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

Abstract

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

Significance Statement

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

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

Introduction

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

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

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

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

Material and methods

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

Subjects and Experimental Procedure

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

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

Speech representations

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

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

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

3. Phoneme onsets (**O**): This vector marks phoneme onsets with a discrete-time unit impulse, corresponding to the half-wave rectified first derivative of F. This is a non-linear transformation of the F features, thus linear models may benefit from the explicit definition of O combined with F.
4. Finally, we propose a novel representation using *phonotactic probabilities* (**P**). Natural languages include various constraints on the permissible phoneme sequences. Probabilities can be derived for a given speech token from this set of constraints. For example, the pseudoword *blick* would "sound" better than *bnick* to a native English speaker, which is reflected by a higher phonotactic probability for the first word. Here, we used a computational model (BLICK; Hayes and Wilson, 2008) based on a combination of explicit theoretical rules from traditional phonology and a maxent grammar (Goldwater and Johnson, 2003), which find the optimal weights for such theoretical constraints to best match the phonotactic intuition of a native speaker. Specifically, given a phoneme sequence $ph_{1..n}$, P is composed of two vectors: a) inverse phonotactic probability (score($ph_{1..n}$) is the output of the BLICK software; it is small for well-formed tokens and large for ill-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 phoneme to the well-formedness of the sequence.

In order to assess and quantify the contribution of each of the features F, O, and P to the speech-EEG mapping, the main analyses were conducted on the cumulative combinations S, FS, OFS, and POFS. The rationale is that, if the new feature carries information not subsumed by the other features, including it will improve the fitting score. To control for any potential effect of the difference in dimensionality of the feature space, we also used variants where the newly introduced feature did not correspond to the auditory stimulus, i.e. was shuffled (the entire procedure, including model fit, was rerun for each shuffled version). These mismatched vectors/matrices were generated by randomly shuffling: a) Phonetic features in the FS speech representation ($F_{shu}S$) (every given phoneme, corresponding to a combination of N_F phonetic features, 1 for vowels and 3 for consonants, was replaced by N_F random phonetic features for its entire duration); b) Onset time in OFS ($O_{shu}FS$) (the onset vector O was replaced by vector with the same number of impulses at random time points); and c) Phonotactic probability values in POFS ($P_{shu}OFS$) (the values in the phonotactic vector P were randomly permuted while keeping the time information).

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

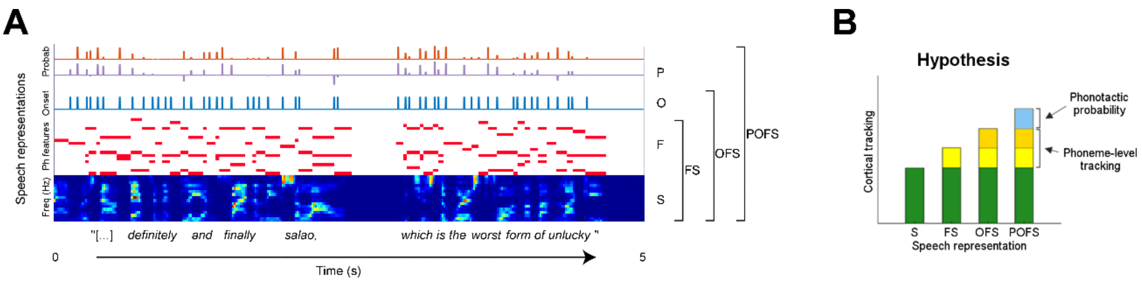

Figure 1. (A) Speech representations for a 5 seconds portion of the stimulus. From bottom to top, the acoustic spectrogram (S) which consists of a 16-channel time series of power within 16 frequency bands; phonetic features (F), whose permissible combinations map to English phonemes; phoneme onsets (O),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

Phonotactic Probability Model

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

Data Acquisition and Preprocessing

Electroencephalographic (EEG) data were recorded from 128 scalp electrodes (plus 2 mastoid channels), filtered over the range 0 - 134 Hz, and digitised with a sampling frequency of 512 Hz using a BioSemi Active Two system. Data were analysed offline using MATLAB software (The Mathworks Inc.). EEG data were digitally filtered between 0.5 and 32 Hz using a Butterworth zero-phase filter (low- and high-pass filters both with order 2; implemented with the function *filtfilt*), and down-sampled to 64 Hz. EEG channels with a variance exceeding three times that of the surrounding channels were replaced by an estimate calculated using spherical spline interpolation (EEGLAB; Delorme and Makeig, 2004). All channels were then re-referenced to the average of the two mastoid channels with the goal of maximizing the EEG responses to the auditory stimuli (Luck, 2005).

Dimensionality reduction

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

Denoising with multiway CCA

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

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

Analysis Procedure

Stimulus-response model based on Canonical Correlation Analysis

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

In its more general definition, given two sets of multichannel data X_1 and X_2 of size $T \times J_1$ and $T \times J_2$, CCA finds linear transformations of both that make them maximally correlated. Specifically, CCA produces the transformation matrices W_1 and W_2 (sizes $J_1 \times J_0$ and $J_2 \times J_0$, where $J_0 < \min(J_1,J_2)$) that maximise the correlation between pairs of columns of X_1W_1 and X_2W_2, while making the columns of each transformed data matrix X_iW_i mutually uncorrelated. The first pair of canonical components (CC) is the linear combination of X_1 and X_2 with highest possible correlation. The next pair of CCs are the most highly correlated combinations orthogonal to the first, and so-on.

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

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

Temporal Response Function Analysis

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

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

Measuring the quality of the speech-EEG mapping

We used two metrics to quantify the quality of the CCA-based speech-EEG mapping model: correlation and discriminability in a match-vs-mismatch classification task. A Pearson's correlation coefficient was calculated for each CC-pair. The first CC-pair is the most relevant, but meaningful speech-EEG correlations can arise for an arbitrary number of components. To obtain a measure sensitive to these multiple dimensions, we introduced a match-vs-mismatch classification task that consisted in deciding whether a segment of EEG (duration $T_{DECODER}$) was produced by the segment of speech that gave rise to it, or by some other segment. Discriminability in this task, measured by *d-prime*, reflects the ability of the model to capture the relation between speech and EEG. The *d-prime* metric was derived from the discriminant function of a support vector machine (SVM) classifier trained on the normalised Euclidean distance between pairs of CCs. A cross-validation procedure (k = 30) was used in which the classifier was trained and evaluated on distinct data to discriminate between match and mismatch segments. $T_{DECODER}$ was set to the value 1 second, which avoided saturation (classification either too easy or too difficult) in both group and single-subject level analyses.

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

Statistical Analyses

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

Results

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

Neural Evidence for the Processing of Probabilistic Phonotactics

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

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

We first estimated the quality of a CCA-based model involving only the phonotactic feature vector (P; **Figure 1A,top**) and EEG. The r-value of 0.42 obtained for the first CC-pair was larger than the 99th percentile of a distribution obtained by shuffling the values of the pulses within the P vector while leaving their times intact (median over 100 shuffles: r = 0.34; 99th percentile: r = 0.35). This result indicates that phonotactic probabilities were reflected by the EEG signals. However the phonotactic feature vector is correlated with other predictive features (such as spectrogram or phonemes), so we cannot be sure that its predictive power stems from phonotactic information per se. For that, we must compare combinations of features that include, or not, the phonotactic vector P. We formed combinations of features including the acoustic spectrogram S (Di Liberto et al., 2015; Lalor et al., 2009; Obleser et al., 2012), a phoneme representation based on phonetic features F (Mesgarani et al., 2014; Di Liberto et al., 2015, 2018a), phoneme onsets O (Brodbeck et al., 2018) and our newly introduced phonotactic features P (see **Figure 1A**). If each of these features carries information complementary to the others, and not captured by them, we expect speech-EEG correlations to monotonically increase with the inclusion of additional features in the analysis: namely S, FS, OFS, and POFS as schematized in **Figure 1B**. Indeed, correlation coefficient values for CCA models based on these four combination of features agree with this prediction (**Figure 2A**; $r_S < r_{FS} < r_{OFS} < r_{POFS}$). Of possible concern is that these models differ in the number of dimensions (and thus parameters) involved. A large number of parameters can lead to overfitting, which should penalise the models with more features, contrary to what we observe. To further exclude such a possibility, we randomly shuffled the values of the pulses within the phonotactic feature vectors while keeping their timing constant. The distribution of correlation scores for P_{shu}OFS obtained by repeated shuffling is indicated in **Figure 2A**. The value obtained for POFS is above the 99th percentile of that distribution. This same control procedure was applied to the F and O features and confirmed that their respective enhancements are driven by the addition of meaningful features, and not by differences in dimensionality, as they produced stronger correlations than the 99th percentile of the corresponding shuffled distributions. In summary, each of these features carries useful information not carried by the others.

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

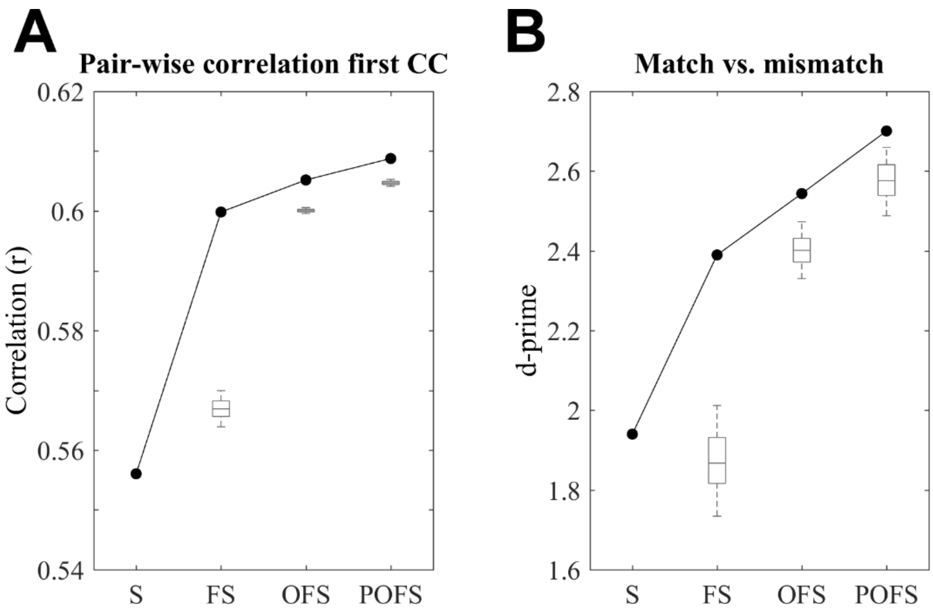

Figure 2: EEG responses to natural speech are best explained when including phonotactic probability among the speech features. Data from all participants were combined using MCCA. This consensus EEG signal (CS) preserves signals that are maximally correlated across subjects. **(A)** A CCA analysis was 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*

= 0.0045, *d* = 1.17; *d-prime*: $p < 0.014$, *d* = 0.85) and, similarly, that the phoneme onsets vector refines the FS representation of speech (OFS > FS; correlations: correlations: $p < 0.0001$, *d* = 1.03; *d-prime*: $p = 0.042$, *d* = 0.65). The average benefits (relative gain) of adding the onset vector O, and the phonotactic vector P, for both measures is plotted in the right-hand panels of **Figure 3A** and **B**. Statistical analysis on these average measures confirms that phonotactic information has a measurable effect on the EEG responses to speech (correlation: $p = 0.0008$, *d* = 1.03; *d-prime*: $p = 0.016$, *d* = 0.65).

Finally, we conducted additional analyses to test whether other models of phonotactic information can explain EEG responses as well, or better, than P. A first single-subject CCA-based analysis compared P to *neighbourhood density* (P_{neigh}). This feature was suggested as a possible neural strategy for an indirect encoding of phonotactic information (Vitevitch et al., 1999; Bailey and Hahn, 2001a). P performed better than this new measure in terms of *d-prime* (POFS > P_{neigh}OFS; one-tailed permutation test: $p = 0.0179$; *d* = 0.68). We performed a similar comparison between P and probabilistic definitions of phoneme surprisal (P_{sur}) and entropy (P_{ent}) (Brodbeck et al., 2018; Gaston and Marantz, 2018). Again, P performed better than these two measures. Specifically, P showed larger *d-prime* values than P_{ent} (POFS > P_{ent}OFS; one-tailed permutation test: $p = 0.037$; *d* = 0.67) and P_{sur} (POFS > P_{sur}OFS; one-tailed permutation test: $p = 0.02$; *d* = 0.73).

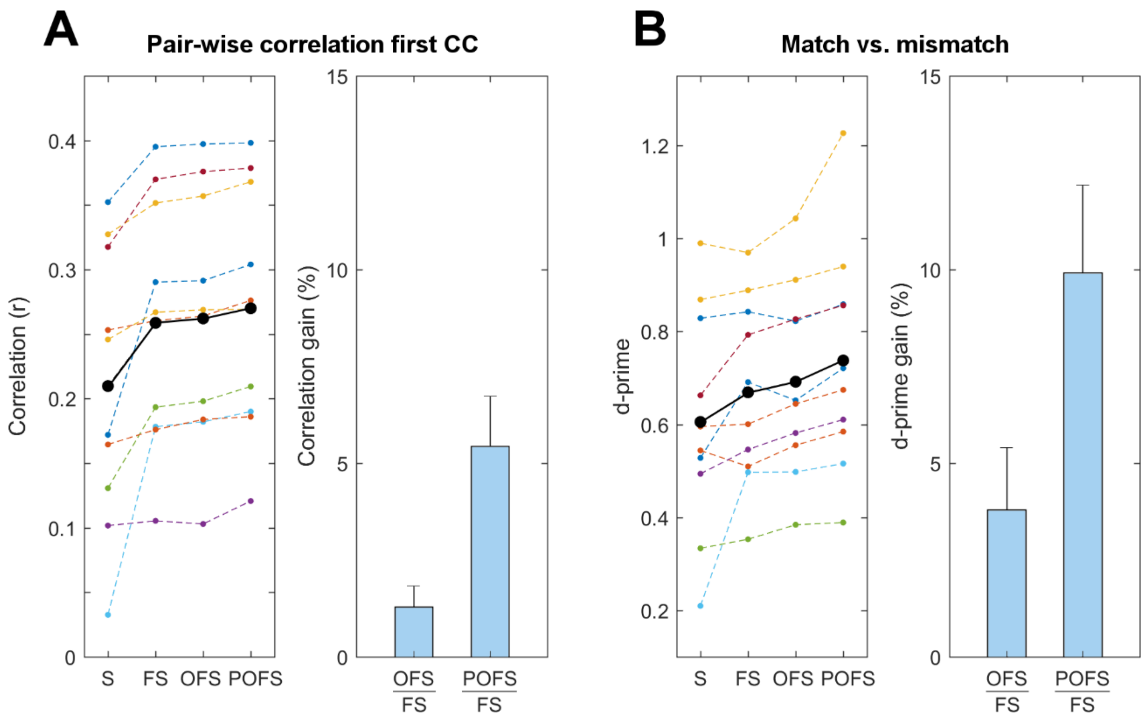

Figure 3: Phonotactic probabilities enhance the speech-EEG mapping at the individual subject level. CCA analyses were conducted between each speech representation and the corresponding EEG responses for each individual subject. **(A)** Speech-EEG correlations for the first canonical component pair were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

Timescale of Cortical Responses to Phonotactics

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

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

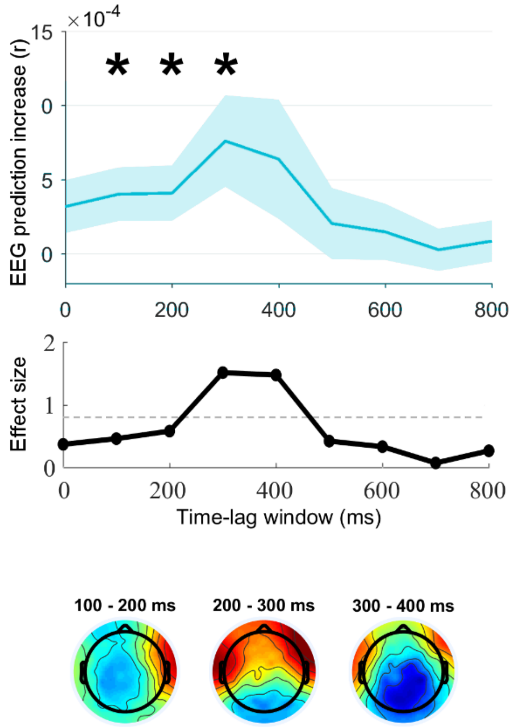

Figure 4: EEG tracking of phonotactic probabilities is specific to speech-brain latencies of 100-400 ms. A temporal response function (TRF) analysis was conducted to estimate the amount of EEG variance explained by phonotactic probabilities for speech-EEG latency windows between 0 and 900 ms and window-size 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 time-lags from 0 to 900 ms are shown for latencies with a significant effect of phonotactic probabilities (100-400 ms) (bottom).

Discussion

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

A novel measure of phonotactic processing

Phonotactic information plays an important role in speech perception. However, crucial questions remain unanswered about the underpinnings of the corresponding cortical processes, mainly due to a lack of tools to extract direct measures of brain responses to phonotactics. Although neurophysiology has partially fulfilled this need (Connolly and Phillips, 1994; Dehaene-Lambertz et al., 2000; Wagner et al., 2012; Cibelli et al., 2015; Leonard et al., 2015), its findings were mainly confined to nonsense words or to the domain of phonotactic violations, which are exceptions in natural speech scenarios. The present study aimed to measure brain signals corresponding to the continuous integration of phonotactic information, which are difficult to isolate when measuring only phonotactic violations. These violations trigger various other processes such as phonological repair, which may emerge in the evoked-response (Dehaene-Lambertz et al., 2000; Dupoux and Pallier, 2001; Domahs et al., 2009). Here, we found evidence that cortical responses to narrative speech reflect the well-formedness of phoneme segments as expressed by probability values, therefore with no (or very few) violations and phonological repair (**Figures 2 and 3**). This finding pushes beyond the phonotactic violation paradigm and provides us with a tool based on linear models to isolate measures of phonotactic-level processing during natural speech perception.

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

Our results provide new insights in this direction, indicating that phonotactic probabilities, as defined by the computational model BLICK, are better represented in the EEG signal than a purely probabilistic definition of phoneme probability (P_{sur}, P_{ent}) (Gaston and Marantz, 2018) and, importantly, than phonological neighbourhood density (P_{neigh}) (Vitevitch et al., 1999; Frisch et al., 2000; Bailey and Hahn, 2001a). While further studies could explore other hypotheses on the encoding and processing of phonotactic information more comprehensively, the present finding is in line with research suggesting distinct roles for phonotactics and neighbourhood density (Vitevitch et al., 1999; Bailey and Hahn, 2001a; Storkel et al., 2006). Specifically, the first would aid speech perception by facilitating processing and triggering learning of new words at early pre-lexical stages, while the latter would influence the integration of new and existing lexical representations at a later stage.

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

Theoretical implications of a rapid time-locked response to phonotactics

Our results have important implications for current theories on phonotactics, by providing insights into both temporal dynamics (when) and neural encoding (how) of this cortical mechanism. Phonotactic information, which aids speech recognition and learning of new words (Mattys and Jusczyk, 2001; Munz, 2017), was suggested to involve one of the following: 1) the phoneme identification stage (one-step models; Dehaene-Lambertz et al., 2000; Dupoux et al., 2011); 2) a pre-lexical stage that occurs after phoneme identification (two-step models; Church, 1987); or 3) a later lexical stage that influences pre-lexical processes through feedback connections (lexicalist models; McClelland et al., 2006; McClelland and Elman, 1986). In this context, a large body of literature in psycholinguistics supports a pre-lexical account of phonotactics (McQueen, 1998; Jusczyk et al., 1999; Sebastián-Gallés, 2007). For example, infants showed sensitivity to phonotactics by 9 months of age, suggesting that this information aids speech segmentation even at early developmental stages, before being able to understand speech (Jusczyk et al., 1994). Similarly, it was shown that humans are sensitive to phonotactic information even when meaning is not involved (nonsense words), pointing to the early implementation of phonotactic repair (Dupoux et al., 1999; Davidson, 2011; Rossi et al., 2013). This indirect evidence for a pre-lexical influence of phonotactic information finds experimental support in both phonotactic violation studies (Dehaene-Lambertz et al., 2000; Pylkkänen et al., 2002) and in the present work, which isolated cortical responses to probabilistic phonotactics showing short speech-EEG latencies (100-400 ms).

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

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

Author Contributions

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

Acknowledgements

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

References

Bailey TM, Hahn U (2001a) Determinants of Wordlikeness: Phonotactics or Lexical Neighborhoods? J Mem Lang 44:568–591 Available at: http://linkinghub.elsevier.com/retrieve/pii/S0749596X00927566 [Accessed May 31, 2018].

Bailey TM, Hahn U (2001b) Determinants of Wordlikeness: Phonotactics or Lexical Neighborhoods? J Mem Lang 44:568–591 Available at: http://www.academicpress.com [Accessed June 1, 2018].

Balling LW, Baayen RH (2012) Probability and surprisal in auditory comprehension of morphologically complex words. Cognition 125:80–106 Available at: http://www.ncbi.nlm.nih.gov/pubmed/22841290 [Accessed May 30, 2018].

Brent MR, Cartwright TA (1996) Distributional regularity and phonotactic constraints are useful for segmentation. Cognition 61:93–125 Available at: https://www.sciencedirect.com/science/article/pii/S0010027796007196 [Accessed June 6, 2018].

Brodbeck C, Hong LE, Simon JZ (2018) Transformation from auditory to linguistic representations across auditory cortex is rapid and attention dependent for continuous speech. bioRxiv:326785 Available at: https://www.biorxiv.org/content/early/2018/05/21/326785.1 [Accessed June 1, 2018].

Broderick MP, Anderson AJ, Di Liberto GM, Crosse MJ, Lalor EC (2018) Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech. Curr Biol.

Carlson MT, Goldrick M, Blasingame M, Fink A (2016) Navigating conflicting phonotactic constraints in bilingual speech perception. Bilingualism 19:939–954 Available at: http://www.journals.cambridge.org/abstract_S1366728915000334 [Accessed June 1, 2018].

Chomsky N, Halle M (1968) The sound pattern of English.

Church KW (1987) Phonological parsing and lexical retrieval. Cognition 25:53–69

Available at: http://www.ncbi.nlm.nih.gov/pubmed/3581729 [Accessed June 2, 2018].

Cibelli ES, Leonard MK, Johnson K, Chang EF (2015) The influence of lexical statistics on temporal lobe cortical dynamics during spoken word listening. Brain Lang 147:66–75 Available at: http://www.ncbi.nlm.nih.gov/pubmed/26072003 [Accessed June 8, 2018].

Connolly JF, Phillips NA (1994) Event-Related Potential Components Reflect Phonological and Semantic Processing of the Terminal Word of Spoken Sentences. J Cogn Neurosci 6:256–266 Available at: http://www.ncbi.nlm.nih.gov/pubmed/23964975 [Accessed May 30, 2018].

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

Crosse MJ, Di Liberto GM, Lalor EC (2016b) Eye can hear clearly now: Inverse effectiveness in natural audiovisual speech processing relies on long-term crossmodal temporal integration. J Neurosci 36.

Davidson L (2006a) Phonotactics and articulatory coordination interact in phonology: Evidence from nonnative production. Cogn Sci 30:837–862 Available at: http://doi.wiley.com/10.1207/s15516709cog0000_73 [Accessed June 6, 2018].

Davidson L (2006b) Phonology, phonetics, or frequency: Influences on the production of non-native sequences. J Phon 34:104–137 Available at: http://linkinghub.elsevier.com/retrieve/pii/S0095447005000240 [Accessed June 6, 2018].

Davidson L (2011) Phonetic, Phonemic, and Phonological Factors in Cross-Language Discrimination of Phonotactic Contrasts. J Exp Psychol Hum Percept Perform 37:270–282 Available at: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0020988 [Accessed June 2, 2018].

de Cheveigné A, Di Liberto GM, Arzounian D, Wong D, Hjortkjaer J, Fuglsang SA, Parra LC (2018a) Multiway Canonical Correlation Analysis of Brain Signals. bioRxiv:344960 Available at:

https://www.biorxiv.org/content/early/2018/06/12/344960 [Accessed June 13, 2018].

de Cheveigné A, Wong DE, Di Liberto GM, Hjortkjær J, Slaney M, Lalor E (2018b) Decoding the auditory brain with canonical component analysis. Neuroimage 172:206–216 Available at: https://www.sciencedirect.com/science/article/pii/S1053811918300338 [Accessed June 1, 2018].

Dehaene-Lambertz G, Dupoux E, Gout A (2000) Electrophysiological correlates of phonological processing: a cross-linguistic study. J Cogn Neurosci 12:635–647 Available at: http://www.unicog.org/publications/Dehaene-LambertzDupoux_EbuzERP_JCogNS2000.pdf [Accessed May 30, 2018].

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9–21.

Di Liberto GM, Crosse MJ, Lalor EC (2018a) Cortical Measures of Phoneme-Level Speech Encoding Correlate with the Perceived Clarity of Natural Speech. Eneuro 5:ENEURO.0084-18.2018 Available at: http://eneuro.sfn.org/lookup/doi/10.1523/ENEURO.0084-18.2018.

Di Liberto GM, Lalor EC (2017) Indexing cortical entrainment to natural speech at the phonemic level: Methodological considerations for applied research. Hear Res 348:70–77.

Di Liberto GM, Lalor EC, Millman RE (2018b) Causal cortical dynamics of a predictive enhancement of speech intelligibility. Neuroimage 166.

Di Liberto GM, O'Sullivan JA, Lalor EC (2015) Low-frequency cortical entrainment to speech reflects phoneme-level processing. Curr Biol 25.

Di Liberto GM, Peter V, Kalashnikova M, Goswami U, Burnham D, Lalor EC (2018c) Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia. Neuroimage NIMG-17-29:70–79 Available at: https://www.sciencedirect.com/science/article/pii/S1053811918302829 [Accessed June 1, 2018].

Ding N, Melloni L, Yang A, Wang Y, Zhang W, Poeppel D (2017) Characterizing Neural Entrainment to Hierarchical Linguistic Units using Electroencephalography (EEG). Front Hum Neurosci 11:481 Available at: http://journal.frontiersin.org/article/10.3389/fnhum.2017.00481/full [Accessed December 30, 2017].

Ding N, Melloni L, Zhang H, Tian X, Poeppel D (2015) Cortical tracking of hierarchical linguistic structures in connected speech. Nat Neurosci 19:158–164 Available at: http://www.nature.com/doifinder/10.1038/nn.4186 [Accessed December 30, 2017].

Ding N, Simon JZ (2012) Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. J Neurophysiol 107:78–89 Available at: http://jn.physiology.org/content/jn/107/1/78.full.pdf.

Ding N, Simon JZ (2014) Cortical Entrainment to Continuous Speech: Functional Roles and Interpretations. Front Hum Neurosci 8 Available at: http://www.frontiersin.org/Journal/Abstract.aspx?s=537&name=human_neuroscience&ART_DOI=10.3389/fnhum.2014.00311.

Domahs U, Kehrein W, Knaus J, Wiese R, Schlesewsky M (2009) Event-related Potentials Reflecting the Processing of Phonological Constraint Violations. Lang Speech 52:415–435 Available at: http://journals.sagepub.com/doi/10.1177/0023830909336581 [Accessed June 2, 2018].

Dupoux E, Kakehi K, Hirose Y, Pallier C, Mehler J (1999) Epenthetic vowels in Japanese: A perceptual illusion? J Exp Psychol Hum Percept Perform 25:1568–1578 Available at: http://cogprints.org/747/5/ebuzo.pdf [Accessed May 30, 2018].

Dupoux E, Pallier C (2001) New evidence for prelexical phonological processing in word recognition. Lang Cogn Process 16:491–505 Available at: http://www.lscp.net/persons/dupoux/papers/Dupoux_PKM_2001_Prelexical_epenthesis_LangCogProc.pdf [Accessed June 1, 2018].

Dupoux E, Parlato E, Frota S, Hirose Y, Peperkamp S (2011) Where do illusory vowels come from? J Mem Lang 64:199–210 Available at:

http://www.lscp.net/persons/peperkamp/Dupoux_Parlato_Frota_Hirose_Peperkamp_(2011)_Where_do_illusory_vowels_come_from.pdf [Accessed June 2, 2018].

Ettinger A, Linzen T, Marantz A (2014) The role of morphology in phoneme prediction: Evidence from MEG. Available at: http://ling.umd.edu/assets/publications/Ettinger-Linzen-Marantz-14-MorphologyInPhonemePrediction.pdf [Accessed May 30, 2018].

Frisch SA, Large NR, Pisoni DB (2000) Perception of Wordlikeness: Effects of Segment Probability and Length on the Processing of Nonwords. J Mem Lang 42:481–496 Available at: http://cas.usf.edu/~frisch/Frisch_Large_Pisoni_00.pdf [Accessed May 18, 2018].

Gaston P, Marantz A (2018) The time course of contextual cohort effects in auditory processing of category-ambiguous words: MEG evidence for a single “clash” as noun or verb. Lang Cogn Neurosci 33:402–423 Available at: https://www.tandfonline.com/doi/full/10.1080/23273798.2017.1395466 [Accessed June 1, 2018].

Goldwater S, Johnson M (2003) Learning OT Constraint Rankings Using a Maximum Entropy Model. Proc Stock Work Var within Optim Theory:111–120 Available at: http://homepages.inf.ed.ac.uk/sgwater/papers/OTvar03.pdf [Accessed June 1, 2018].

Gorman K, Howell J, Wagner M (2011) Prosodylab-aligner: A tool for forced alignment of laboratory speech. Can Acoust - Acoust Can 39:192–193 Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84859480980&partnerID=40&md5=6d828c3caa30dd5c98fdf421f5a0b762.

Greenwood DD (1961) Auditory Masking and the Critical Band. J Acoust Soc Am 33:484–502 Available at: http://scitation.aip.org/content/asa/journal/jasa/33/4/10.1121/1.1908699.

Hallé PA, Dominguez A, Cuetos F, Segui J (2008) Phonological mediation in visual masked priming: Evidence from phonotactic repair. J Exp Psychol Hum Percept Perform 34:177–192 Available at: http://doi.apa.org/getdoi.cfm?doi=10.1037/0096-1523.34.1.177 [Accessed June 1, 2018].

Hammond M (2004) Gradience, Phonotactics, and the Lexicon in English Phonology. Int J English Stud 4 Available at: http://roa.rutgers.edu/files/736-0505/736-HAMMOND-0-0.PDF [Accessed May 30, 2018].

Hayes B (2012) BLICK : a phonotactic probability calculator (manual). Available at: http://linguistics.ucla.edu/people/hayes/BLICK/BLICKManual.pdf [Accessed June 1, 2018].

Hayes B, Wilson C (2008) A Maximum Entropy Model of Phonotactics and Phonotactic Learning. Linguist Inq 39:379–440 Available at: http://linguistics.ucla.edu/people/hayes/papers/HayesAndWilsonPhonotactics2008.pdf [Accessed June 1, 2018].

Hotelling H (1936) Relations Between Two Sets of Variates. Biometrika 28:321 Available at: https://www.jstor.org/stable/2333955?origin=crossref [Accessed June 1, 2018].

Jusczyk PW, Houston DM, Newsome M (1999) The Beginnings of Word Segmentation in English-Learning Infants. Cogn Psychol 39:159–207 Available at: https://www.sciencedirect.com/science/article/pii/S0010028599907168 [Accessed June 2, 2018].

Jusczyk PW, Luce PA, Charles-Luce J (1994) Infants′ Sensitivity to Phonotactic Patterns in the Native Language. J Mem Lang 33:630–645 Available at: https://www.sciencedirect.com/science/article/pii/S0749596X84710308 [Accessed June 2, 2018].

Khalighinejad B, Cruzatto da Silva G, Mesgarani N (2017) Dynamic Encoding of Acoustic Features in Neural Responses to Continuous Speech. J Neurosci.

Kösem A, van Wassenhove V (2016) Distinct contributions of low- and high-frequency neural oscillations to speech comprehension. Lang Cogn Neurosci:1–9 Available at: http://dx.doi.org/10.1080/23273798.2016.1238495.

Kutas M, Federmeier KD (2011) Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu Rev Psychol

62:621–647 Available at: http://www.ncbi.nlm.nih.gov/pubmed/20809790 [Accessed December 30, 2017].

Lalor EC, Pearlmutter BA, Reilly RB, McDarby G, Foxe JJ (2006) The VESPA: a method for the rapid estimation of a visual evoked potential. Neuroimage 32:1549–1561 Available at: http://ac.els-cdn.com/S1053811906006434/1-s2.0-S1053811906006434-main.pdf?_tid=ff77a230-a642-11e4-8c11-00000aacb35e&acdnat=1422376921_3d415c577602776e8b0c6a9e8425ac29.

Lalor EC, Power AJ, Reilly RB, Foxe JJ (2009) Resolving Precise Temporal Processing Properties of the Auditory System Using Continuous Stimuli. J Neurophysiol 102:349–359 Available at: http://jn.physiology.org/content/jn/102/1/349.full.pdf.

Lentz TO, Kager RWJ (2015) Categorical phonotactic knowledge filters second language input, but probabilistic phonotactic knowledge can still be acquired. Lang Speech 58:387–413 Available at: http://www.ncbi.nlm.nih.gov/pubmed/26529903 [Accessed June 6, 2018].

Leonard MK, Bouchard KE, Tang C, Chang EF (2015) Dynamic Encoding of Speech Sequence Probability in Human Temporal Cortex. J Neurosci 35:7203–7214 Available at: http://www.ncbi.nlm.nih.gov/pubmed/25948269 [Accessed June 8, 2018].

Luck SJ (2005) An introduction to the event-related potential technique.

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

Mattys SL, Jusczyk PW (2001) Phonotactic cues for segmentation of fluent speech by infants. Cognition 78:91–121 Available at: http://www.ncbi.nlm.nih.gov/pubmed/11074247 [Accessed June 2, 2018].

Mattys SL, Jusczyk PW, Luce PA, Morgan JL (1999) Phonotactic and Prosodic Effects on Word Segmentation in Infants. Cogn Psychol 38:465–494 Available at: http://www.ncbi.nlm.nih.gov/pubmed/10334878 [Accessed May 30, 2018].

McClelland JL, Elman JL (1986) The TRACE model of speech perception. Cogn Psychol 18:1–86.

McClelland JL, Mirman D, Holt LL (2006) Are there interactive processes in speech perception? Trends Cogn Sci 10:363–369.

McQueen JM (1998) Segmentation of continuous speech using phonotactics. J Mem Lang 39:21–46 Available at: https://www.sciencedirect.com/science/article/pii/S0749596X98925682 [Accessed June 2, 2018].

Mesgarani N, Cheung C, Johnson K, Chang EF (2014) Phonetic Feature Encoding in Human Superior Temporal Gyrus. Science (80-) 343:1006–1010 Available at: http://www.sciencemag.org/content/343/6174/1006.abstract.

Munz ED (2017) Psychotherapie in der Psychiatrie. Nervenheilkunde 36:800–805 Available at: https://www.era.lib.ed.ac.uk/handle/1842/10432 [Accessed June 2, 2018].

O'Sullivan JA, Power AJ, Mesgarani N, Rajaram S, Foxe JJ, Shinn-Cunningham BG, Slaney M, Shamma SA, Lalor EC (2014) Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. Cereb Cortex:bht355.

Obleser J, Herrmann B, Henry MJ (2012) Neural Oscillations in Speech: Don't be Enslaved by the Envelope. Front Hum Neurosci 6:250 Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3431501/.

Obrig H, Mentzel J, Rossi S (2016) Universal and language-specific sublexical cues in speech perception: a novel electroencephalography-lesion approach. Brain 139:1800–1816 Available at: http://www.ncbi.nlm.nih.gov/pubmed/27190021 [Accessed May 30, 2018].

Parker SG (Stephen G (2012) The sonority controversy. De Gruyter Mouton. Available at: https://books.google.fr/books?id=ixpO4NZD2gkC&dq=Parker,+S.+(Ed.).+(2012).+The+Sonority+Controversy&lr=&hl=it&source=gbs_navlinks_s [Accessed May 18, 2018].

Pisoni DB, Remez RE (2005) The handbook of speech perception. Blackwell Pub. Available at: https://books.google.fr/books?id=EwY15naRiFgC&pg=PA619&lpg=PA619&dq=t

race+model+phonotactic&source=bl&ots=0OXeu89-4S&sig=EWPP5YRtV4Odmn1E1aRk0YzJUEY&hl=it&sa=X&ved=0ahUKEwiAuurYqL_bAhWIbRQKHeFICy8Q6AEIWzAG#v=onepage&q=trace model phonotactic&f=false [Accessed June 6, 2018].

Pylkkänen L, Stringfellow A, Flagg E, Marantz A (2000) A neural response sensitive to repetition and phonotactic probability : MEG investigations of lexical access. Proc Biomag 2000, 12th Int Conf Biomagn:1–4 Available at: https://pdfs.semanticscholar.org/5d61/dcc9f304711f79ff230cee855855c149eec5.pdf [Accessed June 1, 2018].

Pylkkänen L, Stringfellow A, Marantz A (2002) Neuromagnetic Evidence for the Timing of Lexical Activation: An MEG Component Sensitive to Phonotactic Probability but Not to Neighborhood Density. Brain Lang 81:666–678 Available at: https://www.sciencedirect.com/science/article/pii/S0093934X01925556 [Accessed June 1, 2018].

Rossi S, Hartmüller T, Vignotto M, Obrig H (2013) Electrophysiological evidence for modulation of lexical processing after repetitive exposure to foreign phonotactic rules. Brain Lang 127:404–414 Available at: http://www.ncbi.nlm.nih.gov/pubmed/23489581 [Accessed June 2, 2018].

Rossi S, Jürgenson IB, Hanulíková A, Telkemeyer S, Wartenburger I, Obrig H (2011) Implicit Processing of Phonotactic Cues: Evidence from Electrophysiological and Vascular Responses. J Cogn Neurosci 23:1752–1764 Available at: http://www.ncbi.nlm.nih.gov/pubmed/20666594 [Accessed May 30, 2018].

Scholes RJ (1966) Phonotactic Grammaticality. Hague Mout Co.

Sebastián-Gallés N (2007) Biased to learn language. Dev Sci 10:713–718 Available at: http://doi.wiley.com/10.1111/j.1467-7687.2007.00649.x [Accessed June 2, 2018].

Storkel HL (2001) Learning New Words. J Speech Lang Hear Res 44:1321 Available at: http://jslhr.pubs.asha.org/article.aspx?doi=10.1044/1092-4388(2001/103) [Accessed December 30, 2017].

Storkel HL (2004) The emerging lexicon of children with phonological delays: phonotactic constraints and probability in acquisition. J Speech Lang Hear Res

47:1194–1212 Available at: http://www.ncbi.nlm.nih.gov/pubmed/15603471 [Accessed May 31, 2018].

Storkel HL, Armbrüster J, Hogan TP (2006) Differentiating Phonotactic Probability and Neighborhood Density in Adult Word Learning. J Speech Lang Hear Res 49:1175 Available at: http://www.ncbi.nlm.nih.gov/pubmed/17197489 [Accessed June 1, 2018].

Storkel HL, Morrisette ML (2002) The Lexicon and Phonology. Lang Speech Hear Serv Sch 33:24 Available at: http://www.ncbi.nlm.nih.gov/pubmed/27764412 [Accessed May 31, 2018].

Storkel HL, Rogers MA (2000) The effect of probabilistic phonotactics on lexical acquistion. Clin Linguist Phon:407–425 Available at: https://wordlearning.ku.edu/storkel-hl-rogers-ma-2000 [Accessed May 31, 2018].

Vanthornhout J, Decruy L, Wouters J, Simon JZ, Francart T (2018) Speech Intelligibility Predicted from Neural Entrainment of the Speech Envelope. J Assoc Res Otolaryngol 19:181–191 Available at: http://link.springer.com/10.1007/s10162-018-0654-z [Accessed June 11, 2018].

Vitevitch MS, Luce PA, Pisoni DB, Auer ET (1999) Phonotactics, neighborhood activation, and lexical access for spoken words. Brain Lang 68:306–311 Available at: http://www.ncbi.nlm.nih.gov/pubmed/10433774 [Accessed May 30, 2018].

Wagner M, Shafer VL, Martin B, Steinschneider M (2012) The phonotactic influence on the perception of a consonant cluster /pt/ by native English and native Polish listeners: A behavioral and event related potential (ERP) study. Brain Lang 123:30–41 Available at: http://www.ncbi.nlm.nih.gov/pubmed/22867752 [Accessed June 11, 2018].

White J, Chiu F (2017) Disentangling phonological well-formedness and attestedness: An ERP study of onset clusters in English. Acta Linguist Acad 64:513–537 Available at: http://www.akademiai.com/doi/10.1556/2062.2017.64.4.2 [Accessed June 1, 2018].

Wiese R, Orzechowska P, Alday PM, Ulbrich C (2017) Structural Principles or Frequency of Use? An ERP Experiment on the Learnability of Consonant Clusters.

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

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

Figures

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

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).

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

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

Extended data

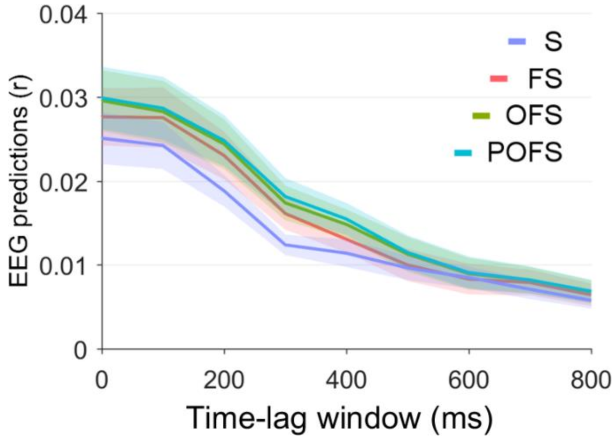

Figure 4-1. A temporal response function (TRF) analysis was conducted to estimate the amount of EEG variance explained by phonotactic probabilities for speech-EEG latency windows between 0 and 900 ms and window-size 100 ms. EEG prediction correlations averaged across all scalp electrodes are shown for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