Low-frequency cortical responses to natural speech reflect probabilistic phonotactics

Giovanni M Di Liberto, Daniel Wong, Gerda Ana Melnik, Alain de Cheveigné

NeuroImage 196 (2019) 237–247

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Low-frequency cortical responses to natural speech reflect probabilistic phonotactics

Giovanni M. Di Liberto [a,b,*], Daniel Wong [a,b], Gerda Ana Melnik [b,c], Alain de Cheveigné [a,b,d]

[a] Laboratoire des Systèmes Perceptifs, UMR 8248, CNRS, France
[b] Département d'Etudes Cognitives, Ecole Normale Supérieure, PSL University, France
[c] Laboratoire de Sciences Cognitives et Psycholinguistique, ENS, EHESS, CNRS, France
[d] UCL Ear Institute, London, United Kingdom

ARTICLE INFO

Keywords:
cortical tracking
EEG
Language
Neighbourhood density
Phonemes

ABSTRACT

Humans comprehend speech despite the various challenges such as mispronunciation and noisy environments. Our auditory system is robust to these thanks to the integration of the 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 demonstrated that violations of these rules modulate brain-evoked responses. However, several fundamental questions remain unresolved, especially regarding the neural encoding and integration strategy of phonotactics in naturalistic conditions, when there are no (or few) violations. Here, we used linear modelling 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 stimulus descriptor includes phonotactic probabilities. This indicates that low-frequency cortical signals (<9 Hz) reflect the integration of phonotactic information during natural speech perception, providing us with a measure of phonotactic processing at the individual subject-level. Furthermore, phonotactics-related signals showed the strongest speech-EEG interactions at latencies of 100–500 ms, supporting a pre-lexical role of phonotactic information.

1. 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). A numerical value can be assigned to a given sequence of phonemes describing its likelihood of belonging to the language, meaning that well-formedness is a graded value (Scholes, 1966; Chomsky and Halle, 1968; Frisch et al., 2000; Bailey and Hahn, 2001; Hammond, 2004). This concept has been linked to our ability to learn language-specific regularities and constraints for phoneme sequences, which are referred to as *phonotactics* (Zamuner and Kharlamov, 2016). 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). Despite the compelling evidence for a role of phonotactic information in speech processing (Vitevitch et al., 1997; Ettinger et al., 2014; Leonard et al., 2015), 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). Hypotheses range from the explicit encoding of phonotactic probabilities to the use of the lexical neighbourhood size as a proxy measure (McClelland and Elman, 1986; Bailey and Hahn, 2001; Pisoni and Remez, 2005; Leonard et al., 2015). Furthermore, part of the debate regards the pre- or post-lexical role of phonotactics, but this question is

[*] Corresponding author. Laboratoire des Systèmes Perceptifs, 29 rue d’Ulm, 75005 Paris. France.
E-mail address: diliberg@tcd.ie (G.M. Di Liberto).

https://doi.org/10.1016/j.neuroimage.2019.04.037
Received 27 October 2018; Received in revised form 18 March 2019; Accepted 11 April 2019
Available online 13 April 2019

hard to resolve given the lack of neural data examining the cortical representation of phonotactic statistics.

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. the phonotactic mismatch response (PMM; Connolly and Phillips, 1994; Dehaene-Lambertz et al., 2000). This paradigm has been largely exploited in the literature, but with results that are sparse and sometimes inconsistent. 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; Ulbrich et al., 2017; White and Chiu, 2017; 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 (nonsense words) and *phonological repair* for ill-formed tokens (non-words) (Bailey and Hahn, 2001; 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 the latter issue, but the paradigm becomes more artificial and one can question whether the cortical mechanisms underlying phonological processing of such stimuli are equivalent to those employed to process natural speech.

Here we use natural speech together with novel stimulus-response modelling techniques (Ding and Simon, 2014). Recent research demonstrated that linear modelling approaches, such as the temporal response function (Lalor et al., 2009; Crosse et al., 2016) and canonical correlation analysis (de Cheveigné et al., 2018), allow to isolate and measure cortical responses to linguistic features of interest (e.g. phonemes) using natural speech stimuli (Di Liberto et al., 2015; Di Liberto, Crosse et al., 2018). Natural speech is generally characterised by well-formed words, thus we cannot rely on responses to ill-formed tokens as in PMM studies. Instead, we use a computational model of phonotactics based on phonological constraints (Hayes and Wilson, 2008) to estimate a graded measure of well-formedness that fits expert intuitions. Then, we test whether these values are reflected in the time-locked brain responses elicited by narrative speech. We characterise the dynamics of cortical signals that are elicited by natural speech stimuli, contributing to the debate on the underpinnings of the cortical processes specific to phonotactics.

2. Materials 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; Broderick et al., 2018). 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 several multiple 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. Both the EEG data and audio are available at https://datadryad.org/resource/%3ca href=*title = "https://datadryad.org/resource/*https://doi.org/10.5061/dryad*.070jc">https://datadryad.org/resource/*https://doi.org/10.5061/dryad*.070jc*. The code used for the data analysis can be found at *http://audition.ens.fr/adc/NoiseTools/*and *https://sourceforge.net/projects/aespa/*. This manuscript reports all information regarding measures, conditions, data exclusions, and sample size.

2.1. Subjects and experimental procedure

Nineteen healthy subjects (13 male) aged between 19 and 38 years old participated in the experiment. Participants were native speakers of English and 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 20 trials, each of ~155 s 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 minimise 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 Broderick et al. (2018).

2.2. Speech representations

The analysis approach used here follows a system identification framework (Lalor et al., 2009; Crosse et al., 2016) that aims at disentangling brain responses to different speech and language properties (Di Liberto et al., 2015). To this end, we first need to define such properties (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 8), 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. The time series of phoneme labels (35 phonemes) was recoded as a multivariate time series of 19 binary feature 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. This information is not a linear transformation of F, thus linear models may benefit from the explicit definition of O combined with F.
4. Finally, we propose a feature vector describing *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 nonsense word *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 (expressed as sequences of phonetic features) and a maxent grammar (Goldwater and Johnson,

2003). A training stage on a separate large corpus of English-language data fits the model parameters. Then, given a new phoneme sequence, BLICK produces probabilities describing the likelihood that a sequence belongs to the language. This feature of the stimulus has been shown to optimally match the well-formedness intuition of native speakers (see *phonotactic probability model* section), thus we expected it to be reflected in the cortical responses to speech.

Given a phoneme sequence $p_{1..n}$, where p_i indicates a single phoneme, P is composed of two vectors: a) inverse phonotactic probability (score($p_{1..n}$) is the output of the BLICK software and corresponds to –log(L($p_{1..n}$)), with L($p_{1..n}$) indicating the likelihood that a token belongs to the language; as such, it is small for well-formed tokens and large for ill-formed ones) and b) within-word derivative of the phonotactic score (score($p_{1..(n-1)}$) - score($p_{1..n}$)), 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 representations F, O, and P to the speech-EEG mapping, the main analyses were conducted on each of the combinations S, FS, OFS, and POFS. The rationale is that, if we extend the speech representation by including information not subsumed by other features, this will improve the fitting score. As an additional test of the value of that information, we also reran this analysis on data where the information was shuffled, thus with the same dimensionality but with additional information that did not correspond to the auditory stimulus (the entire procedure, including model fit, was rerun for each shuffled version). Shuffled vectors/matrices were generated by randomly shuffling: a) Phonetic features (F_{shu}): in every occurrence of a phoneme, coded as a combination of N_F phonetic features (1 for vowels and 3 for consonants) was replaced by the same number of phonetic features chosen randomly for its entire duration; b) Onset time (O_{shu}): the onset vector O was replaced by a vector with the same number of impulses at random time points; and c) Phonotactic probability values (P_{shu}): the values in the phonotactic vector P were randomly permuted while keeping the time information. Shuffling disrupts the value of the feature as a predictor of the EEG response, and the distribution of outcomes of the shuffled analysis thus serves as a baseline against which to judge the significance of analysis results obtained with unshuffled data.

In addition to the phonotactic vector P, we defined three alternative phonotactic vectors that could reflect the encoding of phonotactic information in the brain. First, P_{neigh} is a vector of pulses indicating phoneme onsets with amplitudes determined by the *neighbourhood density*. 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).

2.3. 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, which are expressed as sequences of phonetic features. Weights are associated to each constraint and their weighted sum estimates a penalty score indicating the inverse "goodness" of that token. Specifically, the likelihood L that a token belongs to the language is modelled as $L = e^{-score}$, where X is the score produced by the model. The weights of the phonotactic constraints are determined according to the principle of maximum entropy and, in the present work, were pre-assigned (E.T. Jaynes, 1988; Grendár, 2001; Hayes, 2012) by means of a training stage on a separate large corpus of data (extracted from the CMU Pronouncing Dictionary; *http://www.speech.cs.cmu.edu*). The choice of weights minimised the probability of words that are not plausible tokens for the language according to the set of constraints. The estimated probabilities were shown to accurately match the phonological well-formedness intuition of native speakers ($r = 0.946$; Hayes and Wilson, 2008). This pre-trained grammar was used on the textual transcription of the audio-book stimulus to calculate probability values estimating the well-formedness of each speech token. Given a word, two scores were calculated for each phoneme token. The first indicates the penalty score (inverse log 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 distribution of the inverse probabilities is shown in Supplementary Fig. 1. 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 first 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.

2.4. 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).

2.5. Dimensionality reduction

The analyses that follow involve fitting the stimulus representation to the EEG response using linear models. 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 minimise this redundancy and reduce the 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 PC is the linear transform of the data with the largest possible variance. The second PC 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. A similar dimensionality-reduction procedure is applied at multiple stages of 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.

2.6. 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) to the case of multiple (>2) datasets. Given N multichannel datasets Y_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., 2018). These patterns can be derived by summing the transformed data matrices: $Y = \sum_{i=1}^{N} Y_i W_i$. The columns of the matrix Y, which are mutually orthogonal, are referred to as summary components (SC) (de Cheveigné et al., 2018). Intuitively, the first few components are signals that most strongly reflect the shared information across the several input datasets, thus minimising subject-specific and channel-specific noise. Note that, in the context of MCCA, 'noise' refers to both EEG artifacts (e.g. eye blinks) and EEG response patterns that are not similar across participants. Here, these datasets are EEG responses to a same speech stimulus for 19 subjects.

This technique allows to extract a *consensus* EEG signal that is more reliable than that of any subject. This methodology overcomes limitations of previous studies that attempted to obtain similar consensus responses by averaging data across subjects. Unless such data is coregistered using anatomical information, likely topographical discrepancies between participants may lead to suboptimal averaging and thus loss of information (O'Sullivan et al., 2015; Di Liberto and Lalor, 2017). MCCA accomodates such discrepancies without the need for coregistration. Under the assumption that brain responses to speech share a similar time course 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 spanning the most reliable 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 any remaining irrelevant signals within the retained components will be excluded through the more restrictive CCA analysis that follows.

2.7. Analysis procedure

2.7.1. Stimulus-response model based on two-way 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, for example because it reflects other brain processes, or various forms of noise (e.g. muscle movements). Similarly, certain speech and language properties may have little or no impact on the measured brain responses. Studying the relation between speech and the corresponding EEG responses can 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., 2018). As such, CCA serves both as a denoising procedure and as stimulus-EEG modelling approach.

In its more general definition, given two sets of multichannel data X and Y 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 XW_1 and YW_2, while making the columns of each transformed data matrix mutually uncorrelated. The first pair of canonical components (CC) is the linear combination of X and Y 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 represents the stimulus (both acoustic properties and linguistic abstractions) and Y represents the neural data (either the EEG signal Y_i of a specific subject or the consensus signal Y extracted with MCCA). This basic formulation of CCA can capture the *instantaneous* interaction between stimulus representations and brain response. The basic formulation can readily be extended to convolutional interactions (e.g. delays or filtering) by applying to both stimulus and EEG matrices a set of time shifts. CCA then produces solutions that correspond to optimal finite impulse response filters (FIR) of order n, where n is the number of time shifts. Alternatively, and this is the solution adopted here, the time shifts can be replaced by a bank of filters. CCA then finds optimal filters within this filter space. With the appropriate choice of filterbank, long temporal structures can be captured for both X and Y with a relatively small number of parameters (de Cheveigné et al., 2018), thus capturing the same information but reducing the risk of overfitting and the computation time. 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 each channel of the stimulus representation, yielding a matrix of size $T \times (N_F \times N_{CH})$, and the dimensionality of this filtered representation was reduced by applying PCA and selecting N_{stim} components, as described previously. The same filterbank was also applied to each channel of the MCCA-processed EEG, yielding a matrix of size $T \times (N_{SC} \times N_{CH})$, and the dimensionality of this matrix was reduced by applying PCA and selecting N_{EEG} components. CCA was then applied to these reduced filtered stimulus and EEG data matrices. The CCA models were trained and tested using a leave-one-out nested cross-validation to control for overfitting (cross-validation folds corresponded to trials of the experiment, i.e. parts of the audiobook). For each iteration of the outer cross-validation loop, one fold was held out for testing while an inner 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 inner loop allowed for the tuning of the values N_{stim} and N_{EEG}, while the outer loop controlled for overfitting. The entire process was repeated for a range of time shifts between stimulus and EEG response, to determine a shift that compensates optimally for the latency of neural processing.

2.7.2. Temporal response function analysis

The CCA-based analysis allows to optimally quantify the coupling between a stimulus and the corresponding EEG responses. However, the convolutional steps (filterbank) make it difficult to interpret temporal relations. Therefore, complementary to CCA, the Temporal Response Function (TRF) analysis method was used to compute a channel-specific mapping between each speech feature and the recorded EEG data (Lalor et al., 2006; Ding and Simon, 2012). This method estimates a temporal 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 (Ridge regression), wherein a regularisation parameter was tuned to control for overfitting (Crosse et al., 2016). The simpler nature of the TRF mapping makes the result more easily interpretable than for CCA. Specifically, the spatio-temporal dynamics of the system can be investigated by studying the magnitude of the TRF model weights, which reflects the importance of each feature (channel and time-lag) to the speech-EEG mapping.

This approach is complementary with CCA analysis in that it provides us with detailed insights on the temporal and spatial patterns. The downside is that it does not benefit from the denoising allowed by CCA. For this reason, it is preferable to conduct the analysis on data filtered within the region of greatest SNR (for an example of the effect of EEG filtering on forward TRF models, see Di Liberto et al., 2015). We restricted the analysis to the 0.5–9 Hz frequency band, using separate low- and high-pass fifth-order Butterworth zero-phase filters. The impulse response of the overall band-pass filter is shown in Fig. 4B.

The time course of the EEG response is potentially affected (e.g. smoothing) by filtering (in particular the 0.5 Hz high-pass filter) (VanRullen, 2011; Widmann and Schröger, 2012). However, while TRF components may be smeared, this should not affect the latency of the peaks estimates as the filters were zero-phase (Matlab function *filtfilt*). Moreover, we further controlled for this issue by successfully replicating these results by using different types of filters (Butterworth and Chebyshev) and orders.

2.7.3. 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 (d-prime) in a match-vs-mismatch classification task. A Pearson's correlation coefficient was calculated for each CC-pair. The first CC-pair offers the highest correlation (by construction), but meaningful speech-EEG correlations can arise for an arbitrary number of additional CC pairs. To obtain an aggregate 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 (randomly selected) 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 evaluated on data distinct from that used for the training. This measure, which quantifies how well the cortical signals track the stimulus, will henceforth be referred to as the *Cortical Tracking Index* (CTI). For the main evaluation dataset, the segment duration $T_{DECODER}$ was set to 5 s, which avoided saturation (classification either too easy or too difficult) in both group and single-subject level analyses.

We used the TRF-based analysis to attempt to pinpoint latencies and topographies associated with each feature of the stimulus representation. The TRF analysis involves a lag window of duration 100 ms that was shifted within a 0–1000 ms range (non-overlapping). For each window position, the associated topography was estimated by calculating Pearson's correlation coefficients between the EEG signal and its prediction for each scalp electrode separately. This procedure was repeated for TRF models fit with different stimulus features. Shifting the analysis over time allows us to estimate roughly the latency of cortical processing associated with a particular stimulus feature. Another solution is to investigate the temporal dynamics reflected by CCA models. However, this was impractical in the present application of CCA as the use of a filter-bank rather than time-lags would have complicated the interpretation of the result. Instead, the use of time-lags would have increased the dimensionality and the risk for overfitting, thus potentially reducing the quality of the CCA models (de Cheveigné et al., 2018).

2.8. Statistical analyses

Unless otherwise stated, all statistical analyses were performed using two-tailed permutation tests. For tests involving several contiguous time latencies, false discovery rate (FDR) correction for multiple comparisons was used.

3. Results

Non-invasive EEG signals were recorded from nineteen participants as they listened to an audiobook. We conducted three analyses tackling the questions: 1) Do cortical signals reflect 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?

3.1. Neural evidence for the processing of probabilistic phonotactics

To increase the SNR of the data, brain signals that are common among participants listening to the same speech stimulus were estimated using MCCA (de Cheveigné et al., 2018). This *consensus signal* (CS) has a better signal-to-noise ratio than EEG data of individual subjects. A speech-EEG model based on CCA was then employed to relate this consensus EEG signal to different stimulus representations. The quality of the model (measured by correlation and CTI) was used as a measure of the ability of each feature to capture speech structure predictive of the EEG response. Comparing different representations, a larger CTI (*d-prime*) value indicates that the model is better at matching portions of speech with the corresponding EEG data, suggesting that the EEG signal reflects the speech properties used in that model.

We wish specifically to evaluate the predictive power of phonotactics (P) relative to, and in combination with, other known informations such as spectrogram or phonetic features. For this, we compared multiple combinations of stimulus features, including or not the phonotactic vector P.

We first estimated the quality of a CCA-based model involving only P (Fig. 1A,top) and EEG. The *r*-value of 0.57 obtained for the first CC-pair was larger than the 99th percentile of a distribution obtained by rerunning the CCA analysis after shuffling the values of the pulses within the P vector while leaving their times intact (Fig. 2A; median over 50 shuffles: $r = 0.54$; 99th percentile: $r = 0.55$). This result indicates that phonotactic probabilities were reflected by the EEG signals. However the phonotactic vector is correlated with other predictive properties (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 stimulus representations that include, or not, the phonotactic vector P. We defined a set of descriptors encoding information about the stimulus and linguistic abstractions that we hypothesised are processed during speech perception: 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; Di Liberto, Crosse, et al., 2018), phoneme onsets O (Brodbeck et al., 2018) and our newly introduced phonotactic vector P (see Fig. 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 information in the analysis: namely S, FS, OFS, and POFS as schematised in Fig. 1B. Indeed, correlation coefficient values for CCA models based on these four combination of features agree with this prediction (Fig. 2B; $r_S < r_{FS} < r_{OFS} < r_{POFS}$). 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 (thanks to cross-validation) penalise the models with more dimensions, contrary to what we observe. To further exclude the potential confound of dimensionality, we randomly shuffled the values of the pulses within the phonotactic vectors while keeping their timing constant. The distribution of correlation scores for P_{shu}OFS obtained by repeated shuffling is indicated in Fig. 2B. The value obtained for POFS is above the 99th percentile of that distribution. This same control procedure was applied to the F and O vectors and confirmed that their respective enhancements are driven by the addition of meaningful information as they produced stronger correlations than the 99th percentile of the corresponding shuffled distributions. In summary, each of these properties carries useful information not carried by the others.

The previous analysis was based on correlations for the first CC-pair only. However, some of the other components may also capture relevant stimulus-EEG interactions. To get a more complete picture we performed a similar analysis based on the Cortical Tracking Index (which combines all CC pairs simultaneously; see Methods). These CTI values showed patterns resembling what previously seen for the correlation analysis. Specifically, a *d-prime* of 2.56 resulted from the CCA analysis on P, which was greater than the 99th percentile of the shuffled distribution (Fig. 2C; median over 50 shuffles: *d-prime* = 2.30; 99th percentile: *d-*

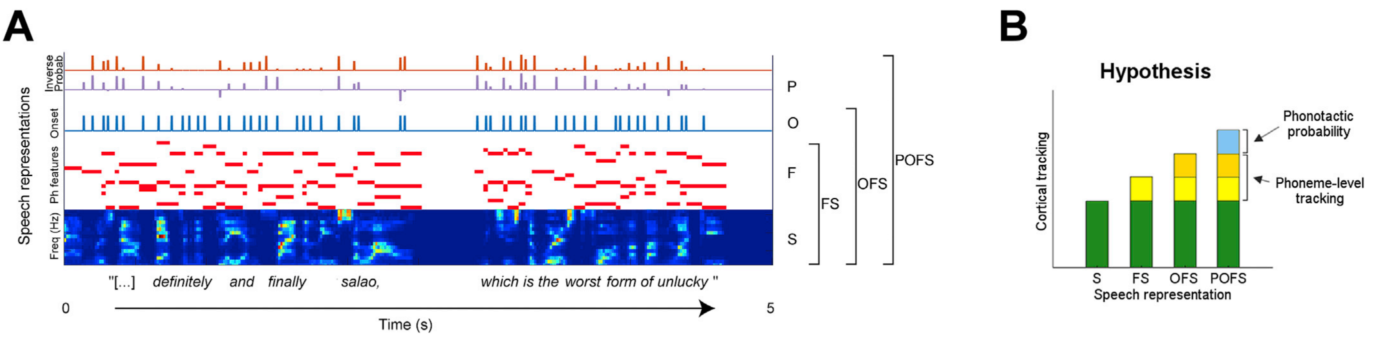

Fig. 1. (A) Speech representations for a 5 s portion of the stimulus. From bottom to top, the acoustic spectrogram (S) which consists of a 16-channel time series of power within 16 frequency bands; phonetic features (F), whose permissible combinations map to English phonemes; phoneme onsets (O), which 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 feature encodes properties not captured by other features, adding it to the others will improve the prediction of cortical responses. In particular we predict an increase in cortical tracking when phonotactic probabilities are added to the mix (POFS – OFS, blue increment).

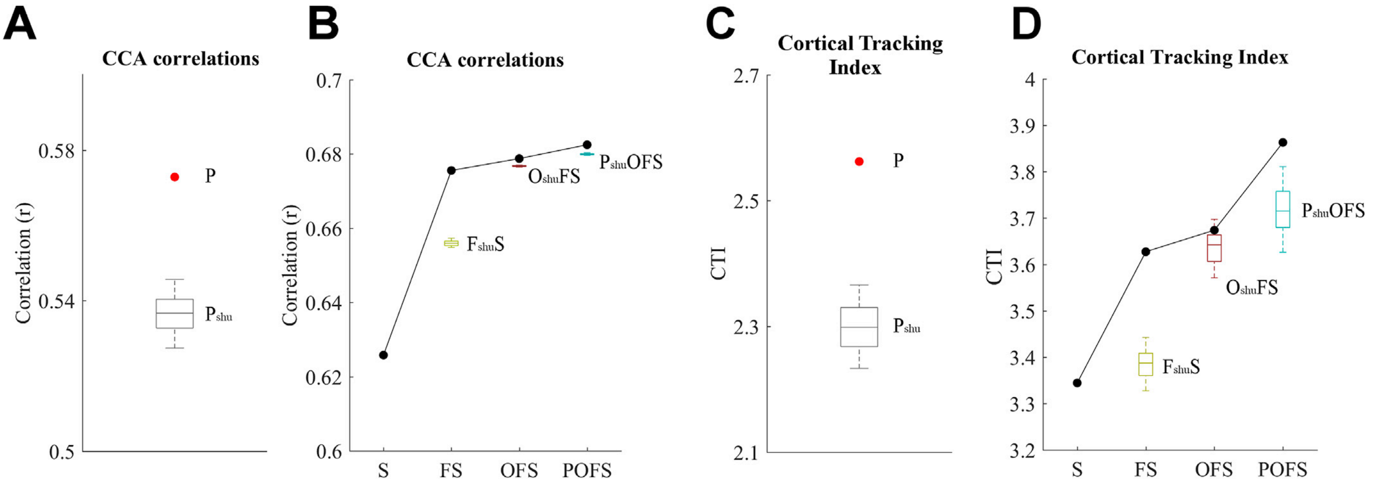

Fig. 2. EEG responses to natural speech are best explained when including phonotactic probability in the stimulus representation. Data from all participants were combined using MCCA. This consensus EEG signal (CS) preserves signals that are maximally correlated across subjects. **(A)** A CCA analysis was conducted between the phonotactic vector (P) and the CS signals. This result was compared to the correlations obtained when using vectors with same time information but shuffled phonotactic values (P_{shu}; 50 versions). Speech-EEG correlations for the first canonical component (CC) pair were best for P than P_{shu}. **(B)** 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). **(C)** The Cortical Tracking Index (CTI; *d-prime* score of a match-vs-mismatch classification test) was larger for P compared to all of its shuffled versions P_{shu}. **(D)** In addition, phonotactic probabilities enhanced the CTI score when combined with acoustic and phoneme-level properties. The box-plots indicate the 99th percentile of the correlation or CTI when using a model (P, FS, OFS, or POFS) after randomly shuffling information for the indicated property (F, O, and P respectively). EEG responses to speech reflect phonotactic probabilities at the individual-subject level.

prime = 2.37). Furthermore, CTI values monotonically increased for S, FS, OFS, and POFS, showing again greater values than the corresponding shuffle distributions (Fig. 2D). The greater value for POFS relative to OFS and P_{shu}OFS reinforces our claim that cortical signals reflect phonotactic probabilities.

The previous analysis provided evidence that the cortical responses to natural speech, measured with non-invasive EEG are coupled to 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. Fig. 3 (left panels of A and B) illustrates both correlation (average correlations for the first five components) and CTI results. The CTI values 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, $p = 0.0006$; $d = 1.00$; POFS > FS, $p = 0.0010$; $d = 0.97$; POFS > S, $p < 0.0001$; $d = 2.24$; FDR-corrected) and *d-prime* (POFS > OFS, $p = 0.0036$; $d = 0.76$; POFS > FS, $p < 0.0001$; $d = 1.33$; POFS > S, $p < 0.0001$; $d = 1.30$; FDR-corrected). In addition, this analysis confirmed that phonetic features explain EEG variance not captured by the acoustic spectrogram (FS > S; correlations: $p < 0.0001$, $d = 2.12$; *d-prime*: $p = 0.0009$, $d = 0.79$; FDR-corrected) and, similarly, that the phoneme onsets vector increase further the amount of EEG variance explained (OFS > FS; correlations: correlations: $p = 0.0190$, $d = 0.58$; *d-prime*: $p = 0.0046$, $d = 0.76$; FDR-corrected). 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 Fig. 3A and B. Statistical analysis on these average measures confirms that phonotactic information has a measurable effect on the EEG responses to speech (correlation: $p = 0.0008$, $d = 0.90$; *d-prime*: $p = 0.0008$, $d = 0.85$).

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 property was suggested as a possible neural strategy for an indirect encoding of phonotactic information

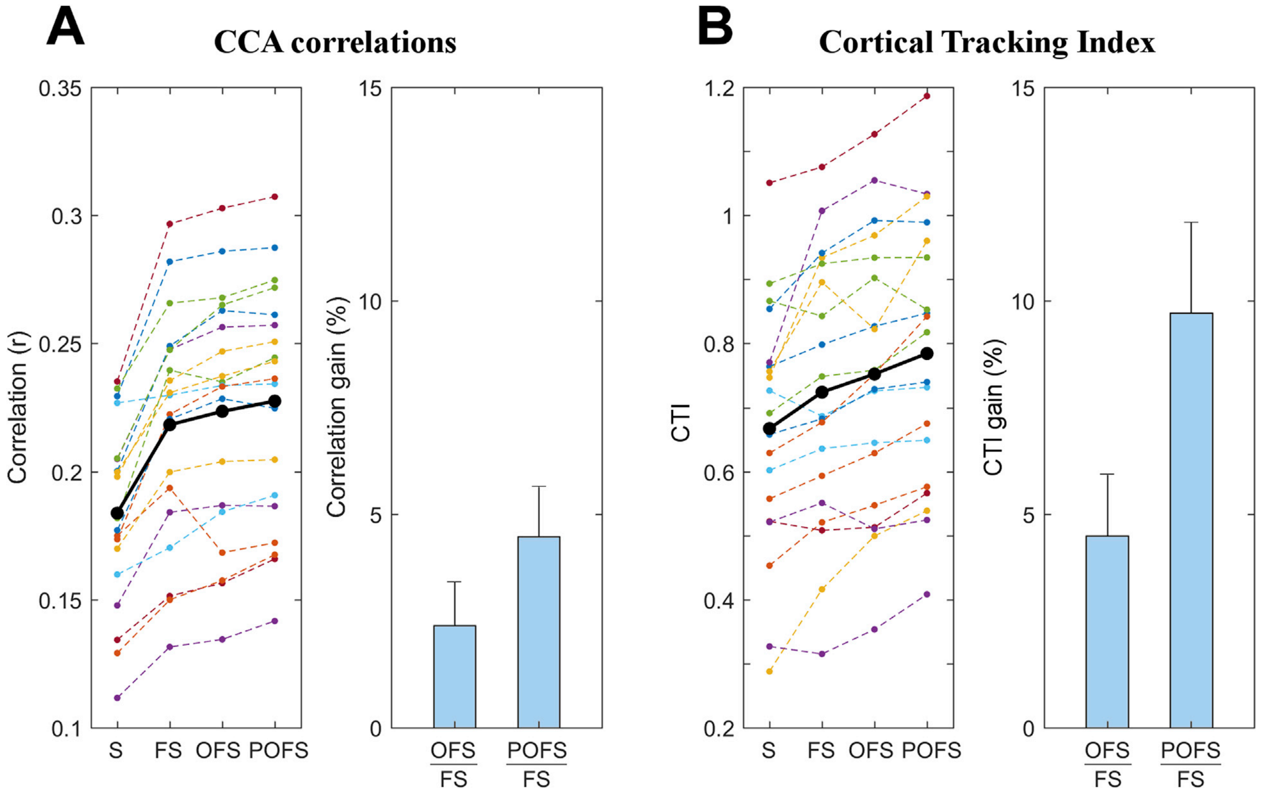

Fig. 3. Phonotactic probabilities enhance the speech-EEG mapping at the individual subject level. CCA analyses were conducted between each speech representation and the corresponding EEG responses for each individual subject. **(A)** Average speech-EEG correlations for the first five canonical component pairs 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 CTI score. 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.

(Vitevitch et al., 1999; Bailey and Hahn, 2001). P had larger CTI values than this new measure (POFS > P_{neigh}OFS; two-tailed permutation test: $p < 0.0001$; $d = 1.19$). We performed a similar comparison between P and probabilistic definitions of phoneme surprisal (P_{sur}) and entropy (P_{ent}) that were derived by means of a completely data-driven approach (Brodbeck et al., 2018; Gaston and Marantz, 2018). Again, P performed better than these two measures. Specifically, P showed larger CTI values than P_{ent} (POFS > P_{ent}OFS; two-tailed permutation test: $p = 0.0150$; $d = 0.62$) and P_{sur} (POFS > P_{sur}OFS; two-tailed permutation test: $p = 0.0031$; $d = 0.69$).

3.2. Timescale of cortical responses to phonotactics

Our results indicate that the language phonotactics is encoded in the cortical signals during natural speech listening. 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 properties 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 coupling of the EEG signal with speech acoustics and phoneme-level properties (Di Liberto et al., 2015; Kösem and van Wassenhove, 2016; Di Liberto, Lalor, et al., 2018; 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 0–1000 ms. Average EEG prediction correlations confirm the hypothesised general trend that emerged also from the CCA analysis (S < FS < OFS < POFS; Supplementary Fig. 2). Crucially, the direct comparison of POFS and OFS reveals a significant effect of phonotactics for speech-EEG latencies between 100 and 500 ms (two-tailed permutation test, FDR corrected, $p < 0.05$), with peak effect-size at the latency-window 300–400 ms ($d = 3.11$) (Fig. 4A). A similar result was observed in the TRF waveforms corresponding to the model weights for P, which show a negative component that peaks at about 300 ms for parietal electrodes (Fig. 4B). Note that the TRF waveforms in Fig. 4B were calculated using the TRF weights for P only, meaning that the result could be influenced also by other EEG responses that correlate to the phonotactic vector. In turn, significant components outside the latencies of interest are likely to be spurious. In particular, the first positive TRF component has both latencies (65 ms) and topographical maps consistent with those of auditory responses. Because of this limitation, we restricted our considerations to the latency interval 100–500 ms.

4. Discussion

Our results indicate that low-frequency 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 100–500 ms are most relevant to those brain responses, suggesting that phonotactic information contributes to natural speech processing at pre-lexical stages.

4.1. A novel measure of phonotactic processing

Phonotactic information has been suggested to play a role in speech perception (Vitevitch et al., 1997; Ettinger et al., 2014; Leonard et al., 2015). However, crucial questions remain unanswered about the underpinnings of the corresponding cortical processes, mainly due to a lack

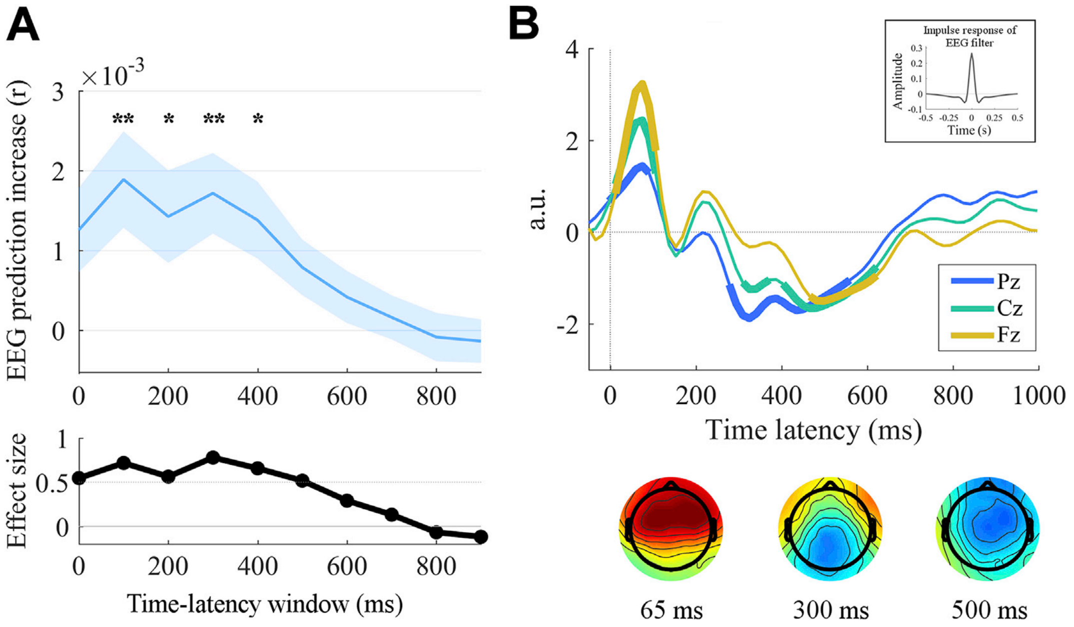

Fig. 4. Robust coupling between EEG signal and phonotactic probabilities at speech-brain latencies of 100–500 ms. (A) A temporal response function (TRF) analysis was conducted to estimate the amount of EEG variance explained by phonotactic probabilities for speech-EEG latency windows between 0 and 1000 ms and window-size 100 ms. EEG prediction correlations were calculated for different speech stimulus representations and for the various speech-EEG latency windows. The enhancement in EEG predictions due to phonotactic probabilities is shown for all latency windows. Shaded areas indicate the standard error of the mean (SE) across subjects. Stars indicate significant enhancement (**$p < 0.01$, *$p < 0.05$; permutation test FDR corrected) (top). Cohen's d was calculated to measure the effect size of the enhancement due to phonotactics. Values above 0.5 are considered as 'medium-large' effects (above dashed grey line) (bottom). **(B)** Grand average TRF waveforms at selected individual channels show the time course of the responses to phonotactics. Thick lines indicate a response that is statistically different from zero across subjects ($p < 0.05$, t-test, FDR corrected). Note that, while panel A isolated the effect of phonotactics, this result is influenced also by lower-level auditory responses. The inset in the top-right corner shows the impulse response of the band-pass filter that was applied to the EEG at the preprocessing stage (top). Topographical patterns of the TRF weights are shown for latencies with significant weights for the phonotactic probability TRF (bottom).

of tools to extract direct measures of brain responses to phonotactics. Although prior studies have partially fulfilled this need (Connolly and Phillips, 1994; Dehaene-Lambertz et al., 2000; Wagner et al., 2012; Cibelli et al., 2015), their 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 reflecting phonotactic information during continuous speech processing, which are difficult to isolate when measuring only phonotactic violations. In fact, those violations trigger various other processes such as phonological repair, which may emerge in the evoked-response (Dehaene-Lambertz et al., 2000; Dupoux et al., 2001; Domahs et al., 2009). Here, we found evidence that low-frequency cortical responses to narrative speech reflect the well-formedness of phoneme segments as expressed by inverse probability values, with no (or very few) violations and thus no need for phonological repair (Figs. 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 using non-invasive EEG.

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; Di Liberto, Peter, et al., 2018), 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 how that information is used to support speech perception, and also how this mechanism plays against us. For example, when learning a second language, phonotactic priors of the native language can 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 vs. others).

Our results provide new insights in this direction, indicating the EEG signal reflects more strongly phonotactic probabilities when linguistic constraints are included in their estimation (using the computational model BLICK). The present finding indicates that cortical signals reflect a process matching the sensory input with expectation at the phonemic time-scale, which is in line with previous work suggesting that phonotactic expectation plays a role in perception by providing, for example, rapid sensory restoration in case of noise (Leonard et al., 2015, 2016). However, it remains unclear how exactly our brain builds such priors. Are sonority patterns part of this process (Parker, 2012)? Is the model provided by BLICK physiologically plausible or do we rely on very different mechanisms to build phoneme expectation? While the first issue may require additional investigation with a separate experiment, here we tackled the second point explicitely by testing alternative models of phoneme expectation based on a probabilistic definition of phoneme probability based on a completely data-driven approach (P_{sur}, P_{ent}) (Brodbeck et al., 2018; Gaston and Marantz, 2018). These models produced probability vectors significantly correlated with the output of BLICK ($r_{blick,sur} = 0.43$, $r_{blick,ent} = 0.47$), but they were less effective at predicting brain activity. These same probability vectors were previously shown to be coupled with MEG cortical activity localised in core auditory cortex and superior temporal sulcus (Brodbeck et al., 2018). However, the effect was most prominent for short latencies (110 ms), while the phonotactic-EEG coupling measured here emerged between 100 and 400 ms and was strongest at latencies between 300 and 400 ms. While we can speculate that our result reflects cortical processes within those same cortical areas, there may be also other brain areas involved, especially in correspondence of long speech-EEG latencies such as 400 ms (**Fig. 4**), which were less prominent or absent in Brodbeck et al. (2018). Our phonotactic vector was also more strongly linked to the EEG signal than phonological neighbourhood density (P_{neigh}) (Vitevitch et al., 1999; Frisch et al., 2000; Bailey and Hahn, 2001). This is in line with previous studies suggesting distinct roles for phonotactics and neighbourhood density (Vitevitch et al., 1999; Bailey and Hahn, 2001; 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.

The speech-brain latencies and spatio-temporal maps that were measured here are in line with and complement the literature on phonotactic information. Previous studies indicated that phonotactic violations contribute to evoked brain components such as N400 and LPC (Friedrich and Friederici, 2005; Domahs et al., 2009; White and Chiu, 2017). It has also been suggested that phonotactic information and the size of the neighbourhood of permissible words may contribute at different stages of the speech perception process, possibly modulating distinct components of the neural response at different latencies (Vitevitch et al., 1999; Storkel et al., 2006). Our results contribute to this debate by suggesting that latencies of 100–500 ms are the most relevant for the processing of phonotactic probabilities. These latencies are consistent with the finding by Leonard et al. (2015) of an effect of phonotactics for latencies up to 400 ms. However, that result emerged specifically for high-frequency cortical signals (high-gamma power) that were recorded invasively (electrocorticography), and the authors did not see any effect for lower frequency data. Here we show for the first time that a negativity corresponding to phonotactics arises at latencies of 300–400 ms in low-frequency cortical responses to continuous speech. The fact that Leonard and colleagues could not measure this effect in superior temporal gyrus (STG) suggests that its cortical sources may be in sites that go beyond their coverage. The advantage of EEG over such invasive recordings is a more comprehensive coverage of the cortex.

Topographical patterns at the latencies where the effect of phonotactic arises present activations over centro-parietal scalp areas that qualitatively resemble that of an N400 component. One possibility is 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 (300–500 ms) (Pylkkänen et al., 2002, 2000). It is also possible that this response is related to an N400 reflecting phonological-level processing in a natural speech scenario (Deacon et al., 2004). However, a more specific experiment should be conducted to more clearly investigate this issue, with a direct comparison of EEG responses to phonotactic probabilities and phonotactic violations (as previously attempted in the similar context of semantic-level processing, Broderick et al., 2018).

4.2. 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 that arise starting from short speech-EEG latencies (100 ms).

Although we cannot be conclusive on this point, an effect of phonotactics at latencies of 100–500 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). Indeed, it is possible that other post-lexical brain responses to phonotactics exist but could not be reliably measured. One possibility is that such higher-level effects exhibit weaker time-locking to the phonotactic vector, which could reflect a different definition of phoneme expectation (different from the one estimated by BLICK), thus hampering the ability to capture them with our framework.

In summary, our results indicate that rapid time-locked brain responses to probabilistic phonotactics emerge for low-frequency cortical signals (<9 Hz). This phenomenon can be reliably measured at the individual subject-level using natural speech stimuli. We also found that the speech-EEG latencies of 100–500 ms most strongly reflect phonotactic information, which is in line with a pre-lexical account of phonotactic processing. This provides the field with new opportunities 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.

Conflicts of interest

None declared.

Acknowledgements

This work 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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.04.037.

References

Bailey, T.M., Hahn, U., 2001. Determinants of wordlikeness: phonotactics or lexical neighborhoods? J. Mem. Lang. 44, 568–591.
Balling, L.W., Baayen, R.H., 2012. Probability and surprisal in auditory comprehension of morphologically complex words. Cognition 125, 80–106.
Brent, M.R., Cartwright, T.A., 1996. Distributional regularity and phonotactic constraints are useful for segmentation. Cognition 61, 93–125.
Brodbeck, C., Hong, L.E., Simon, J.Z., 2018. Transformation from auditory to linguistic representations across auditory cortex is rapid and attention dependent for continuous speech. bioRxiv 326785.
Broderick, M.P., Anderson, A.J., Di Liberto, G.M., Crosse, M.J., Lalor, E.C., 2018. Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. Curr. Biol. 28, 803–809.
Carlson, M.T., Goldrick, M., Blasingame, M., Fink, A., 2016. Navigating conflicting phonotactic constraints in bilingual speech perception. Bilingualism 19, 939–954.

Chomsky, N., Halle, M., 1968. The Sound Pattern of English.
Church, K.W., 1987. Phonological parsing and lexical retrieval. Cognition 25, 53–69.
Cibelli, E.S., Leonard, M.K., Johnson, K., Chang, E.F., 2015. The influence of lexical statistics on temporal lobe cortical dynamics during spoken word listening. Brain Lang. 147, 66–75.
Connolly, J.F., Phillips, N.A., 1994. Event-related potential components reflect phonological and semantic processing of the Terminal word of spoken sentences. J. Cogn. Neurosci. 6, 256–266.
Crosse, M.J., Di Liberto, G.M., Bednar, A., Lalor, E.C., 2016. The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. Front. Hum. Neurosci. 10.
Davidson, L., 2006a. Phonotactics and articulatory coordination interact in phonology: evidence from nonnative production. Cogn. Sci. 30, 837–862.
Davidson, L., 2006b. Phonology, phonetics, or frequency: influences on the production of non-native sequences. J. Phonet. 34, 104–137.
Davidson, L., 2011. Phonetic, phonemic, and phonological factors in cross-language discrimination of phonotactic contrasts. J. Exp. Psychol. Hum. Percept. Perform. 37, 270–282.
de Cheveigné, A., Di Liberto, G.M., Arzounian, D., Wong, D., Hjortkjaer, J., Fuglsang, S.A., Parra, L.C., 2018a. Multiway Canonical Correlation Analysis of Brain Signals. bioRxiv 344960.
de Cheveigné, A., Wong, D.E., Di Liberto, G.M., Hjortkjær, J., Slaney, M., Lalor, E., 2018b. Decoding the auditory brain with canonical component analysis. Neuroimage 172, 206–216.
Deacon, D., Dynowska, A., Ritter, W., Grose-Fifer, J., 2004. Repetition and semantic priming of nonwords: implications for theories of N400 and word recognition. Psychophysiology 41, 60–74.
Dehaene-Lambertz, G., Dupoux, E., Gout, A., 2000. Electrophysiological correlates of phonological processing: a cross-linguistic study. J. Cogn. Neurosci. 12, 635–647.
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, G.M., Lalor, E.C., 2017. Indexing cortical entrainment to natural speech at the phonemic level: methodological considerations for applied research. Hear. Res. 348, 70–77.
Di Liberto, G.M., O'Sullivan, J.A., Lalor, E.C., 2015. Low-frequency cortical entrainment to speech reflects phoneme-level processing. Curr. Biol. 25, 2457–2465.
Di Liberto, G.M., Crosse, M.J., Lalor, E.C., 2018a. Cortical measures of phoneme-level speech encoding correlate with the perceived clarity of natural speech. Eneuro 5. ENEURO.0084-18.2018.
Di Liberto, G.M., Lalor, E.C., Millman, R.E., 2018b. Causal cortical dynamics of a predictive enhancement of speech intelligibility. Neuroimage 166.
Di Liberto, G.M., Peter, V., Kalashnikova, M., Goswami, U., Burnham, D., Lalor, E.C., 2018c. Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia. Neuroimage 70–79. NIMG-17-29.
Ding, N., Simon, J.Z., 2012. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. J. Neurophysiol. 107, 78–89.
Ding, N., Simon, J.Z., 2014. Cortical entrainment to continuous speech: functional roles and interpretations. Front. Hum. Neurosci. 8.
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.
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.
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.
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.
Dupoux, E., Pallier, C., Kakehi, K., Mehler, J., 2001. New evidence for prelexical phonological processing in word recognition. Lang. Cognit. Process. 16, 491–505.
Dupoux, E., Parlato, E., Frota, S., Hirose, Y., Peperkamp, S., 2011. Where do illusory vowels come from? J. Mem. Lang. 64, 199–210.
Ettinger, A., Linzen, T., Marantz, A., 2014. The Role of Morphology in Phoneme Prediction: Evidence from MEG.
Friedrich, M., Friederici, A.D., 2005. Phonotactic knowledge and lexical-semantic processing in one-year-olds: brain responses to words and nonsense words in picture contexts. J. Cogn. Neurosci. 17, 1785–1802.
Frisch, S.A., Large, N.R., Pisoni, D.B., 2000. Perception of wordlikeness: effects of segment probability and length on the processing of nonwords. J. Mem. Lang. 42, 481–496.
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.
Goldwater, S., Johnson, M., 2003. Learning OT constraint rankings using a maximum entropy model. In: Proc Stock Work Var within Optim Theory, pp. 111–120.
Gorman, K., Howell, J., Wagner, M., 2011. Prosodylab-aligner: a tool for forced alignment of laboratory speech. Can Acoust - Acoust Can. 39, 192–193.
Greenwood, D.D., 1961. Auditory masking and the critical band. J. Acoust. Soc. Am. 33, 484–502.
Grendár, M., 2001. Maximum Entropy: Clearing up Mysteries. Entropy 3 (2), 58–63.
Hallé, P.A., 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.
Hammond, M., 2004. Gradience, phonotactics, and the lexicon in English phonology. Int. J. Engl. Stud. 4.
Hayes, B., 2012. BLICK : a Phonotactic Probability Calculator (Manual).
Hayes, B., Wilson, C., 2008. A maximum entropy model of phonotactics and phonotactic learning. Ling. Inq. 39, 379–440.
Hotelling, H., 1936. Relations between two sets of variates. Biometrika 28, 321.
Jaynes, E.T., 1988. The Relation of Bayesian and Maximum Entropy Methods, Maximum-Entropy and Bayesian Methods in Science and Engineering.
Jusczyk, P.W., Luce, P.A., Charles-Luce, J., 1994. Infants' sensitivity to phonotactic patterns in the native language. J. Mem. Lang. 33, 630–645.
Jusczyk, P.W., Houston, D.M., Newsome, M., 1999. The beginnings of word segmentation in English-learning infants. Cogn. Psychol. 39, 159–207.
Khalighinejad, B., Cruzatto da Silva, G., Mesgarani, N., 2017. Dynamic encoding of acoustic features in neural responses to continuous speech. J. Neurosci. 37 (8), 2176–2185.
Kösem, A., van Wassenhove, V., 2016. Distinct contributions of low- and high-frequency neural oscillations to speech comprehension. Lang. Cogn. Neurosci. 1–9.
Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu. Rev. Psychol. 62, 621–647.
Lalor, E.C., Pearlmutter, B.A., Reilly, R.B., McDarby, G., Foxe, J.J., 2006. The VESPA: a method for the rapid estimation of a visual evoked potential. Neuroimage 32, 1549–1561.
Lalor, E.C., Power, A.J., Reilly, R.B., Foxe, J.J., 2009. Resolving precise temporal processing properties of the auditory system using continuous stimuli. J. Neurophysiol. 102, 349–359.
Lentz, T.O., Kager, R.W.J., 2015. Categorical phonotactic knowledge filters second language input, but probabilistic phonotactic knowledge can still be acquired. Lang. Speech 58, 387–413.
Leonard, M.K., Bouchard, K.E., Tang, C., Chang, E.F., 2015. Dynamic encoding of speech sequence probability in human temporal cortex. J. Neurosci. 35, 7203–7214.
Leonard, M.K., Baud, M.O., Sjerps, M.J., Chang, E.F., 2016. Perceptual restoration of masked speech in human cortex. Nat. Commun. 7, 13619.
Luck, S.J., 2005. An Introduction to the Event-Related Potential Technique.
Mattys, S.L., Jusczyk, P.W., 2001. Phonotactic cues for segmentation of fluent speech by infants. Cognition 78, 91–121.
Mattys, S.L., Jusczyk, P.W., Luce, P.A., Morgan, J.L., 1999. Phonotactic and prosodic effects on word segmentation in infants. Cogn. Psychol. 38, 465–494.
McClelland, J.L., Elman, J.L., 1986. The TRACE model of speech perception. Cogn. Psychol. 18, 1–86.
McClelland, J.L., Mirman, D., Holt, L.L., 2006. Are there interactive processes in speech perception? Trends Cognit. Sci. 10, 363–369.
McQueen, J.M., 1998. Segmentation of continuous speech using phonotactics. J. Mem. Lang. 39, 21–46.
Mesgarani, N., Cheung, C., Johnson, K., Chang, E.F., 2014. Phonetic feature encoding in human superior temporal gyrus. Science (80-) 343, 1006–1010.
Munz, E.D., 2017. Psychotherapie in der Psychiatrie. Nervenheilkunde 36, 800–805.
Obleser, J., Herrmann, B., Henry, M.J., 2012. Neural oscillations in speech: don't be enslaved by the envelope. Front. Hum. Neurosci. 6, 250.
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.
O'Sullivan, J.A., Power, A.J., Mesgarani, N., Rajaram, S., Foxe, J.J., Shinn-Cunningham, B.G., Slaney, M., Shamma, S.A., Lalor, E.C., 2015. Attentional selection in a cocktail party environment can Be decoded from single-trial EEG. Cerebr. Cortex 25 (7), 1697–1706 bht355.
Parker, S.G., 2012. The Sonority Controversy. De Gruyter Mouton.
Pisoni, D.B., Remez, R.E., 2005. The Handbook of Speech Perception. Blackwell Pub.
Pylkkänen, L., Stringfellow, A., Flagg, E., Marantz, A., 2000. A neural response sensitive to repetition and phonotactic probability : MEG investigations of lexical access. In: Proc Biomag 2000, 12th Int Conf Biomagn., pp. 1–4.
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.
Rossi, S., Jürgenson, I.B., 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.
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.
Scholes, R.J., 1966. Phonotactic Grammaticality. Hague Mout Co.
Sebastián-Gallés, N., 2007. Biased to learn language. Dev. Sci. 10, 713–718.
Storkel, H.L., 2001. Learning new words. J. Speech Lang. Hear. Res. 44, 1321.
Storkel, H.L., 2004. The emerging lexicon of children with phonological delays: phonotactic constraints and probability in acquisition. J. Speech Lang. Hear. Res. 47, 1194–1212.
Storkel, H.L., Morrisette, M.L., 2002. The lexicon and phonology. Lang. Speech Hear. Serv. Sch. 33, 24.
Storkel, H.L., Rogers, M.A., 2000. The effect of probabilistic phonotactics on lexical acquistion [WWW Document]. Clin Linguist Phon. URL. https://wordlearning.ku.edu/storkel-hl-rogers-ma-2000.
Storkel, H.L., Armbrüster, J., Hogan, T.P., 2006. Differentiating phonotactic probability and neighborhood density in adult word learning. J. Speech Lang. Hear. Res. 49, 1175.
Ulbrich, C., Alday, P.M., Knaus, J., Orzechowska, P., Wiese, R., 2017. Language, Cognition and Neuroscience the Role of Phonotactic Principles in Language Processing.

VanRullen, R., 2011. Four common conceptual fallacies in mapping the time course of recognition. Front. Psychol. 2, 365.

Vanthornhout, J., Decruy, L., Wouters, J., Simon, J.Z., Francart, T., 2018. Speech intelligibility predicted from neural entrainment of the speech envelope. J. Assoc. Res. Otolaryngol. 19, 181–191.

Vitevitch, M.S., Luce, P.A., Charles-Luce, J., Kemmerer, D., 1997. Phonotactics and syllable stress: implications for the processing of spoken nonsense words. Lang. Speech 40, 47–62.

Vitevitch, M.S., Luce, P.A., Pisoni, D.B., Auer, E.T., 1999. Phonotactics, neighborhood activation, and lexical access for spoken words. Brain Lang. 68, 306–311.

Wagner, M., Shafer, V.L., 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.

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.

Widmann, A., Schröger, E., 2012. Filter effects and filter artifacts in the analysis of electrophysiological data. Front. Psychol. 3, 233.

Wiese, R., Orzechowska, P., Alday, P.M., Ulbrich, C., 2017. Structural principles or frequency of use?. In: An ERP Experiment on the Learnability of Consonant Clusters, vol. 7.

Winther Balling, L., Harald Baayen, R., 2008. Morphological effects in auditory word recognition: evidence from Danish. Lang. Cognit. Process. 23, 1159–1190.

Zamuner, T.S., Kharlamov, V., 2016. Phonotactics and Syllable Structure in Infant Speech Perception. Oxford University Press.