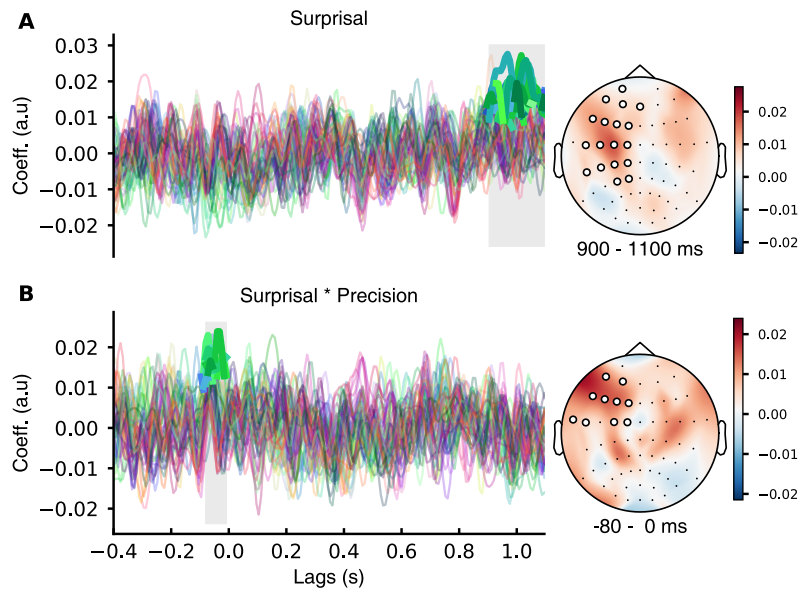
Figure 5: Tracking of surprisal by gamma-band activity. (A), The gamma activity is decreased at around 1,000 ms, mostly in the left temporal and frontal scalp areas. **(B)**, The interaction between precision and surprisal leads to a modulation of the gamma power at the latency of around 0 ms. This modulation occurs predominantly for left temporal and frontal channels as well.

Discussion

We have shown that cortical activity tracks the surprisal of words in speech comprehension. Such cortical tracking has emerged at low frequencies, that is, within the delta band that encompasses a similar frequency range as the rate of words in speech. Importantly, we found that the neural activity in the faster theta, beta and gamma frequency bands tracks the surprisal as well. These frequency bands have previously been suggested to be involved in the bottom-up and top-down propagation of predictions and prediction errors (Lewis & Bastiaansen, 2015).

We have further demonstrated that the cortical tracking of word surprisal is modulated by the precision: the interaction between surprisal and precision lead to responses both in the slow delta band as well as in the power of the faster gamma band. In particular, word predictions that are made with high precision but then lead to large surprisal cause an increased gamma power at zero lag. However, as opposed to a previous study on event-related potentials, we did not observe a significant effect in the theta or alpha bands (Rommers et al., 2017). This difference may be due to our use of naturalistic

399 stimuli, and the inclusion of all words in the analysis, while the previous study used specialized

400 sentences with final words that had either high or low surprisal, and either high or low precision.

401   The cortical tracking of surprisal may indicate predictive processing by the brain. Predictive

402 processing is a framework for perception in which it is assumed that the brain infers hypotheses about

403 a sensory input and that the hypotheses are constantly updated as new sensory information becomes

404 available (Bendixen, SanMiguel, & Schröger, 2012; Friston, 2010; Friston & Kiebel, 2009; Kanai et

405 al., 2015). In particular, the surprisal of a word reflects a prediction error, a key quantity in the

406 framework of predictive coding (Friston, 2010). However, the expectancy of a word based on previous

407 words also correlates with the plausibility of a word in a particular context (DeLong, Quante, & Kutas,

408 2014; Nieuwland et al., 2019). Further studies are therefore required to disentangle neural correlates of

409 actual word prediction from those that do not require predictive processing, such as word plausibility.

410   The surprisal of a word can reflect both its semantic as well as syntactic information, and

411 previous investigations into the neurobiological mechanisms of language comprehension have

412 manipulated both independently (Henderson, Choi, Lowder, & Ferreira, 2016; Humphries, Binder,

413 Medler, & Liebenthal, 2006). In contrast, our approach has taken a naturalistic and holistic approach to

414 surprisal; we employed natural speech without manipulations combined with statistical learning of a

415 rich variety of natural language cues through a recurrent neural network. Because the neural network

416 infers both syntactic rules as well as semantic information from the training of the speech material, the

417 reported neural response to word surprisal can reflect both semantic as well as syntactic information

418 (Collobert et al., 2011).

419   It is instructive to compare the reported neural responses to surprisal to the well-characterized

420 event-related responses that can be elicited by violations of semantics, syntax or morphology in

421 sentences. In particular, semantic violations can cause the N400 response, a negativity at $200 - 500$ ms

422 at the central and parietal scalp area (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980). Syntactic

423 anomalies due to ungrammaticality or temporary misanalysis elicit the P600, a broad positive potential

424 that is located at the posterior scalp area and arises around 600 ms after the anomaly (Friederici, Pfeifer,

425 & Hahne, 1993; Hagoort & Brown, 2000). More specific syntactic anomalies can lead to negative

426   potentials that occur anteriorly and that can be left lateralized, either occurring at 300 – 500 ms ((L)AN)

427   or earlier, at 125 – 150 ms (ELAN) (Friederici, 2002; Rösler, Pechmann, Streb, Röder, & Hennighausen,

428   1998; Steinhauer & Drury, 2012; Van Den Brink, Brown, & Hagoort, 2001).

429   These event-related potentials (ERPs) do presumably not reflect the activation of single static

430   neural sources, but rather waves of neural activity that propagate in time across different brain areas

431   (Kutas & Federmeier, 2011; Maess, Herrmann, Hahne, Nakamura, & Friederici, 2006; Tse et al., 2007).

432   In the case of the N400, for instance, this wave of activity starts at about 250 ms in the left superior

433   temporal gyrus, and then propagates to the left temporal lobe by 365 ms as well as to both frontal lobes

434   by 500 ms (Halgren et al., 2002; Helenius, Salmelin, Service, & Connolly, 1998; Van Petten & Luka,

435   2006). A recent theory suggests that this wave of activity reflects reverberating activity within the

436   inferior, middle and superior temporal gyri that corresponds to the activation of lexical information, the

437   formation of context and the unification of an upcoming word with the context (Baggio & Hagoort,

438   2011).

439   The spatio-temporal characteristics of the responses to surprisal that we have measured here

440   share certain similarities with these ERPs. In particular, we have found neural responses to surprisal at

441   latencies between 300 ms and 600 ms. These responses show a central-parietal negativity that is

442   reminiscent of the N400. However, other features of the neural responses that we describe here appear

443   distinct from these ERPs. The neural response to surprisal in the delta band at the latency of 600 ms

444   does, for instance, not display the posterior positivity of the P600. Moreover, we have identified late

445   responses around 700 ms and 1,000 ms. We have also shown that neural responses to surprisal arise in

446   various frequency bands, beyond the delta band that matters for the ERPs. However, a further

447   comparison of the neural response to surprisal to the related ERPs is hindered by the lack of spatial

448   resolution offered by EEG recordings. Future neuroimaging studies using intracranial recordings or

449   magnetoencephalography may localize the sources of the neural response to surprisal that we have

450   measured here and quantify potential shared sources with the ERPs.

451   The difference of the cortical tracking of surprisal to the well-known neural correlates of

452   semantic, syntactic or morphological anomalies, and in particular the late responses at a delay of around

453   one second, may come as a result of our use of natural speech that differs from the artificially

454   constructed and tightly controlled stimuli used to measure ERPs. First, in our experiment the subjects

455   encountered no violations of semantics, syntax and morphology, but instead heard naturalistic speech,

456   within which the words occurred in context. Second, our stimuli did not contain artificial manipulations

457   of word surprisal or precision. Instead of altering the stimuli, we focused on quantifying surprisal and

458   precision as they varied naturally in the presented stories. Third, we assessed the responses to surprisal

459   and precision at each word in the story, and hence for words in every sentence position, rather than for

460   words at a particular position within each sentence. Because we accounted for word position through a

461   corresponding control feature, we avoided the possibility of sentence position having an effect on the

462   results (Bastiaansen et al., 2010). Fourth, we did not employ isolated sentences but continuous stories

463   so that information of integration occurred over time scales exceeding a few seconds.

464   While our EEG recordings showed the cortical tracking of surprisal in different frequency

465   bands, they did not allow us to precisely localize the sources of the activity in the cortex. Pairing EEG

466   with functional magnetic resonance imaging (fMRI) or employing magnetoencephalographic (MEG)

467   may allow to add spatial information to the temporal tracking that we have assessed here. A recent fMRI

468   study, for instance, found that the left inferior temporal sulcus, the bilateral posterior superior temporal

469   gyri, and the right amygdala responded to surprisal during natural language comprehension, while the

470   left ventral premotor cortex and the left inferior parietal lobule responded to entropy (Willems et al.,

471   2015). Another recent magnetoencephalographic (MEG) measurement of the brain's natural speech

472   processing found that entropy and surprisal play a role in the assembly of phonemes into words, and

473   involves brain areas such as core auditory cortex and the superior temporal sulcus (Brodbeck et al.,

474   2018). Combining the temporal precision of EEG with the spatial precision of fMRI, or harnessing the

475   ability of MEG to locate neural sources temporally and spatially, will allow to further clarify the spatio-

476   temporal mechanisms of natural language comprehension in the brain.

477   In summary, we showed that neural responses to word surprisal can be measured from EEG

478   responses to naturalistic stories. Our results demonstrate that both the slow delta band as well as the

479   power in higher frequency bands, in particular the theta and higher beta band, are shaped by surprisal.

480   Moreover, we also showed that the neural response to surprisal is modulated by the precision of a

481   prediction. In particular, predictions made with high precision which lead to high surprisal modulate

482   gamma power in the left temporal and frontal scalp areas. In addition, we also demonstrated that neural

483   activity in the delta, theta and beta frequency bands is shaped by the precision of word prediction

484   directly. These responses arise at different latencies and at different scalp areas, suggesting a rich spatio-

485   temporal dynamics of neural activity related to word prediction.

486

487   Competing financial interests

488   The authors declare no competing financial interests.

489

References

490   **References**

491   Baggio, G., & Hagoort, P. (2011). The balance between memory and unification in semantics: A

492   dynamic account of the N400. *Lang. Cogn. Proces., 26*(9), 1338-1367.

493   Bastiaansen, M., & Hagoort, P. (2006). Oscillatory neuronal dynamics during language

494   comprehension. *Progr. Brain Res., 159*, 179-196.

495   Bastiaansen, M., Magyari, L., & Hagoort, P. (2010). Syntactic unification operations are reflected in

496   oscillatory dynamics during on-line sentence comprehension. *J. Cogn. Neurosci., 22*(7),

497   1333-1347.

498   Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for

499   predictive processing in audition: a review. *Int. J. Psychophysiol., 83*(2), 120-131.

500   Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *J.

501   Mach. Learn. Res., 3*(Feb), 1137-1155.

502   Brennan, J. R., & Hale, J. T. (2019). Hierarchical structure guides rapid linguistic predictions during

503   naturalistic listening. *PloS One, 14*(1), e0207741.

504   Brodbeck, C., Presacco, A., & Simon, J. Z. (2018). Neural source dynamics of brain responses to

505   continuous stimuli: speech processing from acoustics to comprehension. *NeuroImage, 172*,

506   162-174.

507   Broderick, M. P., Anderson, A. J., Di Liberto, G. M., Crosse, M. J., & Lalor, E. C. (2018).

508   Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural,

509   narrative speech. *Curr. Biol., 28*, 803.

510   Brown, P. F., Desouza, P. V., Mercer, R. L., Pietra, V. J. D., & Lai, J. C. (1992). Class-based n-gram

511   models of natural language. *Comp. Ling., 18*(4), 467-479.

512   Chatterjee, S., & Hadi, A. S. (2015). *Regression analysis by example*: John Wiley & Sons.

513   Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural

514   language processing (almost) from scratch. *J. Mach. Learn. Res., 12*(Aug), 2493-2537.

515   Davidson, D. J., & Indefrey, P. (2007). An inverse relation between event-related and time–frequency

516   violation responses in sentence processing. *Brain Res., 1158*, 81-92.

517   DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP

518   positivities during written sentence comprehension. *Neuropsychol., 61*, 150-162.

519   Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-Frequency Cortical Entrainment to

520   Speech Reflects Phoneme-Level Processing. *Curr. Biol., 25*(19), 2457-2465.

521   Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical

522   linguistic structures in connected speech. *Nat. Neurosci., 19*(1), 158.

523   Ding, N., Pan, X., Luo, C., Su, N., Zhang, W., & Zhang, J. (2018). Attention is required for

524   knowledge-based sequential grouping: Insights from the integration of syllables into words.

525   *Journal of Neuroscience, 38*(5), 1178-1188.

526   Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to

527   competing speakers. *Proc. Natl. Acad. Sci. U.S.A., 109*(29), 11854-11859.

528   Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and

529   interpretations. *Front. Hum. Neurosci., 8*, 311.

530   Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of

531   sentential constraint on word processing. *Brain Res., 1146*, 75-84.

532   Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. *Front. Hum. Neurosci., 4*,

533   215.

534   Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of

535   information conveyed by words in sentences. *Brain Lang., 140*, 1-11.

536   Frank, S. L., & Willems, R. M. (2017). Word predictability and semantic similarity show distinct

537   patterns of brain activity during language comprehension. *Lang. Cogn. Neurosci., 32*(9),

538   1192-1203.

539   Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends Cogn. Sci.,*

540   *6*(2), 78-84.

541   Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech

542   processing: Effects of semantic, morphological and syntactic violations. *Cogn. Brain Res.,*

543   *1*(3), 183-192.

544   Frisch, R., & Waugh, F. V. (1933). Partial time regressions as compared with individual trends.
545   *Econom. J. Econom. Soc.*, 387-401.

546   Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci., 11*(2), 127.

547   Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Phil. Trans. Roy.*
548   *Soc. B: Biol. Sci., 364*(1521), 1211-1221.

549   Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging
550   computational principles and operations. *Nat. Neurosci., 15*, 511-517.

551   Gorman, K., Howell, J., & Wagner, M. (2011). Prosodylab-aligner: A tool for forced alignment of
552   laboratory speech. *Canad. Acoust., 39*(3), 192-193.

553   Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint*
554   *arXiv:1308.0850.*

555   Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the
556   P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation.
557   *Neuropsychol., 38*(11), 1531-1549.

558   Halgren, E., Dhond, R. P., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., et al.
559   (2002). N400-like magnetoencephalography responses modulated by semantic context, word
560   frequency, and lexical class in sentences. *Neuroimage, 17*(3), 1101-1116.

561   Heilbron, M., & Chait, M. (2017). Great expectations: is there evidence for predictive coding in
562   auditory cortex? *Neuroscience.*

563   Helenius, P., Salmelin, R., Service, E., & Connolly, J. F. (1998). Distinct time courses of word and
564   context comprehension in the left temporal cortex. *Brain, 121*(6), 1133-1142.

565   Henderson, J. M., Choi, W., Lowder, M. W., & Ferreira, F. (2016). Language structure in the brain: A
566   fixation-related fMRI study of syntactic surprisal in reading. *Neuroim., 132*, 293-300.

567   Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2006). Syntactic and semantic
568   modulation of neural activity during auditory sentence comprehension. *J. Cogn. Neurosci.,*
569   *18*(4), 665-679.

570   Hyafil, A., Fontolan, L., Kabdebon, C., Gutkin, B., & Giraud, A.-L. (2015). Speech encoding by
571   coupled cortical theta and gamma oscillations. *Elife, 4*, e06213.

572 Kanai, R., Komura, Y., Shipp, S., & Friston, K. (2015). Cerebral hierarchies: predictive processing,
573   precision and the pulvinar. *Phil. Trans. Roy. Soc. B: Biol. Sci., 370*(1668), 20140169.

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

576 Kielar, A., Meltzer, J. A., Moreno, S., Alain, C., & Bialystok, E. (2014). Oscillatory responses to
577   semantic and syntactic violations. *J. Cogn. Neurosci., 26*(12), 2840-2862.

578 Klema, V., & Laub, A. (1980). The singular value decomposition: Its computation and some
579   applications. *IEEE Trans. Aut. Contr., 25*(2), 164-176.

580 Koelsch, S., Vuust, P., & Friston, K. (2018). Predictive processes and the peculiar case of music.
581   *Trends Cogn. Sci.*

582 Kumar, T. K. (1975). Multicollinearity in regression analysis. *Rev. Econ. Stat., 57*(3), 365-366.

583 Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400
584   component of the event-related brain potential (ERP). *Ann. Rev. Psychol., 62*, 621-647.

585 Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic
586   incongruity. *Science, 207*(4427), 203-205.

587 Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and
588   semantic association. *Nature, 307*(5947), 161.

589 Lakatos, P., Chen, C.-M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal
590   oscillations and multisensory interaction in primary auditory cortex. *Neuron, 53*(2), 279-292.

591 Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126-1177.

592 Lewis, A. G., & Bastiaansen, M. (2015). A predictive coding framework for rapid neural dynamics
593   during sentence-level language comprehension. *Cortex, 68*, 155-168.

594 Lovell, M. C. (2008). A simple proof of the FWL theorem. *J. Econ. Ed., 39*(1), 88-91.

595 Maess, B., Herrmann, C. S., Hahne, A., Nakamura, A., & Friederici, A. D. (2006). Localizing the
596   distributed language network responsible for the N400 measured by MEG during auditory
597   sentence processing. *Brain Res., 1096*(1), 163-172.

598 Mahoney, M. (2011). About the test data. *Available: http://mattmahoney. net.*

599   Mikolov, T., Kombrink, S., Burget, L., Černocký, J., & Khudanpur, S. (2011). *Extensions of recurrent*
600   *neural network language model.* Paper presented at the IEEE Conf. Acoust. Speech Sign.
601   Proc. (ICASSP).

602   Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the
603   context of the test materials. *J. Exp. Psych., 41*(5), 329.

604   Miller, G. A., & Isard, S. (1963). Some perceptual consequences of linguistic rules. *J. Verb. Learn.*
605   *Verb. Behav., 2*(3), 217-228.

606   Molinaro, N., Barraza, P., & Carreiras, M. (2013). Long-range neural synchronization supports fast
607   and efficient reading: EEG correlates of processing expected words in sentences.
608   *NeuroImage, 72*, 120-132.

609   Nieuwland, M., Barr, D., Bartolozzi, F., Busch-Moreno, S., Donaldson, D., Ferguson, H. J., et al.
610   (2019). Dissociable effects of prediction and integration during language comprehension:
611   Evidence from a large-scale study using brain potentials. *Proc. Roy. Soc. B: Biol. Sci.*

612   Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: open source software for
613   advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comp. Intell.*
614   *Neurosci., 2011*, 1.

615   Pascanu, R., Mikolov, T., & Bengio, Y. (2013). *On the difficulty of training recurrent neural*
616   *networks.* Paper presented at the Int. Conf. Mach. Learn.

617   Patten, W. (1910). *International Short Stories* (Vol. 2): P.F. Collier & Son.

618   Pennington, J., Socher, R., & Manning, C. (2014). *Glove: Global vectors for word representation.*
619   Paper presented at the Proc. Conf. Emp. Meth. Nat. Lang. Proc. (EMNLP).

620   Rommers, J., Dickson, D. S., Norton, J. J., Wlotko, E. W., & Federmeier, K. D. (2017). Alpha and
621   theta band dynamics related to sentential constraint and word expectancy. *Lang. Cogn.*
622   *Neurosci., 32*(5), 576-589.

623   Rösler, F., Pechmann, T., Streb, J., Röder, B., & Hennighausen, E. (1998). Parsing of sentences in a
624   language with varying word order: Word-by-word variations of processing demands are
625   revealed by event-related brain potentials. *J. Mem. Lang., 38*(2), 150-176.

626 Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic.

627   *Cognition, 128*(3), 302-319.

628 Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies.

629   *Brain Lang., 120*(2), 135-162.

630 Tse, C.-Y., Lee, C.-L., Sullivan, J., Garnsey, S. M., Dell, G. S., Fabiani, M., et al. (2007). Imaging

631   cortical dynamics of language processing with the event-related optical signal. *Proc. Natl.*

632   *Acad. Sci. U.S.A., 104*(43), 17157-17162.

633 Van Den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early

634   contextual influences during spoken-word recognition: N200 versus N400 effects. *J. Cogn.*

635   *Neurosci., 13*(7), 967-985.

636 Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in

637   electromagnetic and hemodynamic studies. *Brain Lang., 97*(3), 279-293.

638 Wang, L., Jensen, O., Van den Brink, D., Weder, N., Schoffelen, J. M., Magyari, L., et al. (2012).

639   Beta oscillations relate to the N400m during language comprehension. *Hum. Brain Map.,*

640   *33*(12), 2898-2912.

641 Wang, L., Zhu, Z., & Bastiaansen, M. (2012). Integration or predictability? A further specification of

642   the functional role of gamma oscillations in language comprehension. *Front. Psych., 3*, 187.

643 Weiss, S., & Mueller, H. M. (2012). "Too many betas do not spoil the broth": the role of beta brain

644   oscillations in language processing. *Front. Psychol., 3*, 201.

645 Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., et al.

646   (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail

647   party". *Neuron, 77*(5), 980-991.

648