

Cortical responses to natural speech reflect probabilistic phonotactics

Giovanni M. Di Liberto^{1,2}, Daniel Wong^{1,2}, Gerda Ana Melnik^{2,3}, Alain de Cheveigné^{1,2,4}

¹ *Laboratoire des Systèmes Perceptifs, UMR 8248, CNRS, France.*

² *Département d'Etudes Cognitives, Ecole Normale Supérieure, PSL University, France.*

³ *Laboratoire de Sciences Cognitives et Psycholinguistique, ENS, EHESS, CNRS, France.*

⁴ *UCL Ear Institute, London, United Kingdom.*

Correspondence:

Giovanni Di Liberto: Laboratoire des Systèmes Perceptifs, 29 rue d'Ulm, 75005 Paris,
diliberg@tcd.ie

Conflicts of interest: none declared.

Funding sources: this study was supported by the EU H2020-ICT grant 644732 (COCOHA)

Word count:

- Abstract: 210
- Significance Statement: 113
- Introduction: 643
- Discussion: 1288

1 **Abstract**

2 Humans comprehend speech despite the various challenges of real-world environments,
3 such as loud noise and mispronunciation. Our auditory system is robust to these thanks
4 to the integration of the upcoming sensory input with prior knowledge and expectations
5 built on language-specific regularities. One such regularity regards the permissible
6 phoneme sequences, which determine the likelihood that a word belongs to a given
7 language (phonotactic probability; “blick” is more likely to be an English word than
8 “bnick”). Previous research suggested that violations of these rules modulate brain
9 evoked responses such as the N400 and the late positive complex. Yet several
10 fundamental questions remain unresolved, especially regarding the neural encoding and
11 integration strategy of phonotactic information. Here, we used linear modelling
12 approaches to assess the influence of phonotactic probabilities on the brain responses to
13 narrative speech measured with non-invasive EEG. We found that the relationship
14 between continuous speech and EEG responses is best described when the speech
15 descriptor includes phonotactic probabilities. This provides us with a methodology to
16 isolate and measure the brain responses to phonotactics using natural speech at the
17 individual subject-level. Furthermore, such low-frequency signals showed the strongest
18 speech-EEG interactions at latencies of 100-400 ms, supporting a pre-lexical role of
19 phonotactic information.

20 **Significance Statement**

21 Speech is composed of basic units, called phonemes, whose combinations comply with
22 language-specific regularities determining whether a sequence “sounds” as a plausible
23 word. Our ability to detect irregular combinations requires matching incoming sequences
24 with our internal expectations, a process that supports speech segmentation and learning.
25 However, the neural mechanisms underlying this phenomenon have not yet been
26 established. Here, we examine this in the human brain using narrative speech. We
27 identified a brain signal reflecting the likelihood that a word belongs to the language,
28 which may offer new opportunities to investigate speech perception, learning,
29 development, and impairment. Our data also suggest a pre-lexical role of this
30 phenomenon, thus supporting and extending current mechanistic perspectives.

Keywords: phonotactics; phonology; natural speech; EEG; cortical tracking; language; predictions; neighbourhood density

32 **Introduction**

33 Speech can be described as a succession of categorical units called *phonemes* that comply
34 with language-specific regularities determining admissible combinations within a word.
35 A sequence is said *well formed* if it sounds plausible as a word to native speakers (e.g.
36 *blick*) and *ill formed* if it is perceived as extraneous to the language (e.g. *bnick*) (Chomsky
37 and Halle, 1968; Parker, 2012). This concept is referred to as *phonotactics*. Well-
38 formedness is gradient (Scholes, 1966; Chomsky and Halle, 1968; Frisch et al., 2000;
39 Bailey and Hahn, 2001a; Hammond, 2004), meaning that we can assign a numerical value
40 to each sequence of phonemes describing its likelihood of belonging to the language.
41 Phonotactics aids lexical access (Vitevitch et al., 1999) and speech segmentation (Brent
42 and Cartwright, 1996; Mattys et al., 1999) by constraining the space of likely upcoming
43 phonemes, thus contributing to the robustness of speech perception to challenges such as
44 noise, competing speakers, and mispronunciation (Davidson, 2006a; Obrig et al., 2016).
45 High phonotactic probability facilitates learning of new words (Storkel and Rogers, 2000;
46 Storkel, 2001, 2004; Storkel and Morrissette, 2002) and low phonotactic probability
47 (violation) may trigger an attempt to repair a sequence into a well-formed word (Dehaene-
48 Lambertz et al., 2000; Hallé et al., 2008; Carlson et al., 2016). However, considerable
49 uncertainty remains about the cortical mechanisms underpinning the contribution of
50 phonotactic information to speech comprehension (Winther Balling and Harald Baayen,
51 2008; Balling and Baayen, 2012; Ettinger et al., 2014). While part of the debate regards
52 the pre- or post-lexical role of phonotactics, there is currently a lack of neurobiological
53 data examining the cortical representation of phonotactic statistics. Hypotheses range
54 from the explicit encoding of phoneme-level probabilities to the use of the lexical
55 neighbourhood size as a proxy measure (McClelland and Elman, 1986; Bailey and Hahn,
56 2001b; Pisoni and Remez, 2005; Leonard et al., 2015).

57 One way to illuminate these issues is through the direct measurement of brain activity
58 using technologies with high-temporal resolution, such as electroencephalography
59 (EEG). Brain responses to phonotactics emerge by contrasting EEG responses to well-
60 and ill-formed speech tokens, i.e. phonotactic mismatch response (PMM; Connolly and
61 Phillips, 1994; Dehaene-Lambertz et al., 2000). This paradigm has been largely exploited
62 in the literature, with somewhat sparse and inconsistent results. EEG responses to these
63 violations emerge at latencies consistent with other well-known brain components, such
64 as the mismatch-negativity (MMN), N400, and late positive complex (LPC) (Dehaene-

65 Lambertz et al., 2000; Wiese et al., 2017). However, various types of confounds hamper
66 the identification of responses specific to phonotactics. One issue is that brain responses
67 to phonotactic probability may overlap with those reflecting subsequent processes, such
68 as *learning* in case of novel well-formed sequences (pseudowords) and *phonological*
69 *repair* for ill-formed tokens (non-words) (Bailey and Hahn, 2001a; White and Chiu,
70 2017). Secondly, if meaningful words are contrasted with ill-formed tokens, lexical-level
71 N400 responses may arise that confound the contrast (Kutas and Federmeier, 2011; Rossi
72 et al., 2011). The use of nonsense words avoids this issue, but the paradigm becomes
73 more artificial. Natural speech may allow to investigate the cortical processing of
74 phonotactics without such confounds; however it is generally characterised by well-
75 formed words, therefore measuring PMM responses may be either not possible or
76 suboptimal.

77 A novel approach to investigate the brain responses to natural speech may provide a
78 solution to these issues (Di Liberto et al., 2015, 2018a; Crosse et al., 2016b; Broderick et
79 al., 2018; de Cheveigné et al., 2018b). This method, based on linear modelling, allows to
80 isolate and measure cortical responses to linguistic features of interest (e.g. phonemes)
81 using natural speech stimuli. Here, we combine this approach with a computational model
82 of phonotactics to test whether narrative speech elicits robust brain responses time-locked
83 to patterns of phonotactic probabilities. We characterise the dynamics of cortical signals
84 that are representative of real-life speech perception, contributing to the debate on the
85 underpinnings of the cortical processes specific to phonotactics.

86 **Material and methods**

87 The present study is based on new analyses of a previously published EEG dataset on
88 natural speech perception (Di Liberto et al., 2015). The data include both the audio
89 stimulus and the EEG response of the subjects listening to that stimulus. Data analysis
90 involves fitting the EEG to various representations of the stimulus using a linear model.
91 The quality of fit is used as an indicator of the relevance of each representation as a
92 predictor of the cortical activity evoked in the listener by the speech stimulus.

93 **Subjects and Experimental Procedure**

94 Ten healthy subjects (7 male) aged between 23 and 38 years old participated in the
95 experiment. Participants reported no history of hearing impairment or neurological
96 disorder. The experiment was carried out in a single session for each subject.

97 Electroencephalographic (EEG) data were recorded from participants as they undertook
98 28 trials, each of ~155 seconds in length, where they were presented with an audiobook
99 version of a classic work of fiction read by a male American English speaker. The trials
100 preserved the storyline, with neither repetitions nor discontinuities. All stimuli were
101 presented monophonically at a sampling rate of 44,100 Hz using Sennheiser HD650
102 headphones and Presentation software from Neurobehavioral Systems
103 (<http://www.neurobs.com>). Testing was carried out in a dark room and subjects were
104 instructed to maintain visual fixation for the duration of each trial on a crosshair centered
105 on the screen, and to minimize eye blinking and all other motor activities. All procedures
106 were undertaken in accordance with the Declaration of Helsinki and were approved by
107 the Ethics Committees of the School of Psychology at Trinity College Dublin, and the
108 Health Sciences Faculty at Trinity College Dublin. Further details about the stimulus and
109 recording are available in Di Liberto et al., (2015) and the data is available at
110 <https://datadryad.org/resource/doi:10.5061/dryad.070jc>.

111 **Speech representations**

112 The approach used here follows a system identification framework that aims at
113 disentangling brain responses to different speech and language features (Di Liberto et al.,
114 2015). To this end, we first need to define such features (note that the first two elements
115 are as in Di Liberto et al., 2015):

- 116 1. Acoustic spectrogram (**S**): This was obtained by filtering the speech stimulus into
117 16 frequency-bands between 250 Hz and 8 kHz distributed according to
118 Greenwood's equation (equal distance on the basilar membrane; Greenwood,
119 1961) using Chebyshev type 2 filters (order 100), and then computing the Hilbert
120 amplitude envelope (the absolute value of the analytical signal obtained by the
121 Hilbert Transform) for each frequency band.
- 122 2. Phonetic features (**F**): This multivariate representation of speech encodes
123 phoneme-level information using phonetic features. The Prosodylab-Aligner
124 software (Gorman et al., 2011) was used to partition each word into phonemes
125 from the American English International Phonetic Alphabet (IPA) and align the
126 speech stimulus with its textual transcription. This procedure returns estimates of
127 the starting and ending time-points for each phoneme. Indicator functions for each
128 of the 35 phonemes were recoded as a multivariate time series of 19 indicator

129 variables, one for each of 19 phonetic features (based on the University of Iowa's
130 phonetics project <http://soundsofspeech.uiowa.edu/>) coding the manner of
131 articulation (plosive, fricative, affricate, nasal, liquid, and glide), place of
132 articulation (bilabial, labio-dental, lingua-dental, lingua-alveolar, lingua-palatal,
133 lingua-velar, and glottal), voicing of a consonant (voiced and voiceless), and
134 backness of a vowel (front, central, and back). Also, a specific feature was
135 reserved for diphthongs. Each indicator variable took the value 1 between the start
136 and the end of the phoneme (if relevant) and 0 elsewhere. Each phoneme was
137 characterised by a value of 1 for some combination of indicator variables; not all
138 such combinations map to permissible phonemes.

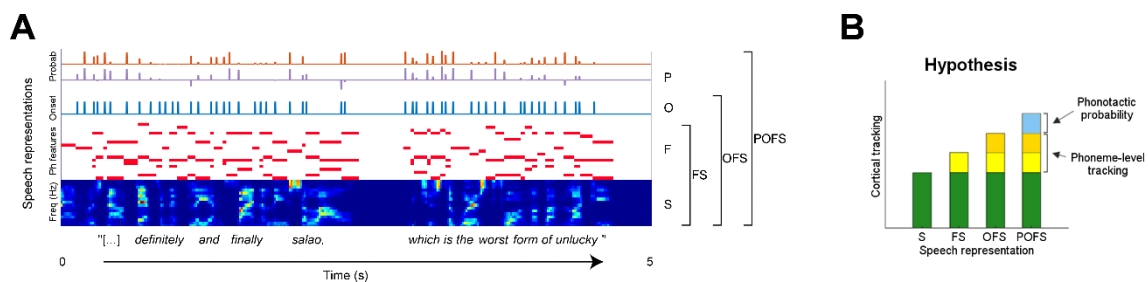
139 3. Phoneme onsets (**O**): This vector marks phoneme onsets with a discrete-time unit
140 impulse, corresponding to the half-wave rectified first derivative of F. This is a
141 non-linear transformation of the F features, thus linear models may benefit from
142 the explicit definition of O combined with F.

143 4. Finally, we propose a novel representation using *phonotactic probabilities* (**P**).
144 Natural languages include various constraints on the permissible phoneme
145 sequences. Probabilities can be derived for a given speech token from this set of
146 constraints. For example, the pseudoword *blick* would “sound” better than *bnick*
147 to a native English speaker, which is reflected by a higher phonotactic probability
148 for the first word. Here, we used a computational model (BLICK; Hayes and
149 Wilson, 2008) based on a combination of explicit theoretical rules from traditional
150 phonology and a maxent grammar (Goldwater and Johnson, 2003), which find the
151 optimal weights for such theoretical constraints to best match the phonotactic
152 intuition of a native speaker. Specifically, given a phoneme sequence $ph_{1..n}$, P is
153 composed of two vectors: a) inverse phonotactic probability ($score(ph_{1..n})$ is the
154 output of the BLICK software; it is small for well-formed tokens and large for ill-
155 formed ones) and b) within-word derivative of the phonotactic probability ($score(ph_{1..(n-1)}) - score(ph_{1..n})$), which describes the contribution of the latest
156 phoneme to the well-formedness of the sequence.
157

158 In order to assess and quantify the contribution of each of the features F, O, and P to the
159 speech-EEG mapping, the main analyses were conducted on the cumulative combinations
160 S, FS, OFS, and POFS. The rationale is that, if the new feature carries information not
161 subsumed by the other features, including it will improve the fitting score. To control for

162 any potential effect of the difference in dimensionality of the feature space, we also used
 163 variants where the newly introduced feature did not correspond to the auditory stimulus,
 164 i.e. was shuffled (the entire procedure, including model fit, was rerun for each shuffled
 165 version). These mismatched vectors/matrices were generated by randomly shuffling: a)
 166 Phonetic features in the FS speech representation ($F_{shu}S$) (every given phoneme,
 167 corresponding to a combination of N_F phonetic features, 1 for vowels and 3 for
 168 consonants, was replaced by N_F random phonetic features for its entire duration); b) Onset
 169 time in OFS ($O_{shu}FS$) (the onset vector O was replaced by vector with the same number
 170 of impulses at random time points); and c) Phonotactic probability values in POFS
 171 ($P_{shu}OFS$) (the values in the phonotactic vector P were randomly permuted while keeping
 172 the time information).

173 In addition to the phonotactic vector P , we defined three other representations that could
 174 reflect the encoding of phonotactic information in the brain. First, P_{neigh} is a vector of
 175 phoneme onsets amplitude-modulated using *neighborhood density* values. This
 176 information indicates the number of phonological neighbours given a speech token, where
 177 a phonological “neighbour” is a sequence of phonemes that can be obtained from the
 178 given token by deletion, addition, or substitution of a single phoneme. Similarly, P_{sur} and
 179 P_{ent} are vectors of phoneme onsets that are amplitude-modulated using phoneme *surprisal*
 180 and *entropy* respectively. These were calculated using the purely probabilistic measures
 181 “phoneme surprisal” and “cohort entropy” as defined by Gaston and Marantz (2018).



182 **Figure 1. (A) Speech representations** for a 5 seconds portion of the stimulus. From bottom to top, the
 183 acoustic spectrogram (S) which consists of a 16-channel time series of power within 16 frequency bands;
 184 phonetic features (F), whose permissible combinations map to English phonemes; phoneme onsets (O),
 185 which mark the beginning of each phoneme; and the probabilistic phonotactic vector (P), a representation
 186 indicating the inverse likelihood of a sequence (from the beginning of a word to each of its phonemes). **(B)**
 187 **Expected outcomes:** We hypothesise that, if a stimulus representation encodes features not captured by
 188 other representations, adding it to the others will improve the prediction of cortical responses. In particular
 189 we predict an increase in cortical tracking due when phonotactic probabilities are added to the mix (POFS
 190 – OFS, blue increment).
 191

192

193 *Phonotactic Probability Model*

194 Phonotactic probability vectors were derived using the BLICK algorithm (Hayes and
195 Wilson, 2008), a state-of-the-art tool based on explicit theories of phonology.
196 Specifically, the BLICK algorithm constructs maxent grammars (e.g. Goldwater and
197 Johnson, 2003) consisting of a set of numerically weighted phonological constraints. A
198 training stage identifies weights that optimally match the phonotactic well-formedness
199 intuition of experts. These weights are determined according to the principle of maximum
200 entropy and, in the present work, were pre-assigned using an English grammar model
201 (Hayes, 2012). Combining this pre-trained grammar with the textual transcription of the
202 audio-book stimulus, BLICK performs a weighted sum of its constraint violations to
203 calculate probability values reflecting the well-formedness of each speech token. Given
204 a word, two scores were calculated for each phoneme token. The first indicates the inverse
205 probability of the word segment up to that phoneme (e.g. the scores for /b/, /b l/, /b l ɪ /,
206 and /b l ɪ k/ were calculated in correspondence of the four phonemes of the word ‘blick’).
207 This time series of inverse probabilities was coded by the amplitudes of a series of pulses
208 synchronous with those of the onset vector. The second is the finite difference of
209 consecutive inverse probability values within a word (starting from the second phoneme
210 of each word, e.g. $P(/b/)-P(/b l/)$, $P(/b l/)-P(/b l ɪ /)$, $P(/b l ɪ /)-P(/b l ɪ k/)$; the score for
211 the first phoneme of a word was assigned to the same value as in the phonotactic
212 probability vector). The time series of difference measures was also coded as a time series
213 of pulses synchronous with O. The concatenation of these two pulse trains constitutes the
214 2-dimensional phonotactic probability vector P.

215 **Data Acquisition and Preprocessing**

216 Electroencephalographic (EEG) data were recorded from 128 scalp electrodes (plus 2
217 mastoid channels), filtered over the range 0 - 134 Hz, and digitised with a sampling
218 frequency of 512 Hz using a BioSemi Active Two system. Data were analysed offline
219 using MATLAB software (The Mathworks Inc.). EEG data were digitally filtered
220 between 0.5 and 32 Hz using a Butterworth zero-phase filter (low- and high-pass filters
221 both with order 2; implemented with the function *filtfilt*), and down-sampled to 64 Hz.
222 EEG channels with a variance exceeding three times that of the surrounding channels
223 were replaced by an estimate calculated using spherical spline interpolation (EEGLAB;
224 Delorme and Makeig, 2004). All channels were then re-referenced to the average of the

225 two mastoid channels with the goal of maximizing the EEG responses to the auditory
226 stimuli (Luck, 2005).

227 *Dimensionality reduction*

228 The analyses that follow involve fitting the stimulus representation to the EEG response
229 using a linear model. Both the stimulus and the EEG include a large number of
230 dimensions (channels) many of which are correlated. To limit the risk of overfitting, it is
231 useful to reduce their dimensionality. This is typically performed using principal
232 component analysis (PCA). PCA finds a matrix of size $N \times N$ (if the data have N channels)
233 that transforms the data to N ‘principal components’ (PC). The variance of the PCs sum
234 up to the variance of the data. Subject to that constraint, the first principal component is
235 the linear transform of the data with the largest possible variance. The second has the
236 largest variance of transforms orthogonal to the first and so on. The first few PCs pack
237 most of the variance, and so little variance is lost if a subset of $N_{PC} < N$ PCs are selected
238 and the remainder discarded. This procedure is applied repeatedly in the following
239 analyses. In each case N_{PC} is tuned as a hyperparameter in a crossvalidation procedure to
240 optimise the tradeoff between information retained and overfitting.

241 *Denoising with multiway CCA*

242 Our goal of evaluating the relevance of high-level speech structure representations by
243 measuring their ability to predict cortical responses is hampered by the high level of noise
244 and artifact in the EEG. We use a novel tool, multiway canonical correlation analysis
245 (MCCA) to merge EEG data across subjects so as to factor out the noise. MCCA is an
246 extension of canonical correlation analysis (CCA; Hotelling, 1936; de Cheveigné et al.,
247 2018a) to the case of multiple (> 2) datasets. Given N multichannel datasets X_i with size
248 $T \times J_i$, $1 \leq i \leq N$ (time x channels), MCCA finds a linear transform W_i (sizes $J_i \times J_0$, where
249 $J_0 < \min(J_i)_{1 \leq i \leq N}$) that, when applied to the corresponding data matrices, aligns them to
250 common coordinates and reveals shared patterns (de Cheveigné et al., 2018a). These
251 patterns can be derived by summing the transformed data matrices: $Y = \sum_{i=1}^N X_i W_i$. The
252 columns of the matrix Y , which are mutually orthogonal, are referred to as summary
253 components (SC) (de Cheveigné et al., 2018a). Intuitively, the first few components are
254 signals that most strongly reflect the shared information across the several input datasets.
255 Here, these datasets are EEG responses to a same speech stimulus for 10 subjects.

256 This technique allows to extract a *consensus* signal that is shared across participants. The
257 present study utilises this approach to test whether EEG responses to speech reflect
258 phonotactic information. This methodology overcomes limitations of previous studies
259 that attempted to obtain similar consensus responses by averaging data across subjects,
260 which could not perform coregistration because of the lack of anatomical information
261 and, therefore, ignored the likely topographical discrepancies between participants EEG
262 signals (O’Sullivan et al., 2014; Di Liberto and Lalor, 2017). MCCA accounts for such
263 discrepancies without the need for coregistration. Under the assumption that brain
264 responses to speech share some fundamental similarities within a homogeneous group of
265 normal hearing young adults, the MCCA procedure allows us to extract such common
266 responses to the stimulus from other, more variable aspects of the EEG signals, such as
267 subject-specific noise. For this reason, our analysis focuses on the first N_{SC} summary
268 components, which we can consider as reflecting a ground truth EEG response to speech.
269 N_{SC} was arbitrarily set to the number of dimensions for a single subject after
270 dimensionality reduction (N_{PC} ; see the following section). This conservative choice was
271 made by taking into consideration that the irrelevant signals within the retained
272 components are excluded through the more restrictive CCA analysis that follows.

273 **Analysis Procedure**

274 *Stimulus-response model based on Canonical Correlation Analysis*

275 Speech elicits brain responses that can be recorded with EEG. However, a large part of
276 the EEG signal is unrelated to the stimulus as it may reflect other brain processes , as well
277 as various forms of noise (e.g. muscle movements). Similarly, certain features of the
278 speech input may have little or no impact on the measured brain responses. Studying the
279 relation between speech and the corresponding EEG responses would greatly benefit from
280 the ability to remove those unrelated portions of speech and EEG. This can be done by
281 using canonical correlation analysis (CCA), a powerful technique that linearly transforms
282 both stimulus and brain measurements so as to minimise irrelevant variance (Hotelling,
283 1936; de Cheveigné et al., 2018b) .

284 In its more general definition, given two sets of multichannel data X_1 and X_2 of size $T \times$
285 J_1 and $T \times J_2$, CCA finds linear transformations of both that make them maximally
286 correlated. Specifically, CCA produces the transformation matrices W_1 and W_2 (sizes J_1
287 $\times J_0$ and $J_2 \times J_0$, where $J_0 < \min(J_1, J_2)$) that maximise the correlation between pairs of

288 columns of X_1W_1 and X_2W_2 , while making the columns of each transformed data matrix
289 X_iW_i mutually uncorrelated. The first pair of canonical components (CC) is the linear
290 combination of X_1 and X_2 with highest possible correlation. The next pair of CCs are the
291 most highly correlated combinations orthogonal to the first, and so-on.

292 In the present study, X_1 and X_2 represent speech features and the EEG signal respectively.
293 This basic formulation of CCA can be used directly to study the instantaneous interaction
294 between stimulus features and brain response. However, a stimulus at time t affects the
295 brain signals for a certain length of time (a few hundreds of milliseconds). Although CCA
296 is a linear approach, simple manipulations of the data allow for its extension to the study
297 of non-linear and convolutional relations, therefore capturing the stimulus-to-brain
298 interaction for a given set of latencies (or time-lags). Here, this is achieved by using a set
299 of filters that capture increasingly long temporal structures (de Cheveigné et al., 2018b).
300 Specifically, we used a dyadic bank of FIR bandpass filters with characteristics (center
301 frequency, bandwidth, duration of impulse response) approximately uniformly
302 distributed on a logarithmic scale. There was a total of 15 channels (N_{CH}) with impulse
303 response durations ranging from 2 to 128 samples (2 s). The filterbank was applied to
304 both stimulus and EEG matrices, largely increasing the dimensionality of the data.

305 Dimensionality reduction was applied to both stimulus and EEG matrices (of size $T \times N_F$
306 and $T \times N_{EL}$, where T , N_F , and N_{EL} indicate numbers of time-samples, stimulus features,
307 and EEG electrodes respectively). First, we used PCA and retained $N_{PC} < N_{EL}$ principal
308 components to spatially whiten the EEG data, whose neighbouring channels are largely
309 correlated. The value of this parameter was adjusted using a grid search procedure.
310 Second, the filterbank was applied to both stimulus and EEG data. Finally, PCA was used
311 to reduce the dimensionality of both stimulus and EEG matrices, by retaining $N_{stim} <$
312 $N_F * N_{CH}$ and $N_{EEG} < N_{PC} * N_{CH}$ components respectively. The CCA models were all trained
313 and tested using a leave-one-out nested cross-validation to control for overfitting. For
314 each outer cross-validation loop, one fold was held-out for testing while a second cross-
315 validation loop was run on the remaining data. In this inner loop, the model
316 hyperparameters were tuned on a held-out validation fold to maximise the sum of the
317 correlation coefficients for the CC-pairs. This framework allowed for the tuning of the
318 values N_{stim} and N_{EEG} . In addition, the validation folds at each cross-validation step were
319 used to determine the optimal *shift* between stimulus and neural signals.

320 *Temporal Response Function Analysis*

321 Complementary to CCA, a system identification technique was used to compute a
322 channel-specific mapping between each speech representation and the recorded EEG
323 data. This method, commonly referred to as the temporal response function (TRF)
324 analysis (forward model) (Lalor et al., 2006; Ding and Simon, 2012), estimates a filter
325 that optimally describes how the brain transforms the speech features of interest $S(t)$ into
326 the corresponding continuous neural responses $R(t)$, over a series of pre-specified time-
327 lags: $R(t) = \text{TRF} * S(t)$, where ‘*’ indicated the convolution operator. The TRF values, or
328 weights, were estimated using a regularised linear regression approach, wherein a
329 regularisation parameter was tuned to control for overfitting (Crosse et al., 2016a). One
330 way to use this approach is to study the model weights to identify scalp areas and time-
331 lags that are of particular importance for the specific speech-EEG mapping. A second
332 approach consists of predicting the EEG signals at each channel of interest.

333 This approach is complementary with CCA analysis in that it provides us with detailed
334 insights on the temporal and spatial patterns. This is possible at the cost of additional
335 constraints, specifically on the frequency-bands of interest and on the magnitude of the
336 prediction correlation values which, since they are calculate in the noisy EEG channel-
337 space (rather than the denoised CCA-space), are usually in the order of 0.05. For this
338 reason, it is preferable to conduct the analysis on the most relevant part of the EEG
339 signals, which can be achieved with a more confined temporal filtering (for an example
340 of the effect of EEG filtering on forward TRF models see Di Liberto et al., 2015). In
341 particular, we restricted the analysis to the frequency-band 0.5-9 Hz (we applied separate
342 low- and high-pass fifth-order Butterworth zero-phase filters).

343 *Measuring the quality of the speech-EEG mapping*

344 We used two metrics to quantify the quality of the CCA-based speech-EEG mapping
345 model: correlation and discriminability in a match-vs-mismatch classification task. A
346 Pearson’s correlation coefficient was calculated for each CC-pair. The first CC-pair is the
347 most relevant, but meaningful speech-EEG correlations can arise for an arbitrary number
348 of components. To obtain a measure sensitive to these multiple dimensions, we
349 introduced a match-vs-mismatch classification task that consisted in deciding whether a
350 segment of EEG (duration T_{DECODER}) was produced by the segment of speech that gave
351 rise to it, or by some other segment. Discriminability in this task, measured by *d-prime*,

352 reflects the ability of the model to capture the relation between speech and EEG. The *d-*
353 *prime* metric was derived from the discriminant function of a support vector machine
354 (SVM) classifier trained on the normalised Euclidean distance between pairs of CCs. A
355 cross-validation procedure ($k = 30$) was used in which the classifier was trained and
356 evaluated on distinct data to discriminate between match and mismatch segments.
357 T_{DECODER} was set to the value 1 second, which avoided saturation (classification either
358 too easy or too difficult) in both group and single-subject level analyses.

359 The quality of the TRF-based speech-EEG mapping was assessed using a correlation
360 metric. Specifically, Pearson's correlation coefficients were calculated between the EEG
361 signal and its prediction for each scalp electrode separately. This procedure was repeated
362 for TRF models fit using various time-latency windows and stimulus feature-sets, which
363 allowed the pinpointing of latencies that were most relevant to the interaction between
364 EEG and particular speech features of interest (the time-lag windows were within the
365 interval 0 – 900 ms, non-overlapping, and of duration 100 ms). A similar analysis was
366 conducted to investigate topographical patterns corresponding to the various speech-EEG
367 latencies. Specifically, TRF models were fit using a single time-latency window between
368 0 and 900 ms. Topographical patterns of the corresponding TRF weights were averaged
369 for intervals of interest.

370 *Statistical Analyses*

371 Unless otherwise stated, all statistical analyses were performed using two-tailed
372 permutation tests. For tests involving several contiguous time latencies, a cluster-mass
373 non-parametric analysis was conducted, with one as the minimum cluster size (Maris and
374 Oostenveld, 2007). This statistical test takes into consideration the scalp distribution of
375 the measure of interest by performing a permutation test on the cluster of electrodes with
376 the highest score, i.e., the most important cluster according to the metric of interest. This
377 approach provides a solution to the multiple comparison problem by including
378 biophysically-motivated constraints that increase the sensitivity of this statistical test in
379 comparison with a standard Bonferroni correction.

380

381

382 **Results**

383 Non-invasive EEG signals were recorded from ten participants as they listened to an
384 audiobook. We conducted three analyses tackling the questions: 1) Do cortical signals
385 track the small changes in phonotactic probability that characterise natural speech? 2)
386 Can we measure these phonotactic responses at the individual-subject level? And 3) do
387 these signals reflect a pre-lexical influence of phonotactics in speech comprehension?
388

389 **Neural Evidence for the Processing of Probabilistic Phonotactics**

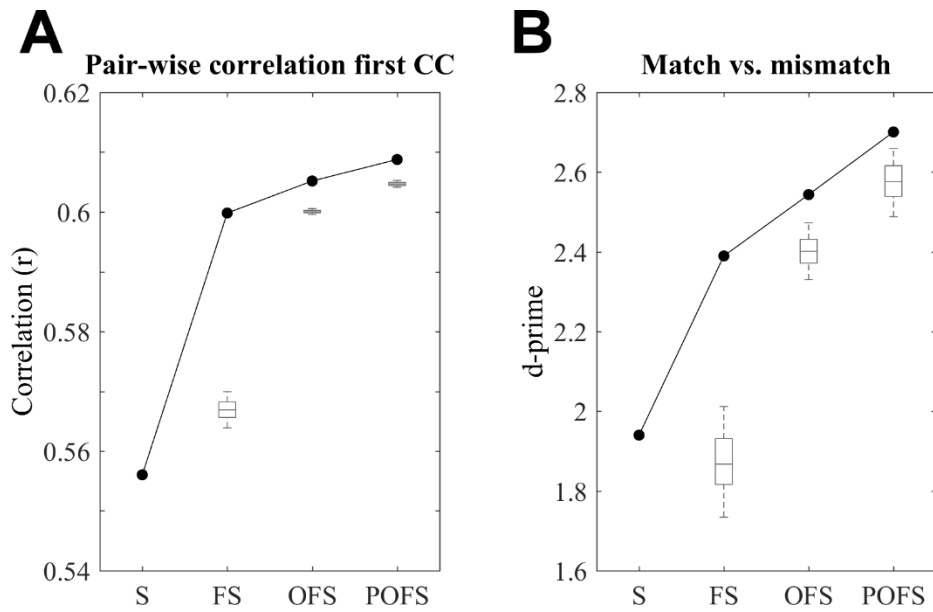
390 Brain signals that are common among participants listening to the same speech stimulus
391 were estimated using MCCA (de Cheveigné et al., 2018a). This *consensus signal* (CS)
392 can be thought of as a ground truth cortical response with better signal-to-noise ratio than
393 EEG data of individual subjects. A speech-EEG model based on CCA was then employed
394 to related this consensus EEG signal to different speech feature sets. The quality of the
395 model (measured by correlation and *d-prime* metrics) was used as a measure of the ability
396 of each feature set to capture speech structure predictive of the EEG response.

397 We wish specifically to evaluate the predictive power of the phonotactic feature set P
398 relative to, and in combination with, other known feature sets such as spectrogram of
399 phonetic features.

400 We first estimated the quality of a CCA-based model involving only the phonotactic
401 feature vector (P; **Figure 1A,top**) and EEG. The r -value of 0.42 obtained for the first CC-
402 pair was larger than the 99th percentile of a distribution obtained by shuffling the values
403 of the pulses within the P vector while leaving their times intact (median over 100
404 shuffles: $r = 0.34$; 99th percentile: $r = 0.35$). This result indicates that phonotactic
405 probabilities were reflected by the EEG signals. However the phonotactic feature vector
406 is correlated with other predictive features (such as spectrogram or phonemes), so we
407 cannot be sure that its predictive power stems from phonotactic information per se. For
408 that, we must compare combinations of features that include, or not, the phonotactic
409 vector P. We formed combinations of features including the acoustic spectrogram S (Di
410 Liberto et al., 2015; Lalor et al., 2009; Obleser et al., 2012), a phoneme representation
411 based on phonetic features F (Mesgarani et al., 2014; Di Liberto et al., 2015, 2018a),
412 phoneme onsets O (Brodbeck et al., 2018) and our newly introduced phonotactic features
413 P (see **Figure 1A**). If each of these features carries information complementary to the
414 others, and not captured by them, we expect speech-EEG correlations to monotonically

415 increase with the inclusion of additional features in the analysis: namely S, FS, OFS, and
416 POFS as schematized in **Figure 1B**. Indeed, correlation coefficient values for CCA
417 models based on these four combination of features agree with this prediction (**Figure**
418 **2A**; $r_S < r_{FS} < r_{OFS} < r_{POFS}$). Of possible concern is that these models differ in the number
419 of dimensions (and thus parameters) involved. A large number of parameters can lead to
420 overfitting, which should penalise the models with more features, contrary to what we
421 observe. To further exclude such a possibility, we randomly shuffled the values of the
422 pulses within the phonotactic feature vectors while keeping their timing constant. The
423 distribution of correlation scores for $P_{\text{shu}}\text{OFS}$ obtained by repeated shuffling is indicated
424 in **Figure 2A**. The value obtained for POFS is above the 99th percentile of that
425 distribution. This same control procedure was applied to the F and O features and
426 confirmed that their respective enhancements are driven by the addition of meaningful
427 features, and not by differences in dimensionality, as they produced stronger correlations
428 than the 99th percentile of the corresponding shuffled distributions. In summary, each of
429 these features carries useful information not carried by the others.

430 The previous analysis was based on correlations for the first CC-pair only, but other
431 components may carry relevant information as well. To get a more complete picture we
432 performed a similar analysis based on the *d-prime* measure for a match-vs-mismatch trial
433 classification, which combines all components simultaneously (see Methods). The *d-*
434 *prime* values showed patterns resembling what previously seen for the correlation
435 analysis. Specifically, a *d-prime* of 0.704 resulted from the CCA analysis on P, which
436 was greater than the 99th percentile of the shuffled distribution (median over 100 shuffles:
437 *d-prime* = 0.504; 99th percentile: *d-prime* = 0.544). Furthermore, *d-prime* values
438 monotonically increased for S, FS, OFS, and POFS, showing again greater values than
439 the corresponding shuffle distributions (**Figure 2B**). The greater value for POFS relative
440 to OFS and $P_{\text{shu}}\text{OFS}$ reinforces our claim that cortical signals track phonotactic
441 probabilities.



442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

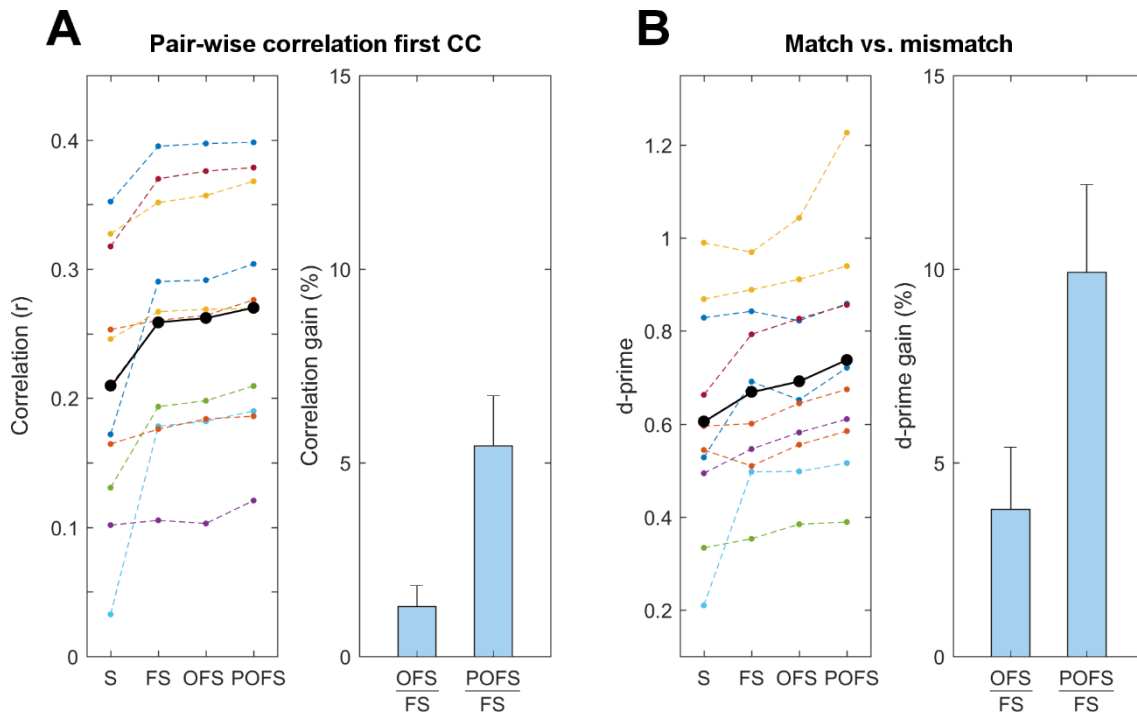
466

Figure 2: EEG responses to natural speech are best explained when including phonotactic probability among the speech features. Data from all participants were combined using MCCA. This consensus EEG signal (CS) preserves signals that are maximally correlated across subjects. **(A)** A CCA analysis was conducted between each speech representation and the CS signals. Speech-EEG correlations for the first canonical component (CC) pair were best when using the combined model POFS, indicating that phonotactic probabilities explain EEG variance that was not captured by the purely acoustic-phonemic models (S, FS, and OFS). **(B)** In addition, phonotactic probabilities enhanced the *d-prime* score of a match-vs-mismatch classification test. The box-plots indicate the 99th percentile of the performance when using a combined model (FS, OFS, or POFS) after randomly shuffling information for the newly added feature (F, O, and P respectively).

Robust Individual-subject EEG Tracking of Phonotactic Probabilities

The previous analysis provided evidence that the cortical responses to natural speech, measured with non-invasive EEG, are influenced by phonotactic probabilities. To test whether such responses can be reliably measured at the individual-subject level, we conducted the same CCA analysis as in the previous section on the brain recordings from each individual. **Figure 3** (left panels of A and B) illustrates both correlation and *d-prime* results. The scores are overall smaller than for the analysis based on the consensus signal, reflecting the greater amount of noise in the subject-specific data, but the same trends are observed. POFS is the best performing model in terms of both correlation (POFS > OFS, $p = 0.0008$; $d = 1.35$; POFS > FS, $p < 0.0001$; $d = 1.89$; POFS > S, $p < 0.0001$; $d = 1.20$) and *d-prime* (POFS > OFS, $p = 0.027$; $d = 0.60$; POFS > FS, $p = 0.0012$; $d = 0.86$; POFS > S, $p = 0.0049$; $d = 0.98$). In addition, this analysis confirmed that phonetic features explain EEG variance not captured by the acoustic spectrogram (FS > S; correlations: p

467 = 0.0045, $d = 1.17$; d -prime: $p < 0.014$, $d = 0.85$) and, similarly, that the phoneme onsets
 468 vector refines the FS representation of speech (OFS > FS; correlations: correlations: $p <$
 469 0.0001, $d = 1.03$; d -prime: $p = 0.042$, $d = 0.65$). The average benefits (relative gain) of
 470 adding the onset vector O, and the phonotactic vector P, for both measures is plotted in
 471 the right-hand panels of **Figure 3A** and **B**. Statistical analysis on these average measures
 472 confirms that phonotactic information has a measurable effect on the EEG responses to
 473 speech (correlation: $p = 0.0008$, $d = 1.03$; d -prime: $p = 0.016$, $d = 0.65$).
 474 Finally, we conducted additional analyses to test whether other models of phonotactic
 475 information can explain EEG responses as well, or better, than P. A first single-subject
 476 CCA-based analysis compared P to *neighbourhood density* (P_{neigh}). This feature was
 477 suggested as a possible neural strategy for an indirect encoding of phonotactic
 478 information (Vitevitch et al., 1999; Bailey and Hahn, 2001a). P performed better than this
 479 new measure in terms of d -prime (POFS > P_{neigh} OFS; one-tailed permutation test: $p =$
 480 0.0179; $d = 0.68$). We performed a similar comparison between P and probabilistic
 481 definitions of phoneme surprisal (P_{sur}) and entropy (P_{ent}) (Brodbeck et al., 2018; Gaston
 482 and Marantz, 2018). Again, P performed better than these two measures. Specifically, P
 483 showed larger d -prime values than P_{ent} (POFS > P_{ent} OFS; one-tailed permutation test: p
 484 = 0.037; $d = 0.67$) and P_{sur} (POFS > P_{sur} OFS; one-tailed permutation test: $p = 0.02$; $d =$
 485 0.73).



487 **Figure 3: Phonotactic probabilities enhance the speech-EEG mapping at the individual subject level.**

488 CCA analyses were conducted between each speech representation and the corresponding EEG responses
489 for each individual subject. **(A)** Speech-EEG correlations for the first canonical component pair were
490 greatest when using the combined model POFS (left panel). The thick black line indicates the average
491 across subjects while the coloured dots/lines refer to the individual subjects. The bar-plot shows the relative
492 correlation gain (%) of the combined models OFS and POFS with FS (i.e. the contribution given by O and
493 P respectively). **(B)** Similar results are shown for the *d-prime* scores of a match-vs-mismatch classification
494 test. Results for individual subjects are colour-coded (same colors as for A). Phonotactic probabilities
495 enhance the single-subject scores for FS and also show significant improvement compared to OFS.

496

497 **Timescale of Cortical Responses to Phonotactics**

498 Our results suggests that phonotactic probabilities influence the cortical processing of
499 natural speech. We conducted further analyses to assess the temporal dynamics of this
500 effect. Linear forward models were fit using the TRF approach to describe how speech
501 features are transformed into EEG signals. Because of the sensitivity of the forward TRF
502 method to EEG noise, we restricted the analysis to the frequencies 0.5-9 Hz, which are
503 most relevant for the EEG tracking of speech acoustic and phoneme-level features (Di
504 Liberto et al., 2015, 2018b; Kösem and van Wassenhove, 2016; Vanthornhout et al.,
505 2018).

506 Forward encoding models were fit for each speech representation (S, FS, OFS, POFS)
507 using non-overlapping time-lag windows of duration 100 ms within the interval 0 – 900
508 ms. Average EEG prediction correlations confirm the hypothesised general trend that
509 emerged also from the CCA analysis ($S < FS < OFS < POFS$; **Figure 4-1**). Crucially, the
510 direct comparison of POFS and OFS reveals a significant effect of phonotactics for a
511 cluster of speech-EEG latencies between 100 and 400 ms (cluster statistics, $p < 0.05$),
512 with peak effect-size at the latency-window 300 – 400 ms ($d = 1.53$) (**Figure 4**).

513

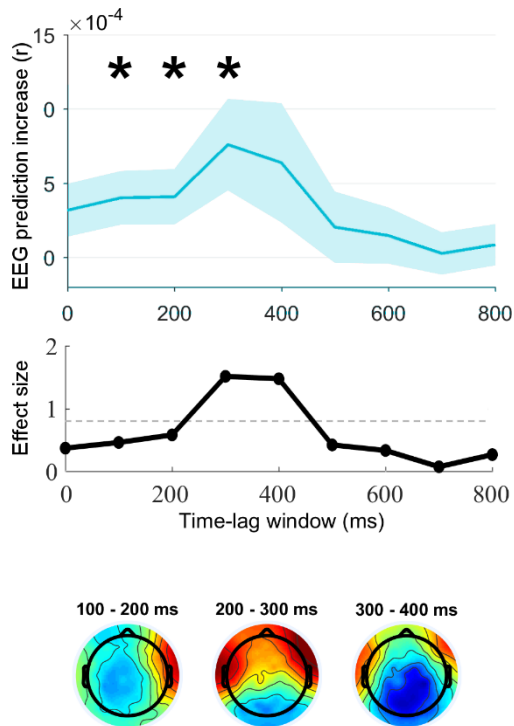


Figure 4: EEG tracking of phonotactic probabilities is specific to speech-brain latencies of 100-400 ms. A temporal response function (TRF) analysis was conducted to estimate the amount of EEG variance explained by phonotactic probabilities for speech-EEG latency windows between 0 and 900 ms and window-size 100 ms. EEG prediction correlations were calculated for different speech feature-sets and for the various speech-EEG latencies. The enhancement in EEG predictions due to phonotactic probabilities is shown for all time-latency windows. Shaded areas indicate the standard error of the mean (SE) across subjects. Stars indicate significant enhancement ($*p < 0.05$) as a result of a cluster mass statistics (top). Cohen's d was calculated to measure the effect size of the enhancement due to phonotactics. Values above 0.8 are considered as 'large' effects (above dashed grey line) (centre). Topographical patterns of the TRF weights for a model fit over

530 time-lags from 0 to 900 ms are shown for latencies with a significant effect of phonotactic probabilities
531 (100-400 ms) (bottom).

532

533 Discussion

534 Our results demonstrate that cortical responses to natural speech reflect probabilistic
535 phonotactics. First, linear modelling revealed a time-locked interaction between
536 phonotactic information and low-frequency EEG. Then, we established that brain
537 responses to phonotactics can be measured at the individual subject-level. Finally, we
538 found that speech-EEG latencies of 100-400 ms are most relevant to those brain
539 responses, suggesting that phonotactic information contributes to natural speech
540 processing at pre-lexical stages.

541

542 A novel measure of phonotactic processing

543 Phonotactic information plays an important role in speech perception. However, crucial
544 questions remain unanswered about the underpinnings of the corresponding cortical
545 processes, mainly due to a lack of tools to extract direct measures of brain responses to
546 phonotactics. Although neurophysiology has partially fulfilled this need (Connolly and
547 Phillips, 1994; Dehaene-Lambertz et al., 2000; Wagner et al., 2012; Cibelli et al., 2015;
548 Leonard et al., 2015), its findings were mainly confined to nonsense words or to the
549 domain of phonotactic violations, which are exceptions in natural speech scenarios. The

550 present study aimed to measure brain signals corresponding to the continuous integration
551 of phonotactic information, which are difficult to isolate when measuring only
552 phonotactic violations. These violations trigger various other processes such as
553 phonological repair, which may emerge in the evoked-response (Dehaene-Lambertz et
554 al., 2000; Dupoux and Pallier, 2001; Domahs et al., 2009). Here, we found evidence that
555 cortical responses to narrative speech reflect the well-formedness of phoneme segments
556 as expressed by probability values, therefore with no (or very few) violations and
557 phonological repair (**Figures 2 and 3**). This finding pushes beyond the phonotactic
558 violation paradigm and provides us with a tool based on linear models to isolate measures
559 of phonotactic-level processing during natural speech perception.

560 This work constitutes a further step towards the characterisation of brain responses to
561 natural speech, adding to recent work aimed at isolating brain responses to distinct
562 processing stages, involving speech acoustics (Ding and Simon, 2014), phonemes (Di
563 Liberto et al., 2015, 2018c), sentence structure (Ding et al., 2015, 2017), and semantic
564 similarity (Broderick et al., 2018). The ability to simultaneously account for and
565 disentangle brain responses to continuous speech at different processing stages constitutes
566 a novel and powerful tool to study the neurophysiology of speech. In particular, isolating
567 brain responses to phonotactics could provide new insights on the positive impact of this
568 mechanism in case of language impairment, and also when the phenomenon plays against
569 us. For example, when learning a second language, these brain mechanisms cause
570 misperception and mispronunciation, and contribute to stereotypical accents (Davidson,
571 2006a, 2006b; Lentz and Kager, 2015). In addition, the present framework produces
572 objective measures indicating how strongly EEG responses to speech correspond with a
573 particular phonotactic model, thus offering a new opportunity to test the
574 neurophysiological validity of theoretical and computational models (e.g. BLICK).

575 Our results provide new insights in this direction, indicating that phonotactic
576 probabilities, as defined by the computational model BLICK, are better represented in the
577 EEG signal than a purely probabilistic definition of phoneme probability (P_{sur} , P_{ent})
578 (Gaston and Marantz, 2018) and, importantly, than phonological neighbourhood density
579 (P_{neigh}) (Vitevitch et al., 1999; Frisch et al., 2000; Bailey and Hahn, 2001a). While further
580 studies could explore other hypotheses on the encoding and processing of phonotactic
581 information more comprehensively, the present finding is in line with research suggesting
582 distinct roles for phonotactics and neighbourhood density (Vitevitch et al., 1999; Bailey

583 and Hahn, 2001a; Storkel et al., 2006). Specifically, the first would aid speech perception
584 by facilitating processing and triggering learning of new words at early pre-lexical stages,
585 while the latter would influence the integration of new and existing lexical representations
586 at a later stage.

587 A similar issue relates to the speech-brain latencies associated with phonotactic
588 information. Indeed, previous research found interactions between phonotactic violations
589 and evoked brain components such as N400 and LPC (Domahs et al., 2009; White and
590 Chiu, 2017). It has also been suggested that the N400 magnitude may be directly linked
591 to phonotactic information, while effects at the longer LPC latencies may be spurious
592 and, instead, reflect other related processes, such as changes in cognitive load related to
593 the size of the neighbourhood of permissible words (Dupoux et al., 1999; Vitevitch et al.,
594 1999; Dupoux and Pallier, 2001; Storkel et al., 2006). Our results contribute to this debate
595 by suggesting that latencies of 100-400 ms are the most relevant for the processing of
596 phonotactic probabilities. Furthermore, topographical patterns at those latencies present
597 activations over centro-parietal scalp areas that qualitatively resemble that of an N400
598 component. One possibility is that this response is related to an early N400, whose
599 latencies reflect the rapid processing of phonotactics in a natural speech scenario. It is
600 also possible that this response reflects multiple cortical correlates, one in correspondence
601 with the earlier weaker effect (100-300 ms) (Brodbeck et al., 2018), and a separate one
602 with a larger effect-size at longer latencies (300-400 ms) (Pykkänen et al., 2002, 2000).
603 A direct comparison between EEG responses to phonotactic probabilities and phonotactic
604 violations could clarify some of these issues, as previously attempted in the similar
605 context of semantic-level processing (Broderick et al., 2018).

606 **Theoretical implications of a rapid time-locked response to phonotactics**

607 Our results have important implications for current theories on phonotactics, by providing
608 insights into both temporal dynamics (when) and neural encoding (how) of this cortical
609 mechanism. Phonotactic information, which aids speech recognition and learning of new
610 words (Mattys and Jusczyk, 2001; Munz, 2017), was suggested to involve one of the
611 following: 1) the phoneme identification stage (one-step models; Dehaene-Lambertz et
612 al., 2000; Dupoux et al., 2011); 2) a pre-lexical stage that occurs after phoneme
613 identification (two-step models; Church, 1987); or 3) a later lexical stage that influences
614 pre-lexical processes through feedback connections (lexicalist models; McClelland et al.,
615 2006; McClelland and Elman, 1986). In this context, a large body of literature in

616 psycholinguistics supports a pre-lexical account of phonotactics (McQueen, 1998;
617 Jusczyk et al., 1999; Sebastián-Gallés, 2007). For example, infants showed sensitivity to
618 phonotactics by 9 months of age, suggesting that this information aids speech
619 segmentation even at early developmental stages, before being able to understand speech
620 (Jusczyk et al., 1994). Similarly, it was shown that humans are sensitive to phonotactic
621 information even when meaning is not involved (nonsense words), pointing to the early
622 implementation of phonotactic repair (Dupoux et al., 1999; Davidson, 2011; Rossi et al.,
623 2013). This indirect evidence for a pre-lexical influence of phonotactic information finds
624 experimental support in both phonotactic violation studies (Dehaene-Lambertz et al.,
625 2000; Pykkänen et al., 2002) and in the present work, which isolated cortical responses
626 to probabilistic phonotactics showing short speech-EEG latencies (100-400 ms).

627 Indeed, it is possible that other post-lexical brain responses to phonotactics exist but could
628 not be measured. In fact, such higher-level effects could exhibit weaker time-locking,
629 which would hamper the ability to capture them with our framework. Indeed, this
630 hypothesis should be tested with more controlled experimental paradigms, possibly by
631 making less assumptions on the time-locking between phonotactics and brain signals.
632 Although we cannot be conclusive on this point, the latencies of 100-400 ms could be in
633 line with one-step models (Dehaene-Lambertz et al., 2000; Dupoux et al., 2011), which
634 hypothesise that phonotactic processing occurs pre-lexically and together with phoneme
635 identification, whose EEG responses were measured for latencies up to 300 ms (Di
636 Liberto et al., 2015; Khalighinejad et al., 2017).

637 In summary, our results indicate rapid time-locked brain responses to probabilistic
638 phonotactics. This phenomenon emerged for low-frequency cortical signals (< 9 Hz) and
639 were reliably measured at the individual subject-level. We also found that the speech-
640 EEG latencies of 100-400 ms most strongly reflects phonotactic information, which is in
641 line with a pre-lexical account of phonotactic processing. This provides the field with a
642 new tool to study the brain processing of phonotactics using natural speech.

643

644 **Author Contributions**

645 G.D.L. conceived the study and collected the data. G.D.L., A.d.C., and D.W. formulated
646 the data analysis procedure. G.D.L. analysed the data. G.D.L. wrote the first draft of the
647 manuscript. A.d.C., G.A.M., and D.W. edited the manuscript.

648

649 **Acknowledgements**

650 This study was supported by the EU H2020-ICT grant 644732 (COCOHA). The authors
651 would like to thank Dorothée Arzounian for useful discussions at the start of this study.

652 **References**

- 653 Bailey TM, Hahn U (2001a) Determinants of Wordlikeness: Phonotactics or Lexical
654 Neighborhoods? *J Mem Lang* 44:568–591 Available at:
655 <http://linkinghub.elsevier.com/retrieve/pii/S0749596X00927566> [Accessed May
656 31, 2018].
- 657 Bailey TM, Hahn U (2001b) Determinants of Wordlikeness: Phonotactics or Lexical
658 Neighborhoods? *J Mem Lang* 44:568–591 Available at:
659 <http://www.academicpress.com> [Accessed June 1, 2018].
- 660 Balling LW, Baayen RH (2012) Probability and surprisal in auditory comprehension of
661 morphologically complex words. *Cognition* 125:80–106 Available at:
662 <http://www.ncbi.nlm.nih.gov/pubmed/22841290> [Accessed May 30, 2018].
- 663 Brent MR, Cartwright TA (1996) Distributional regularity and phonotactic constraints
664 are useful for segmentation. *Cognition* 61:93–125 Available at:
665 <https://www.sciencedirect.com/science/article/pii/S0010027796007196> [Accessed
666 June 6, 2018].
- 667 Brodbeck C, Hong LE, Simon JZ (2018) Transformation from auditory to linguistic
668 representations across auditory cortex is rapid and attention dependent for
669 continuous speech. *bioRxiv*:326785 Available at:
670 <https://www.biorxiv.org/content/early/2018/05/21/326785.1> [Accessed June 1,
671 2018].
- 672 Broderick MP, Anderson AJ, Di Liberto GM, Crosse MJ, Lalor EC (2018)
673 Electrophysiological Correlates of Semantic Dissimilarity Reflect the
674 Comprehension of Natural, Narrative Speech. *Curr Biol*.
- 675 Carlson MT, Goldrick M, Blasingame M, Fink A (2016) Navigating conflicting
676 phonotactic constraints in bilingual speech perception. *Bilingualism* 19:939–954
677 Available at: http://www.journals.cambridge.org/abstract_S1366728915000334
678 [Accessed June 1, 2018].
- 679 Chomsky N, Halle M (1968) The sound pattern of English.
- 680 Church KW (1987) Phonological parsing and lexical retrieval. *Cognition* 25:53–69

- 681 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/3581729> [Accessed June 2,
682 2018].
- 683 Cibelli ES, Leonard MK, Johnson K, Chang EF (2015) The influence of lexical
684 statistics on temporal lobe cortical dynamics during spoken word listening. *Brain*
685 *Lang* 147:66–75 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26072003>
686 [Accessed June 8, 2018].
- 687 Connolly JF, Phillips NA (1994) Event-Related Potential Components Reflect
688 Phonological and Semantic Processing of the Terminal Word of Spoken Sentences.
689 *J Cogn Neurosci* 6:256–266 Available at:
690 <http://www.ncbi.nlm.nih.gov/pubmed/23964975> [Accessed May 30, 2018].
- 691 Crosse MJ, Di Liberto GM, Bednar A, Lalor EC (2016a) The multivariate temporal
692 response function (mTRF) toolbox: A MATLAB toolbox for relating neural
693 signals to continuous stimuli. *Front Hum Neurosci* 10.
- 694 Crosse MJ, Di Liberto GM, Lalor EC (2016b) Eye can hear clearly now: Inverse
695 effectiveness in natural audiovisual speech processing relies on long-term
696 crossmodal temporal integration. *J Neurosci* 36.
- 697 Davidson L (2006a) Phonotactics and articulatory coordination interact in phonology:
698 Evidence from nonnative production. *Cogn Sci* 30:837–862 Available at:
699 http://doi.wiley.com/10.1207/s15516709cog0000_73 [Accessed June 6, 2018].
- 700 Davidson L (2006b) Phonology, phonetics, or frequency: Influences on the production
701 of non-native sequences. *J Phon* 34:104–137 Available at:
702 <http://linkinghub.elsevier.com/retrieve/pii/S0095447005000240> [Accessed June 6,
703 2018].
- 704 Davidson L (2011) Phonetic, Phonemic, and Phonological Factors in Cross-Language
705 Discrimination of Phonotactic Contrasts. *J Exp Psychol Hum Percept Perform*
706 37:270–282 Available at: <http://doi.apa.org/getdoi.cfm?doi=10.1037/a0020988>
707 [Accessed June 2, 2018].
- 708 de Cheveigné A, Di Liberto GM, Arzounian D, Wong D, Hjortkjaer J, Fuglsang SA,
709 Parra LC (2018a) Multiway Canonical Correlation Analysis of Brain Signals.
710 bioRxiv:344960 Available at:

- 711 <https://www.biorxiv.org/content/early/2018/06/12/344960> [Accessed June 13,
712 2018].
- 713 de Cheveigné A, Wong DE, Di Liberto GM, Hjortkjær J, Slaney M, Lalor E (2018b)
714 Decoding the auditory brain with canonical component analysis. *Neuroimage*
715 172:206–216 Available at:
716 <https://www.sciencedirect.com/science/article/pii/S1053811918300338> [Accessed
717 June 1, 2018].
- 718 Dehaene-Lambertz G, Dupoux E, Gout A (2000) Electrophysiological correlates of
719 phonological processing: a cross-linguistic study. *J Cogn Neurosci* 12:635–647
720 Available at: [http://www.unicog.org/publications/Dehaene-](http://www.unicog.org/publications/Dehaene-LambertzDupoux_EbuzERP_JCogNS2000.pdf)
721 [LambertzDupoux_EbuzERP_JCogNS2000.pdf](http://www.unicog.org/publications/Dehaene-LambertzDupoux_EbuzERP_JCogNS2000.pdf) [Accessed May 30, 2018].
- 722 Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-
723 trial EEG dynamics including independent component analysis. *J Neurosci*
724 *Methods* 134:9–21.
- 725 Di Liberto GM, Crosse MJ, Lalor EC (2018a) Cortical Measures of Phoneme-Level
726 Speech Encoding Correlate with the Perceived Clarity of Natural Speech. *Eneuro*
727 5:ENEURO.0084-18.2018 Available at:
728 <http://eneuro.sfn.org/lookup/doi/10.1523/ENEURO.0084-18.2018>.
- 729 Di Liberto GM, Lalor EC (2017) Indexing cortical entrainment to natural speech at the
730 phonemic level: Methodological considerations for applied research. *Hear Res*
731 348:70–77.
- 732 Di Liberto GM, Lalor EC, Millman RE (2018b) Causal cortical dynamics of a
733 predictive enhancement of speech intelligibility. *Neuroimage* 166.
- 734 Di Liberto GM, O’Sullivan JA, Lalor EC (2015) Low-frequency cortical entrainment to
735 speech reflects phoneme-level processing. *Curr Biol* 25.
- 736 Di Liberto GM, Peter V, Kalashnikova M, Goswami U, Burnham D, Lalor EC (2018c)
737 Atypical cortical entrainment to speech in the right hemisphere underpins
738 phonemic deficits in dyslexia. *Neuroimage NIMG-17-29:70–79* Available at:
739 <https://www.sciencedirect.com/science/article/pii/S1053811918302829> [Accessed
740 June 1, 2018].

- 741 Ding N, Melloni L, Yang A, Wang Y, Zhang W, Poeppel D (2017) Characterizing
742 Neural Entrainment to Hierarchical Linguistic Units using Electroencephalography
743 (EEG). *Front Hum Neurosci* 11:481 Available at:
744 <http://journal.frontiersin.org/article/10.3389/fnhum.2017.00481/full> [Accessed
745 December 30, 2017].
- 746 Ding N, Melloni L, Zhang H, Tian X, Poeppel D (2015) Cortical tracking of
747 hierarchical linguistic structures in connected speech. *Nat Neurosci* 19:158–164
748 Available at: <http://www.nature.com/doi/10.1038/nn.4186> [Accessed
749 December 30, 2017].
- 750 Ding N, Simon JZ (2012) Neural coding of continuous speech in auditory cortex during
751 monaural and dichotic listening. *J Neurophysiol* 107:78–89 Available at:
752 <http://jn.physiology.org/content/jn/107/1/78.full.pdf>.
- 753 Ding N, Simon JZ (2014) Cortical Entrainment to Continuous Speech: Functional Roles
754 and Interpretations. *Front Hum Neurosci* 8 Available at:
755 http://www.frontiersin.org/Journal/Abstract.aspx?s=537&name=human_neuroscience&ART_DOI=10.3389/fnhum.2014.00311.
- 757 Domahs U, Kehrein W, Knaus J, Wiese R, Schlesewsky M (2009) Event-related
758 Potentials Reflecting the Processing of Phonological Constraint Violations. *Lang*
759 *Speech* 52:415–435 Available at:
760 <http://journals.sagepub.com/doi/10.1177/0023830909336581> [Accessed June 2,
761 2018].
- 762 Dupoux E, Kakehi K, Hirose Y, Pallier C, Mehler J (1999) Epenthetic vowels in
763 Japanese: A perceptual illusion? *J Exp Psychol Hum Percept Perform* 25:1568–
764 1578 Available at: <http://cogprints.org/747/5/ebuzo.pdf> [Accessed May 30, 2018].
- 765 Dupoux E, Pallier C (2001) New evidence for prelexical phonological processing in
766 word recognition. *Lang Cogn Process* 16:491–505 Available at:
767 http://www.lscop.net/persons/dupoux/papers/Dupoux_PKM_2001_Prelexical_epenthesis_LangCogProc.pdf [Accessed June 1, 2018].
- 769 Dupoux E, Parlato E, Frota S, Hirose Y, Peperkamp S (2011) Where do illusory vowels
770 come from? *J Mem Lang* 64:199–210 Available at:

- 771 http://www.lscpl.net/persons/peperkamp/Dupoux_Parlato_Frota_Hirose_Peperkam
772 [p_\(2011\)_Where_do_illusory_vowels_come_from.pdf](http://www.lscpl.net/persons/peperkamp/Dupoux_Parlato_Frota_Hirose_Peperkam_p_(2011)_Where_do_illusory_vowels_come_from.pdf) [Accessed June 2, 2018].
- 773 Ettinger A, Linzen T, Marantz A (2014) The role of morphology in phoneme prediction:
774 Evidence from MEG. Available at:
775 [http://ling.umd.edu/assets/publications/Ettinger-Linzen-Marantz-14-](http://ling.umd.edu/assets/publications/Ettinger-Linzen-Marantz-14-MorphologyInPhonemePrediction.pdf)
776 [MorphologyInPhonemePrediction.pdf](http://ling.umd.edu/assets/publications/Ettinger-Linzen-Marantz-14-MorphologyInPhonemePrediction.pdf) [Accessed May 30, 2018].
- 777 Frisch SA, Large NR, Pisoni DB (2000) Perception of Wordlikeness: Effects of
778 Segment Probability and Length on the Processing of Nonwords. *J Mem Lang*
779 42:481–496 Available at: http://cas.usf.edu/~frisch/Frisch_Large_Pisoni_00.pdf
780 [Accessed May 18, 2018].
- 781 Gaston P, Marantz A (2018) The time course of contextual cohort effects in auditory
782 processing of category-ambiguous words: MEG evidence for a single “clash” as
783 noun or verb. *Lang Cogn Neurosci* 33:402–423 Available at:
784 <https://www.tandfonline.com/doi/full/10.1080/23273798.2017.1395466> [Accessed
785 June 1, 2018].
- 786 Goldwater S, Johnson M (2003) Learning OT Constraint Rankings Using a Maximum
787 Entropy Model. *Proc Stock Work Var within Optim Theory*:111–120 Available at:
788 <http://homepages.inf.ed.ac.uk/sgwater/papers/OTvar03.pdf> [Accessed June 1,
789 2018].
- 790 Gorman K, Howell J, Wagner M (2011) Prosodylab-aligner: A tool for forced
791 alignment of laboratory speech. *Can Acoust - Acoust Can* 39:192–193 Available
792 at: [https://www.scopus.com/inward/record.uri?eid=2-s2.0-](https://www.scopus.com/inward/record.uri?eid=2-s2.0-84859480980&partnerID=40&md5=6d828c3caa30dd5c98fdf421f5a0b762)
793 [84859480980&partnerID=40&md5=6d828c3caa30dd5c98fdf421f5a0b762](https://www.scopus.com/inward/record.uri?eid=2-s2.0-84859480980&partnerID=40&md5=6d828c3caa30dd5c98fdf421f5a0b762).
- 794 Greenwood DD (1961) Auditory Masking and the Critical Band. *J Acoust Soc Am*
795 33:484–502 Available at:
796 <http://scitation.aip.org/content/asa/journal/jasa/33/4/10.1121/1.1908699>.
- 797 Hallé PA, Dominguez A, Cuetos F, Segui J (2008) Phonological mediation in visual
798 masked priming: Evidence from phonotactic repair. *J Exp Psychol Hum Percept*
799 *Perform* 34:177–192 Available at:
800 <http://doi.apa.org/getdoi.cfm?doi=10.1037/0096-1523.34.1.177> [Accessed June 1,

- 801 2018].
- 802 Hammond M (2004) Gradience, Phonotactics, and the Lexicon in English Phonology.
803 Int J English Stud 4 Available at: <http://roa.rutgers.edu/files/736-0505/736->
804 HAMMOND-0-0.PDF [Accessed May 30, 2018].
- 805 Hayes B (2012) BLICK : a phonotactic probability calculator (manual). Available at:
806 <http://linguistics.ucla.edu/people/hayes/BLICK/BLICKManual.pdf> [Accessed June
807 1, 2018].
- 808 Hayes B, Wilson C (2008) A Maximum Entropy Model of Phonotactics and
809 Phonotactic Learning. *Linguist Inq* 39:379–440 Available at:
810 [http://linguistics.ucla.edu/people/hayes/papers/HayesAndWilsonPhonotactics2008.](http://linguistics.ucla.edu/people/hayes/papers/HayesAndWilsonPhonotactics2008.pdf)
811 pdf [Accessed June 1, 2018].
- 812 Hotelling H (1936) Relations Between Two Sets of Variates. *Biometrika* 28:321
813 Available at: <https://www.jstor.org/stable/2333955?origin=crossref> [Accessed June
814 1, 2018].
- 815 Jusczyk PW, Houston DM, Newsome M (1999) The Beginnings of Word Segmentation
816 in English-Learning Infants. *Cogn Psychol* 39:159–207 Available at:
817 <https://www.sciencedirect.com/science/article/pii/S0010028599907168> [Accessed
818 June 2, 2018].
- 819 Jusczyk PW, Luce PA, Charles-Luce J (1994) Infants' Sensitivity to Phonotactic
820 Patterns in the Native Language. *J Mem Lang* 33:630–645 Available at:
821 <https://www.sciencedirect.com/science/article/pii/S0749596X84710308> [Accessed
822 June 2, 2018].
- 823 Khalighinejad B, Cruzatto da Silva G, Mesgarani N (2017) Dynamic Encoding of
824 Acoustic Features in Neural Responses to Continuous Speech. *J Neurosci*.
- 825 Kösem A, van Wassenhove V (2016) Distinct contributions of low- and high-frequency
826 neural oscillations to speech comprehension. *Lang Cogn Neurosci*:1–9 Available
827 at: <http://dx.doi.org/10.1080/23273798.2016.1238495>.
- 828 Kutas M, Federmeier KD (2011) Thirty years and counting: finding meaning in the
829 N400 component of the event-related brain potential (ERP). *Annu Rev Psychol*

- 830 62:621–647 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20809790>
831 [Accessed December 30, 2017].
- 832 Lalor EC, Pearlmutter BA, Reilly RB, McDarby G, Foxe JJ (2006) The VESPA: a
833 method for the rapid estimation of a visual evoked potential. *Neuroimage* 32:1549–
834 1561 Available at: [http://ac.els-cdn.com/S1053811906006434/1-s2.0-
835 S1053811906006434-main.pdf?_tid=ff77a230-a642-11e4-8c11-
836 00000aacb35e&acdnat=1422376921_3d415c577602776e8b0c6a9e8425ac29](http://ac.els-cdn.com/S1053811906006434/1-s2.0-S1053811906006434-main.pdf?_tid=ff77a230-a642-11e4-8c11-00000aacb35e&acdnat=1422376921_3d415c577602776e8b0c6a9e8425ac29).
- 837 Lalor EC, Power AJ, Reilly RB, Foxe JJ (2009) Resolving Precise Temporal Processing
838 Properties of the Auditory System Using Continuous Stimuli. *J Neurophysiol*
839 102:349–359 Available at: <http://jn.physiology.org/content/jn/102/1/349.full.pdf>.
- 840 Lentz TO, Kager RWJ (2015) Categorical phonotactic knowledge filters second
841 language input, but probabilistic phonotactic knowledge can still be acquired. *Lang*
842 *Speech* 58:387–413 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26529903>
843 [Accessed June 6, 2018].
- 844 Leonard MK, Bouchard KE, Tang C, Chang EF (2015) Dynamic Encoding of Speech
845 Sequence Probability in Human Temporal Cortex. *J Neurosci* 35:7203–7214
846 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25948269> [Accessed June 8,
847 2018].
- 848 Luck SJ (2005) An introduction to the event-related potential technique.
- 849 Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG-and MEG-data.
850 *J Neurosci Methods* 164:177–190.
- 851 Mattys SL, Jusczyk PW (2001) Phonotactic cues for segmentation of fluent speech by
852 infants. *Cognition* 78:91–121 Available at:
853 <http://www.ncbi.nlm.nih.gov/pubmed/11074247> [Accessed June 2, 2018].
- 854 Mattys SL, Jusczyk PW, Luce PA, Morgan JL (1999) Phonotactic and Prosodic Effects
855 on Word Segmentation in Infants. *Cogn Psychol* 38:465–494 Available at:
856 <http://www.ncbi.nlm.nih.gov/pubmed/10334878> [Accessed May 30, 2018].
- 857 McClelland JL, Elman JL (1986) The TRACE model of speech perception. *Cogn*
858 *Psychol* 18:1–86.

- 859 McClelland JL, Mirman D, Holt LL (2006) Are there interactive processes in speech
860 perception? *Trends Cogn Sci* 10:363–369.
- 861 McQueen JM (1998) Segmentation of continuous speech using phonotactics. *J Mem*
862 *Lang* 39:21–46 Available at:
863 <https://www.sciencedirect.com/science/article/pii/S0749596X98925682> [Accessed
864 June 2, 2018].
- 865 Mesgarani N, Cheung C, Johnson K, Chang EF (2014) Phonetic Feature Encoding in
866 Human Superior Temporal Gyrus. *Science* (80-) 343:1006–1010 Available at:
867 <http://www.sciencemag.org/content/343/6174/1006.abstract>.
- 868 Munz ED (2017) Psychotherapie in der Psychiatrie. *Nervenheilkunde* 36:800–805
869 Available at: <https://www.era.lib.ed.ac.uk/handle/1842/10432> [Accessed June 2,
870 2018].
- 871 O’Sullivan JA, Power AJ, Mesgarani N, Rajaram S, Foxe JJ, Shinn-Cunningham BG,
872 Slaney M, Shamma SA, Lalor EC (2014) Attentional Selection in a Cocktail Party
873 Environment Can Be Decoded from Single-Trial EEG. *Cereb Cortex*:bht355.
- 874 Obleser J, Herrmann B, Henry MJ (2012) Neural Oscillations in Speech: Don’t be
875 Enslaved by the Envelope. *Front Hum Neurosci* 6:250 Available at:
876 <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3431501/>.
- 877 Obrig H, Mentzel J, Rossi S (2016) Universal and language-specific sublexical cues in
878 speech perception: a novel electroencephalography-lesion approach. *Brain*
879 139:1800–1816 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27190021>
880 [Accessed May 30, 2018].
- 881 Parker SG (Stephen G (2012) *The sonority controversy*. De Gruyter Mouton. Available
882 at:
883 [https://books.google.fr/books?id=ixpO4NZD2gkC&dq=Parker,+S.+\(Ed.\).+\(2012\).
884 +The+Sonority+Controversy&lr=&hl=it&source=gbs_navlinks_s](https://books.google.fr/books?id=ixpO4NZD2gkC&dq=Parker,+S.+(Ed.).+(2012).+The+Sonority+Controversy&lr=&hl=it&source=gbs_navlinks_s) [Accessed May
885 18, 2018].
- 886 Pisoni DB, Remez RE (2005) *The handbook of speech perception*. Blackwell Pub.
887 Available at:
888 <https://books.google.fr/books?id=EwY15naRiFgC&pg=PA619&lpg=PA619&dq=t>

- 889 race+model+phonotactic&source=bl&ots=0OXeu89-
890 4S&sig=EWPP5YRtV4Odmn1E1aRk0YzJUEY&hl=it&sa=X&ved=0ahUKEwiA
891 uurYqL_bAhWIbRQKHeFICy8Q6AEIWzAG#v=onepage&q=trace model
892 phonotactic&f=false [Accessed June 6, 2018].
- 893 Pylkkänen L, Stringfellow A, Flagg E, Marantz A (2000) A neural response sensitive to
894 repetition and phonotactic probability : MEG investigations of lexical access. Proc
895 Biomag 2000, 12th Int Conf Biomagn:1–4 Available at:
896 <https://pdfs.semanticscholar.org/5d61/dcc9f304711f79ff230cee855855c149eec5.pdf>
897 f [Accessed June 1, 2018].
- 898 Pylkkänen L, Stringfellow A, Marantz A (2002) Neuromagnetic Evidence for the
899 Timing of Lexical Activation: An MEG Component Sensitive to Phonotactic
900 Probability but Not to Neighborhood Density. *Brain Lang* 81:666–678 Available
901 at: <https://www.sciencedirect.com/science/article/pii/S0093934X01925556>
902 [Accessed June 1, 2018].
- 903 Rossi S, Hartmüller T, Vignotto M, Obrig H (2013) Electrophysiological evidence for
904 modulation of lexical processing after repetitive exposure to foreign phonotactic
905 rules. *Brain Lang* 127:404–414 Available at:
906 <http://www.ncbi.nlm.nih.gov/pubmed/23489581> [Accessed June 2, 2018].
- 907 Rossi S, Jürgenson IB, Hanulíková A, Telkemeyer S, Wartenburger I, Obrig H (2011)
908 Implicit Processing of Phonotactic Cues: Evidence from Electrophysiological and
909 Vascular Responses. *J Cogn Neurosci* 23:1752–1764 Available at:
910 <http://www.ncbi.nlm.nih.gov/pubmed/20666594> [Accessed May 30, 2018].
- 911 Scholes RJ (1966) *Phonotactic Grammaticality*. Hague Mout Co.
- 912 Sebastián-Gallés N (2007) Biased to learn language. *Dev Sci* 10:713–718 Available at:
913 <http://doi.wiley.com/10.1111/j.1467-7687.2007.00649.x> [Accessed June 2, 2018].
- 914 Storkel HL (2001) Learning New Words. *J Speech Lang Hear Res* 44:1321 Available
915 at: [http://jslhr.pubs.asha.org/article.aspx?doi=10.1044/1092-4388\(2001/103\)](http://jslhr.pubs.asha.org/article.aspx?doi=10.1044/1092-4388(2001/103))
916 [Accessed December 30, 2017].
- 917 Storkel HL (2004) The emerging lexicon of children with phonological delays:
918 phonotactic constraints and probability in acquisition. *J Speech Lang Hear Res*

- 919 47:1194–1212 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15603471>
920 [Accessed May 31, 2018].
- 921 Storkel HL, Armbrüster J, Hogan TP (2006) Differentiating Phonotactic Probability and
922 Neighborhood Density in Adult Word Learning. *J Speech Lang Hear Res* 49:1175
923 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17197489> [Accessed June 1,
924 2018].
- 925 Storkel HL, Morrisette ML (2002) The Lexicon and Phonology. *Lang Speech Hear Serv*
926 *Sch* 33:24 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27764412> [Accessed
927 May 31, 2018].
- 928 Storkel HL, Rogers MA (2000) The effect of probabilistic phonotactics on lexical
929 acquisition. *Clin Linguist Phon*:407–425 Available at:
930 <https://wordlearning.ku.edu/storkel-hl-rogers-ma-2000> [Accessed May 31, 2018].
- 931 Vanthornhout J, Decruy L, Wouters J, Simon JZ, Francart T (2018) Speech
932 Intelligibility Predicted from Neural Entrainment of the Speech Envelope. *J Assoc*
933 *Res Otolaryngol* 19:181–191 Available at:
934 <http://link.springer.com/10.1007/s10162-018-0654-z> [Accessed June 11, 2018].
- 935 Vitevitch MS, Luce PA, Pisoni DB, Auer ET (1999) Phonotactics, neighborhood
936 activation, and lexical access for spoken words. *Brain Lang* 68:306–311 Available
937 at: <http://www.ncbi.nlm.nih.gov/pubmed/10433774> [Accessed May 30, 2018].
- 938 Wagner M, Shafer VL, Martin B, Steinschneider M (2012) The phonotactic influence
939 on the perception of a consonant cluster /pt/ by native English and native Polish
940 listeners: A behavioral and event related potential (ERP) study. *Brain Lang*
941 123:30–41 Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22867752>
942 [Accessed June 11, 2018].
- 943 White J, Chiu F (2017) Disentangling phonological well-formedness and attestedness:
944 An ERP study of onset clusters in English. *Acta Linguist Acad* 64:513–537
945 Available at: <http://www.akademai.com/doi/10.1556/2062.2017.64.4.2> [Accessed
946 June 1, 2018].
- 947 Wiese R, Orzechowska P, Alday PM, Ulbrich C (2017) Structural Principles or
948 Frequency of Use? An ERP Experiment on the Learnability of Consonant Clusters.

949 7 Available at:
950 <https://pdfs.semanticscholar.org/a581/f7fd74aa28e80ea6ce5eff8d0fbfde7c79e8.pdf>
951 [Accessed May 30, 2018].

952 Winther Balling L, Harald Baayen R (2008) Morphological effects in auditory word
953 recognition: Evidence from Danish. *Lang Cogn Process* 23:1159–1190 Available
954 at: <http://www.tandfonline.com/doi/abs/10.1080/01690960802201010> [Accessed
955 May 30, 2018].

956

957

958 **Figures**

959 **Figure 1. (A) Speech representations** for a 5 seconds portion of the stimulus. From bottom to top, the
960 acoustic spectrogram (S) which consists of a 16-channel time series of power within 16 frequency bands;
961 phonetic features (F), whose permissible combinations map to English phonemes; phoneme onsets (O),
962 which mark the beginning of each phoneme; and the probabilistic phonotactic vector (P), a representation
963 indicating the inverse likelihood of a sequence (from the beginning of a word to each of its phonemes). **(B)**
964 **Expected outcomes:** We hypothesise that, if a stimulus representation encodes features not captured by
965 other representations, adding it to the others will improve the prediction of cortical responses. In particular
966 we predict an increase in cortical tracking due when phonotactic probabilities are added to the mix (POFS
967 – OFS, blue increment).

968 **Figure 2: EEG responses to natural speech are best explained when including phonotactic probability**
969 **among the speech features.** Data from all participants were combined using MCCA. This consensus EEG
970 signal (CS) preserves signals that are maximally correlated across subjects. **(A)** A CCA analysis was
971 conducted between each speech representation and the CS signals. Speech-EEG correlations for the first
972 canonical component (CC) pair were best when using the combined model POFS, indicating that
973 phonotactic probabilities explain EEG variance that was not captured by the purely acoustic-phonemic
974 models (S, FS, and OFS). **(B)** In addition, phonotactic probabilities enhanced the *d-prime* score of a match-
975 vs-mismatch classification test. The box-plots indicate the 99th percentile of the performance when using a
976 combined model (FS, OFS, or POFS) after randomly shuffling information for the newly added feature (F,
977 O, and P respectively).

978 **Figure 3: Phonotactic probabilities enhance the speech-EEG mapping at the individual subject level.**
979 CCA analyses were conducted between each speech representation and the corresponding EEG responses
980 for each individual subject. **(A)** Speech-EEG correlations for the first canonical component pair were
981 greatest when using the combined model POFS (left panel). The thick black line indicates the average
982 across subjects while the coloured dots/lines refer to the individual subjects. The bar-plot shows the relative
983 correlation gain (%) of the combined models OFS and POFS with FS (i.e. the contribution given by O and
984 P respectively). **(B)** Similar results are shown for the *d-prime* scores of a match-vs-mismatch classification
985 test. Results for individual subjects are colour-coded (same colors as for A). Phonotactic probabilities
986 enhance the single-subject scores for FS and also show significant improvement compared to OFS.

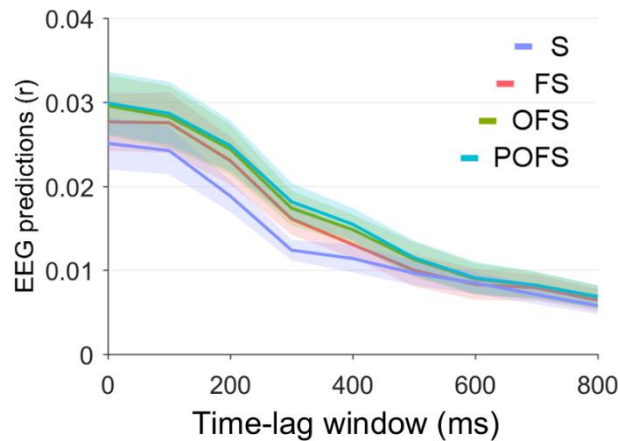
987

988 **Figure 4: EEG tracking of phonotactic probabilities is specific to speech-brain latencies of 100-400**
989 **ms.** A temporal response function (TRF) analysis was conducted to estimate the amount of EEG variance
990 explained by phonotactic probabilities for speech-EEG latency windows between 0 and 900 ms and
991 window-size 100 ms. EEG prediction correlations were calculated for different speech feature-sets and for
992 the various speech-EEG latencies. The enhancement in EEG predictions due to phonotactic probabilities is
993 shown for all time-latency windows. Shaded areas indicate the standard error of the mean (SE) across
994 subjects. Stars indicate significant enhancement ($*p < 0.05$) as a result of a cluster mass statistics (top).
995 Cohen's *d* was calculated to measure the effect size of the enhancement due to phonotactics. Values above

996 0.8 are considered as 'large' effects (above dashed grey line) (centre). Topographical patterns of the TRF
997 weights for a model fit over time-lags from 0 to 900 ms are shown for latencies with a significant effect of
998 phonotactic probabilities (100-400 ms) (bottom).

999

1000 **Extended data**



1001

1002 **Figure 4-1.** A temporal response function (TRF) analysis was conducted to estimate the amount of EEG
1003 variance explained by phonotactic probabilities for speech-EEG latency windows between 0 and 900 ms
1004 and window-size 100 ms. EEG prediction correlations averaged across all scalp electrodes are shown for
1005 different speech feature-sets and for the various speech-EEG latencies. Shaded areas indicate the standard
1006 error of the mean (SE) across subjects. The contrast between EEG prediction values for POFS and OFS is
1007 shown in Figure 4.