



HAL
open science

Converging intracortical signatures of two separated processing timescales in human early auditory cortex.

Fabiano Baroni, Benjamin Morillon, Agnès Trébuchon, Catherine Liégeois-Chauvel, Itsaso Olasagasti, Anne-Lise Giraud

► **To cite this version:**

Fabiano Baroni, Benjamin Morillon, Agnès Trébuchon, Catherine Liégeois-Chauvel, Itsaso Olasagasti, et al.. Converging intracortical signatures of two separated processing timescales in human early auditory cortex.. NeuroImage, 2020, pp.116882. 10.1016/j.neuroimage.2020.116882 . hal-02613983

HAL Id: hal-02613983

<https://hal.science/hal-02613983>

Submitted on 4 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Converging intracortical signatures of two separated processing timescales in human early auditory cortex



Fabiano Baroni^{a,b,*}, Benjamin Morillon^c, Agnès Trébuchon^{c,d}, Catherine Liégeois-Chauvel^{c,e}, Itsaso Olasagasti^a, Anne-Lise Giraud^a

^a Department of Fundamental Neuroscience, University of Geneva, Geneva, Switzerland

^b School of Engineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

^c Aix Marseille Université, Institut National de la Santé et de la Recherche Médicale (INSERM), Institut de Neurosciences des Systèmes (INS), Marseille, France

^d Clinical Neurophysiology and Epileptology Department, Timone Hospital, Assistance Publique Hôpitaux de Marseille, Marseille, France

^e Department of Neurological Surgery, University of Pittsburgh, PA, 15213, USA

ARTICLE INFO

Keywords:

iEEG
Speech perception
Spectral analysis
Brain decoding
Computational modeling
Auditory cortex

ABSTRACT

Neural oscillations in auditory cortex are argued to support parsing and representing speech constituents at their corresponding temporal scales. Yet, how incoming sensory information interacts with ongoing spontaneous brain activity, what features of the neuronal microcircuitry underlie spontaneous and stimulus-evoked spectral fingerprints, and what these fingerprints entail for stimulus encoding, remain largely open questions. We used a combination of human invasive electrophysiology, computational modeling and decoding techniques to assess the information encoding properties of brain activity and to relate them to a plausible underlying neuronal micro-architecture. We analyzed intracortical auditory EEG activity from 10 patients while they were listening to short sentences. Pre-stimulus neural activity in early auditory cortical regions often exhibited power spectra with a shoulder in the delta range and a small bump in the beta range. Speech decreased power in the beta range, and increased power in the delta-theta and gamma ranges. Using multivariate machine learning techniques, we assessed the spectral profile of information content for two aspects of speech processing: detection and discrimination. We obtained better phase than power information decoding, and a bimodal spectral profile of information content with better decoding at low (delta-theta) and high (gamma) frequencies than at intermediate (beta) frequencies. These experimental data were reproduced by a simple rate model made of two subnetworks with different timescales, each composed of coupled excitatory and inhibitory units, and connected via a negative feedback loop. Modeling and experimental results were similar in terms of pre-stimulus spectral profile (except for the iEEG beta bump), spectral modulations with speech, and spectral profile of information content. Altogether, we provide converging evidence from both univariate spectral analysis and decoding approaches for a dual timescale processing infrastructure in human auditory cortex, and show that it is consistent with the dynamics of a simple rate model.

1. Introduction

The brains of humans and other animals generate electrical activity that often exhibits rhythmic patterns, which are apparent as shoulders or small bumps in the power spectrum of electrophysiological signals on top of the $1/f^\alpha$ profile (Buzsáki, 2006; Buzsáki and Draguhn, 2004). It has been suggested that rhythmic activity constitutes the neural basis of rhythmic and pseudo-rhythmic motor actions such as breathing, locomotion, chewing, peristalsis and the generation of vocalizations and

other communication signals, the latter being more prominently developed in primates and birds.

Neural oscillations correspond to rhythmic activity contributed by several thousand neurons, astrocytes and possibly other cell types. They constitute a compact, low-dimensional signature of the local network state that can be informative about contextual cognitive state (e.g. arousal (Steriade et al., 1993; McGinley et al., 2015), attention (Fries et al., 2001; Besle et al., 2011; Ding and Simon, 2012; Zion Golumbic et al., 2013; Klimesch, 2012; Clayton et al., 2015; Calderone et al., 2014)

* Corresponding author. Department of Fundamental Neuroscience, University of Geneva, Geneva, Switzerland.

E-mail address: fabianobaroni@gmail.com (F. Baroni).

¹ Current address: Escuela Politécnica Superior, Universidad Autónoma de Madrid, Madrid, Spain.

and expectation (Saleh et al., 2010; Stefanics et al., 2010; Cravo et al., 2011; Rohenkohl and Nobre, 2011; Arnal and Giraud, 2012; Arnal et al., 2015; Morillon et al., 2015; Morillon and Baillet, 2017; Breska and Deouell, 2017)) and even provide diagnostic information about pathological conditions such as schizophrenia, autism and dyslexia (Roopun et al., 2008; Lehongre et al., 2011, 2013; Calderone et al., 2014; Uhlhaas et al., 2010; Uhlhaas and Singer, 2012; Sun et al., 2013; Voytek and Knight, 2015; Soltész et al., 2013; Jochaut et al., 2015; Simon and Wallace, 2016).

Accordingly, alterations of spectral features observed in clinical populations have been related to microscopic anomalies in interneuronal function (Gonzalez-Burgos and Lewis, 2008; Pizzarelli and Cherubini, 2011) and/or in the local balance and coordination between synaptic excitation and inhibition (Fenton, 2015; Gao and Penzes, 2015). Macroscopic features related to rhythmic brain activity could hence reflect microscopic anomalies at the neuronal level and, at least in some cases, be related to specific sets of susceptibility genes (Ramamoorthi and Lin, 2011; Gao and Penzes, 2015; Benítez-Burraco and Murphy, 2016), further enhancing their interest for both basic and clinical research.

In humans, multiple pieces of evidence suggest that auditory perception, and its associated brain activity, is not a scale-free process, but presents at least two separated frequency bands, approximately located near the classically defined delta-theta (1–8 Hz) and gamma (30–60 Hz) bands, where perception and brain entrainment surpass those observed in intermediate frequencies (Poeppl, 2003; Boemio et al., 2005; Luo and Poeppl, 2012; Edwards and Chang, 2013; Ross et al., 2014; Teng et al., 2016, 2017).

It has been suggested that the motor and the auditory system co-evolved neural mechanisms for the generation of oscillations in corresponding frequency bands, matching the natural resonance properties of the jaw, the tongue and other components of the vocal apparatus that are critical for the generation of articulated vocalizations (Morillon et al., 2010; Kotz et al., 2018; Ravignani et al., 2019). In particular, neural oscillations in early auditory cortical regions (i.e. primary and secondary auditory cortex) are thought to be involved in the parsing and representation of speech constituents at the corresponding temporal scale, particularly at the syllabic (~ 2–8 Hz) and phonemic (~ 30–50 Hz) scales (Giraud and Poeppl, 2012).

In this work, we present intracranial data recorded from ten epilepsy patients while they were listening to short sentences. Spectral analysis revealed that neural activity in early auditory cortical regions in the absence of sensory stimulation does not conform to a purely scale-free process (i.e., one characterized by a $1/f^\alpha$ power spectrum profile): most individual subjects exhibited power spectra with a shoulder in the delta range and a small bump in the beta range. These were apparent as convex deflections of the power spectrum, which could be as pronounced as to give rise to a local maximum (a spectral peak). Speech stimulation decreased power in the beta range, and increased power in the delta-theta and gamma ranges.

By performing a series of spectrally-resolved decoding analyses of brain signals for speech detection and discrimination, we demonstrated that not only the information required for speech detection (i.e., distinguishing speech from pre-stimulus activity), but also information enabling speech discrimination (i.e., distinguishing between different speech segments) is preferentially conveyed in single-trial patterns of power values in low (delta - theta) and high (gamma - high-gamma) frequency bands. In some subjects, the information conveyed in single-trial patterns of phase values also exhibited such a bimodal profile whereas, at the group level, phase information was best described by a low-pass profile, with decoding accuracy decreasing with increasing frequency.

Based on our experimental observations and consistently with previous electrophysiological studies in non-human animals (Lakatos et al., 2005), we implemented a simple rate model comprising two distinct subnetworks with different timescales, each composed of coupled excitatory and inhibitory units. The slower subnetwork exhibits a

resonance in the delta-theta range (3.8 Hz), while the faster subnetwork exhibits a resonance in the gamma - high-gamma range (85.9 Hz). These frequency values roughly correspond to the delta-theta shoulder in the iEEG power spectrum in the absence of sensory stimulation, and the gamma - high-gamma bump in the iEEG power spectrum difference between speech and pre-stimulus activity, respectively. They are connected via an inter-subnetwork connectivity pattern implementing a negative feedback loop between the fast and the slow subnetwork. We show that this simple model bears a remarkable resemblance to intracranial electroencephalography (iEEG) activity, both in terms of pre-stimulus spectral profile, of spectral modulations with speech, and of spectral profile of information content for speech detection and discrimination as assessed by multivariate machine learning techniques. These features do not require fine-tuning of the model parameters and are observed in a broad region of parameter space.

The current experimental findings challenge the view that extracellular field potentials in general, and electrocorticography (ECoG) and iEEG signals in particular, could be scale-free processes (Miller et al., 2009; He et al., 2010; He, 2014; Podvalny et al., 2015), and rather reveal preferred timescales in which neural responses to sensory stimuli are predominantly expressed and information about sensory inputs more effectively encoded. Our modeling investigations further show that macroscopic features of neural activity and information encoding as recorded by intracranial EEG from human early auditory cortex during speech perception can be faithfully approximated by a simple population rate model incorporating specific synaptic interactions between a slower (delta - theta) and a faster (gamma - high-gamma) neural subnetwork.

2. Materials and methods

2.1. Data acquisition

We recorded intracranially with iEEG electrodes from 10 epilepsy patients undergoing pre-surgical monitoring. Two (six) patients were implanted with EEG electrodes on the left (right) hemisphere only, and 2 patients were implanted bilaterally. Multi-contact electrodes (0.8 mm diameter, 10 or 15 contacts of 2 mm length each with 1.5 mm spacing between contacts) were orthogonally introduced in the stereotactic space (Szikla et al., 1977; Talairach and Tournoux, 1988). Electrode location was based solely on clinical criteria. Patient age, sex, educational level, language dominance, presence of brain MRI abnormalities and locations of seizure foci are reported in Table S1. The anatomical position of each contact was then identified on the basis of (i) an axial scanner image acquired before the removal of electrodes, and (ii) an MRI scan performed after the removal of electrodes (Liegeois-Chauvel et al., 1991). Only electrodes passing through Heschl's gyrus were considered.

The iEEG recordings were monopolar, with each contact of a given depth electrode referenced to an extra-dural lead using acquisition software and a 128-channel SynAmps EEG amplification system from NeuroScan Labs (Neurosoft Inc.). During the acquisition, the iEEG signal was high-pass filtered at 0.5 Hz and amplified with an anti-aliasing filter at 200 Hz (temporal resolution of 1 ms and amplitude resolution of 1 μ V). Sampling rate for the iEEG signal was 1000 Hz.

We did not record for 12 h after any generalized seizure event. Voltage traces from each electrode and each trial were visually inspected; one trial from subject S8 included large amplitude high-frequency deflections of clearly non-physiological origin, and was thus excluded from the analyses. Intracranial recordings exhibit very little contamination from non-neural (e.g., muscular) sources (with the exception of specific regions of the temporal pole that can exhibit activity related to eye movements (Jerbi et al., 2009; Kovach et al., 2011)). Isolated interictal epileptiform discharges could have occurred; however, their timing, frequency and amplitude are expected to be unrelated to our variables of interest.

2.2. Experimental protocols

In this study we considered data from 2 different auditory protocols: a functional localizer and a speech (phrase) protocol.

In the functional localizer, 24 ms pure tone sounds with 2 ms on and off ramps were presented binaurally (except for subject S7, who received monaural stimuli). The interstimulus interval was uniformly distributed in the interval 1030 ms + [-200,200] ms for the bilaterally implanted patients, while it was constant and fixed at 1030 ms for the other patients. Pure tone frequencies varied across subjects and are reported in [Table S2](#) together with the number of trials in each protocol.

In the phrase protocol, subjects listened to several repetitions of two 2.5 s long sentences in French, uttered by a French female whose voice had a fundamental frequency of 201 Hz. Both trial types correspond to the French sentence “le nouveau garde la porte”, pronounced with different prosody: in the first trial type (hereafter P1), a short pause occurred after “garde”, while in the second trial type (hereafter P2), a short pause occurred after “nouveau”. Stimuli were presented monaurally in a pseudo-randomized order which was the same for every subject (P1, P1, P2, P2, P2, ...) with an interstimulus interval of 4135 ms, and only the (dominant) contra-lateral response was taken into account (as the auditory ascending pathway is partly decussated). In both protocols, stimuli were delivered through headphones at a comfortable hearing level (mean \sim 30 dB sensation level) in a pseudo-randomized order at a 22 kHz rate using E-prime 1.1 (Psychology Software Tools Inc., Pittsburgh, PA, USA). Patients were instructed to passively listen and concentrate on what they heard. A subset of this dataset has previously been reported in ([Morillon et al., 2012](#); [Fontolan et al., 2014](#)).

3. Data analysis

3.1. Data preprocessing

We bipolar re-referenced the original iEEG signals to remove low spatial frequency components and hence obtain a more localized signal to better exploit the fine spatial resolution of iEEG recordings; that is, for each multi-contact electrode with N contacts (where N equals 10 or 15), we generated a set of N-1 bipolar channels by subtracting the voltage traces acquired by neighboring contacts.

3.2. ERP analysis and channel selection

We measured Event-Related Potentials (ERP) to the pure tone stimuli to functionally select the channels that recorded activity from early auditory cortical areas (i.e. primary and secondary auditory cortex, corresponding to Brodmann areas 41 and 42). The raw voltage traces were aligned with respect to tone onset and averaged across trials for each frequency value. ERP amplitude was defined as the difference between the maximum and the minimum value in the time interval [0,200] ms post-tone onset, expressed in units of baseline standard deviation (SD). Baseline SD was measured by calculating the standard deviation across trials for each time point, and then averaging across time points. For each channel, the maximum ERP amplitude across the tested frequencies was defined as the ERP amplitude for that channel. For each multi-contact electrode, only the channel with the largest ERP amplitude was retained for further analysis (one channel per patient for the unilaterally implanted patients, two channels for the bilaterally implanted patients). The data corresponding to right ear stimulus presentation for one of the bilateral patients was lost, and the selected channel corresponding to right ear stimulus presentation for the other bilateral patient was discarded because of non-significant ERP size (1.57 SD), resulting in a total number of 10 channels, one for each subject, that entered the main analysis. ERP amplitudes for these channels were in the interval [4,18] SD. Spectral tuning curves at non-selected channels were the same as those at the selected channel for each multi-contact electrode, but exhibited ERP of lesser amplitude (i.e., lower signal-to-noise ratio);

hence, larger inter-trial variability. This indicates that our iEEG multi-contact electrodes were not sampling from cortical regions with different preferred frequency, at least within the frequency resolution tested.

We also assessed high-gamma power responses by first wavelet transforming each individual trial as described below, and then averaging $10\log_{10}$ power values in the frequency range 60–150 Hz and time interval [20,100] ms post stimulus onset for each pure tone frequency. The resulting values were then z-scored with respect to the mean and standard deviation across trials of the average $10\log_{10}$ power in the same frequency range for the pre-stimulus time interval [-280,-200] ms. High-gamma power responses exhibited very similar frequency selectivity, and resulted in the same selected channel for 6/12 multi-contact electrodes or an immediately adjacent one for 4/12 multi-contact electrodes. The two remaining multi-contact electrodes either did not exhibit significant high-gamma responses (S7-RH), or did not exhibit significant responses in either of the two measures (S7-LH). Large ERP responses are a classic feature of neural responses in early auditory cortex due to synchronous excitatory volleys from the medial geniculate nuclei which are highly time-locked across trials ([Liegeois-Chauvel et al., 1991](#)), as opposed to higher order auditory regions where ERPs can be small or absent, and relatively dissociated from high-gamma responses ([Nourski et al., 2015](#); [Brugge et al., 2009](#)). Neural responses were stronger when considering bipolar rather than monopolar channels for both measures, further justifying our choice of bipolar re-referencing.

3.3. Spectral analyses

We considered spectrotemporal representations of iEEG or simulated signals based on the Continuous Wavelet Transform (CWT), using a complex Morlet wavelet with bandwidth parameter equal 1.5. To improve visualization and yield a distribution that is closer to normal, we applied a non-linearity ($10\log_{10}$) to the power values. For the estimation of phase consistency across trials, we calculated the Phase Locking Value (PLV) ([Tallon-Baudry et al., 1996](#); [Lachaux et al., 1999](#)):

$$PLV(t, f) = \frac{1}{N_{\text{Trials}}} \left| \sum_{j=1}^{N_{\text{Trials}}} e^{i\Phi(t, f, j)} \right|, \quad (1)$$

where N_{Trials} is the number of trials and $\Phi(t, f, j)$ is the phase angle in radians for the considered time point, frequency value and trial.

To quantify the average power spectra during the pre-stimulus baseline and during speech stimulation, we estimated spectral power separately for the baseline ([-1000,0] ms) and the speech period ([0,1000] ms) by computing CWT-transformed signals for each time window separately, extracting $10\log_{10}$ power, and then averaging over trials and time points within each time window. Voltage traces were normalized in each trial before CWT transformation by removing the mean value in each time window, and dividing by the standard deviation of the pre-stimulus baseline window.

For the estimation of phase consistency across trials during the pre-stimulus baseline and during speech stimulation, phase was extracted from the CWT-transformed signals calculated separately for the baseline ([-1000,0] ms) and the speech period ([0,1000] ms) for each trial, each frequency value and each time point. Then, the PLV value for each time window was calculated as

$$PLV(f) = \frac{1}{N_{\text{Trials}} N_t} \sum_{t=1}^{N_t} \left| \sum_{j=1}^{N_{\text{Trials}}} e^{i\Phi(t, f, j)} \right|, \quad (2)$$

where N_t is the number of time points.

3.4. Model

We implemented a neuronal network composed of rate units of the

Wilson-Cowan type (Wilson and Cowan, 1972), adapted from (Mejias et al., 2016).

The network is composed of two subnetworks: a fast (G) subnetwork and a slow (T) subnetwork, exhibiting activity in the gamma - high-gamma and delta-theta range, respectively. Each subnetwork comprises an excitatory and an inhibitory population, mutually connected, whose average firing rates evolve according to the following equations:

$$\tau_E \frac{dr_E(t)}{dt} = -r_E + \Phi(I_E^{net} + I_E^{ext}) + \sqrt{\tau_E} \eta_E(t) \quad (3)$$

$$\tau_I \frac{dr_I(t)}{dt} = -r_I + \Phi(I_I^{net} + I_I^{ext}) + \sqrt{\tau_I} \eta_I(t), \quad (4)$$

where $r_{E(I)}$ is a dynamical variable describing the (dimensionless) average firing rate of the excitatory (inhibitory) population, and $\tau_{E(I)}$ is the corresponding time constant. The input term I^{net} represents the synaptic inputs arriving from other populations in the network, while the input term I^{ext} represents the inputs from external sources, such as sensory stimuli or other areas not explicitly included in the model. Input terms are passed through a rectifying non-linearity $\Phi(x) = x / (1 + e^{-x})$. Taking into account only local, intra-subnetwork, contributions, the network input is given by

$$I_E^{net} = J_{EE}r_E + J_{EI}r_I \quad (5)$$

$$I_I^{net} = J_{IE}r_E + J_{II}r_I \quad (6)$$

In the linear approximation (which is perfectly realized when $I^{ext} \gg I^{net}$) and ignoring the inter-subnetwork coupling, each subnetwork is characterized by a pair of complex conjugate eigenvalues with negative real part (indicating stable dynamics), and imaginary parts corresponding to a natural frequency of 85.9 Hz for the fast subnetwork, and of 3.8 Hz for the slow subnetwork.

Each population also receives a noise term $\eta(t)$, which represents intrinsic and synaptic noise and other sources of variability. It is implemented as colored noise (i.e., a Ornstein-Uhlenbeck process), which evolves according to the equation

$$\frac{d\eta(t)}{dt} = -\frac{1}{\tau_{bg}} \eta(t) + \sqrt{\frac{2\sigma_{bg}^2}{\tau_{bg}}} \zeta(t) \quad (7)$$

where τ_{bg} is the auto-correlation time constant, and σ_{bg} is the standard deviation. In the canonical model used throughout this study, the only inter-subnetwork connections are those from Ge to Te, and from Te to Gi. This inter-subnetwork connectivity scheme implements a negative feedback loop from the fast to the slow subnetwork (from Ge to Te), and back to the fast subnetwork (from Te to Gi). This choice is motivated by dynamical systems and control theoretical arguments indicating that negative feedback loops from fast to slow subsystems constitute dynamical building blocks for fast and stable responses to inputs, yielding the balance between sensitivity and stability that is common to many biological systems (e.g., (Bhalla and Iyengar, 1999; Rene and D'Ari, 1990; Edelstein-Keshet, 2005)), and is consistent with connectivity results derived from anatomical studies (Thomson and Bannister, 2003; Barbour and Callaway, 2008) and previous modeling work (Lee et al., 2013, 2015).

To deliver speech input to the model, we first extracted the envelope of the speech signal. We used a time-domain method implemented in Matlab (the ‘‘peak’’ method for the function envelope.m), which estimates the upper and lower envelopes of a signal by using spline interpolation of the maxima and minima of the signal, respectively. After extraction of the upper and lower envelopes, we took the maximum of the absolute value of the two at each time point as an estimate of the total envelope $I_{sc}(t)$.

Speech input to the network is mediated by short-term synaptic

depression, as often observed in thalamocortical synapses (Thomson and Deuchars, 1994; Elhilali et al., 2004). Short-term synaptic depression has been implemented according to the Tsodyks-Markram model (Tsodyks et al., 1998): we considered synaptic transmission to be mediated by a dynamical variable x representing the average fraction of available neurotransmitters in the bottom-up pathway conveying the speech input, which evolves as

$$\frac{dx}{dt} = \frac{1-x}{\tau_D} - Ux(t)f(I_{sc}(t)), \quad (8)$$

where f is the sigmoid function $f(x) = 1 / (1 + e^{-k_D(x-x_0)})$. The expression above describes how the fraction of available neurotransmitters decays with incoming sensory input with rate U , and it recovers with time constant τ_D . Finally, the speech input to the network, delivered to the population Ge, is calculated as follows:

$$I_{in}(t) = \mathcal{A}x(t)f(I_{sc}(t)). \quad (9)$$

For each simulated trial, rate variables were initialized at random values extracted from a Gaussian distribution with zero mean and unit variance. Then, model equations were integrated with a fourth-order fixed step-size Runge-Kutta algorithm with time step $\Delta t = 0.1$ ms. The first second of simulated activity was considered a transient period and was hence discarded. After one additional second of spontaneous activity, speech input started, thus recapitulating the structure of experimental trials. Most model results are illustrated using a set of 30 simulated trials per phrase, which is in the range of experimental trials available per subject and phrase. However, decoding results are computed using 100 trials per phrase to yield more accurate estimates. Parameter values and descriptions are provided in Table S3.

3.5. Multivariate classification analyses

We estimated the amount of information conveyed by neural signals using binary Regularized Least-Square Classifiers (RLSC (Rifkin et al., 2003),) with regularization parameter $\lambda = 10^6$. Regularized Least-Square Classification is a machine learning technique that estimates the linear separability between patterns according to their class.

Our goal was to assess the amount of information conveyed by a spectrotemporal representation of the iEEG signal in each trial, obtained by applying a complex Morlet wavelet transformation, about the presented acoustic stimulus. In particular, we considered either $10 \log_{10}$ power, or phase angle (after taking sine and cosine), or both, at 12×17 (time \times frequency) points for each time window in each trial, sampled from a grid with temporal duration equal 360 ms, as the input to the classifiers. Sampled points in time were evenly spaced, while points in frequency were sampled with higher resolution for lower frequencies (4 Hz intervals from 4 to 20 Hz, 10 Hz interval from 20 to 80 Hz, 20 Hz interval from 80 to 200 Hz).

This approach is similar to previous work that used multivariate patterns obtained from multi-taper spectral estimates (Tsuchiya et al., 2008; Baroni et al., 2017), but here we considered wavelet-based spectral estimates to leverage their excellent time-frequency resolution especially at intermediate and high frequencies, which could be beneficial given the highly dynamic nature of speech stimuli, and the fact that beta and gamma events tend to occur in short-lived bursts, as observed in our dataset and previously reported by others (e.g. (Sherman et al., 2016; Lundqvist et al., 2016)).

We also performed a set of spectrally-resolved decoding analyses where either power or phase patterns were sampled across time for a set of frequency values, unevenly spaced as described above, hence resulting in 12-dimensional patterns as inputs to the classifiers.

A set of weights that optimally separate trials according to their class is determined using a subset of the available trials, denoted as training set. The performance of the classifier is defined using a different set of trials, denoted as test set, as the area under the

Receiver Operating Characteristic (ROC) curve, which we refer to as A' (A' prime). We report the average A' values over N_{iter} cross-validations. In each cross-validation, we randomly chose a set of $0.7 \times \min(N_1, N_2)$ (rounded to the nearest integer) trials of each class as the training set, where N_1 and N_2 are the number of trials in class 1 and 2, respectively. As the test set, we chose $\min(N_1, N_2) - \text{round}(0.7 \times \min(N_1, N_2))$ trials of each class among those that are not already included in the training set. Before being fed to the classifier, inputs were z-transformed: the mean and standard deviation of the relevant variable at each time-frequency point in the training set was calculated, and used to transform both training and test sets. Then, optimal RLSC weights were estimated using training trials, and their capacity to separate test trials according to their class was measured as the area under the ROC curve (A'). The number of cross-validations N_{iter} was set to 1000 for all the decoding analyses.

Significance of A' values was estimated via a permutation-based statistics. For each classification considered, the class labels were randomly shuffled. Then, the average A' value over N_{iter} realizations of training and test sets was calculated as described above. This procedure was repeated $N_{\text{perm}} = 1000$ times, yielding a probability distribution of average A' values corresponding to the null hypothesis of lack of linear separability between the two classes. An empirical average A' value was considered significant at level p if it exceeded the p -percentile of the corresponding null distribution ($p = 0.01$, $p = 0.001$). Significance thresholds at $p = 0.01$ were estimated separately for each classification considered. In order to improve the estimation of the significance threshold at $p = 0.001$, null A' values were pooled across decoding types, and the corresponding significance threshold was calculated from the resulting null distribution.

We considered 3 distinct time windows for each sentence: a baseline time window ([-680,-320] ms with respect to speech onset), a speech onset time window, corresponding to the utterance of the speech token “le nou” (hereafter T1; [0,360] ms from speech onset), and a later speech time window, corresponding to the utterance of the speech token “la po” (hereafter T2; [1704,2064] ms for P1, [1589,1949] ms for P2). We considered all possible within-phrase binary decoding (T1 vs. baseline, T2 vs. baseline, and T1 vs. T2, for both P1 and P2), as well as the binary decoding analyses between corresponding speech tokens in P1 and P2.

To simplify the presentation of our results, we further grouped these decoding analyses into three categories: detection (speech token vs. baseline, comprising 4 decoding analyses: T1 vs. baseline and T2 vs. baseline for both P1 and P2), easy discrimination (between different speech tokens, or the same speech token immediately following continuous speech or occurring after a pause, comprising 3 decoding analyses: T1P1 vs. T2P1, T1P2 vs. T2P2, and T2P1 vs. T2P2), and hard discrimination (between the same speech token occurring at sentence onset in the two phrases: T1P1 vs. T1P2).

In spite of its stochastic nature, the rate model presented in the previous subsection does not fully capture the extent and complexity of the inter-trial variability observed in the iEEG dataset; in fact, the rate model yields perfect decoding accuracy in all the detection and discrimination decoding analyses just described. Hence, in order to probe the encoding properties of the model with stimuli that are similar to those employed in the iEEG experiments, we generated a set of stimuli by interpolating between P1 and P2 for each time point: $P_{\text{interp}}^i(t) = (N_{\text{interp}} - i) / (N_{\text{interp}} - 1) P1(t) + (i - 1) / (N_{\text{interp}} - 1) P2(t)$, with $i = 1, 2, \dots, N_{\text{interp}}$ and $N_{\text{interp}} = 10$.

To assess the degree of specificity vs. generalization in the read-out of the information conveyed by patterns of either power or phase values, we performed a series of decoding analyses where training and test patterns were extracted from different frequency values. This approach resembles the temporal generalization method (King and Dehaene, 2014), but generalization is assessed across the frequency instead of the time dimension.

4. Results

4.1. Spectral analysis of neural responses to speech

Most selected channels from early auditory cortex exhibited trajectories with marked and reproducible deflections in correspondence to certain acoustic landmarks, most prominently sentence onset and speech tokens that follow a pause (Fig. 1B). Another distinctive feature was the presence of a power spectral peak or a shoulder in the beta range, which decreased in amplitude with speech presentation. Speech stimulation also resulted in power increases in low (delta-theta) and high (gamma - high-gamma) frequency ranges.

Every subject exhibited this bimodal profile of spectral power changes with speech stimulation; however, the frequency ranges where power increased with stimulation and those where power decreased with stimulation exhibited some inter-individual variability (Fig. 1C-E; see also Fig. 2F). In particular, the transition between the low-frequency activation and the mid-frequency deactivation could vary between 7 and 22 Hz, and the transition between the mid-frequency deactivation and the high-frequency activation could vary between 20 and 40 Hz.

In spite of inter-individual variability in these frequency bands, the average power spectra across the population of selected channels exhibited features that are consistent with the single-channel analyses: the presence of a beta peak during baseline, which decreases with stimulation, and a bimodal profile of activation in low (delta-theta) and high (gamma - high-gamma) frequencies, separated by an intermediate deactivated frequency band.

The pattern of phase-locking was also variable across subjects; however, most subjects as well as the mean profile across subjects exhibited a low-pass profile with strong phase-locking in the delta-theta range, and progressively lower phase-locking at higher frequencies (Fig. 1E, right panel). It is worth noting, however, that some individual subjects exhibited a band-pass PLV profile, with a peak ~ 2 Hz. Given previous work it is likely that a band-pass PLV profile with a peak $\sim 1-2$ Hz could have been identified in all subjects if longer continuous speech segments had been used, allowing for the assessment of phase-locking at lower frequencies (Edwards and Chang, 2013). As our main interest lies in the characterization of neural processes that could potentially underlie speech perception at the (sub-)phonemic, syllabic and word-level scales, frequencies below 1 Hz were not investigated.

A population rate model reproduces the spectral features observed in the iEEG data.

We assessed the extent to which a simple population rate model is capable of reproducing the spectral features of neural responses presented in the previous section. To this end, we implemented a population rate model following the formalism initially proposed by Wilson and Cowan (1972). The model comprises two subnetworks, each composed by an excitatory and inhibitory unit with self and reciprocal connections. Further details, including the model equations, are presented in subsection “Model”.

In spite of its simplicity, the population rate model can qualitatively reproduce several features of iEEG recordings. In particular, the power spectrum during baseline exhibits a knee in the delta-theta range, and an approximately $1/f^\alpha$ profile at higher frequencies. However, it lacks a beta peak, which was observed in most subjects. In response to speech, spectral power increases in low (delta-theta) and high (gamma - high-gamma) frequencies, with no significant change in the intermediate (alpha-beta) range. This profile is similar to what observed in the iEEG data. However, while all iEEG subjects presented a significant beta power reduction with speech, the model does not exhibit significant power changes in this frequency range (Fig. 2F).

The rate model fits the iEEG data to a degree that is comparable to inter-subject variability.

To quantify the similarity between model and iEEG trajectories, we conducted a spectrally-resolved comparison between the model and each iEEG subject, and an analogous spectrally-resolved comparison between

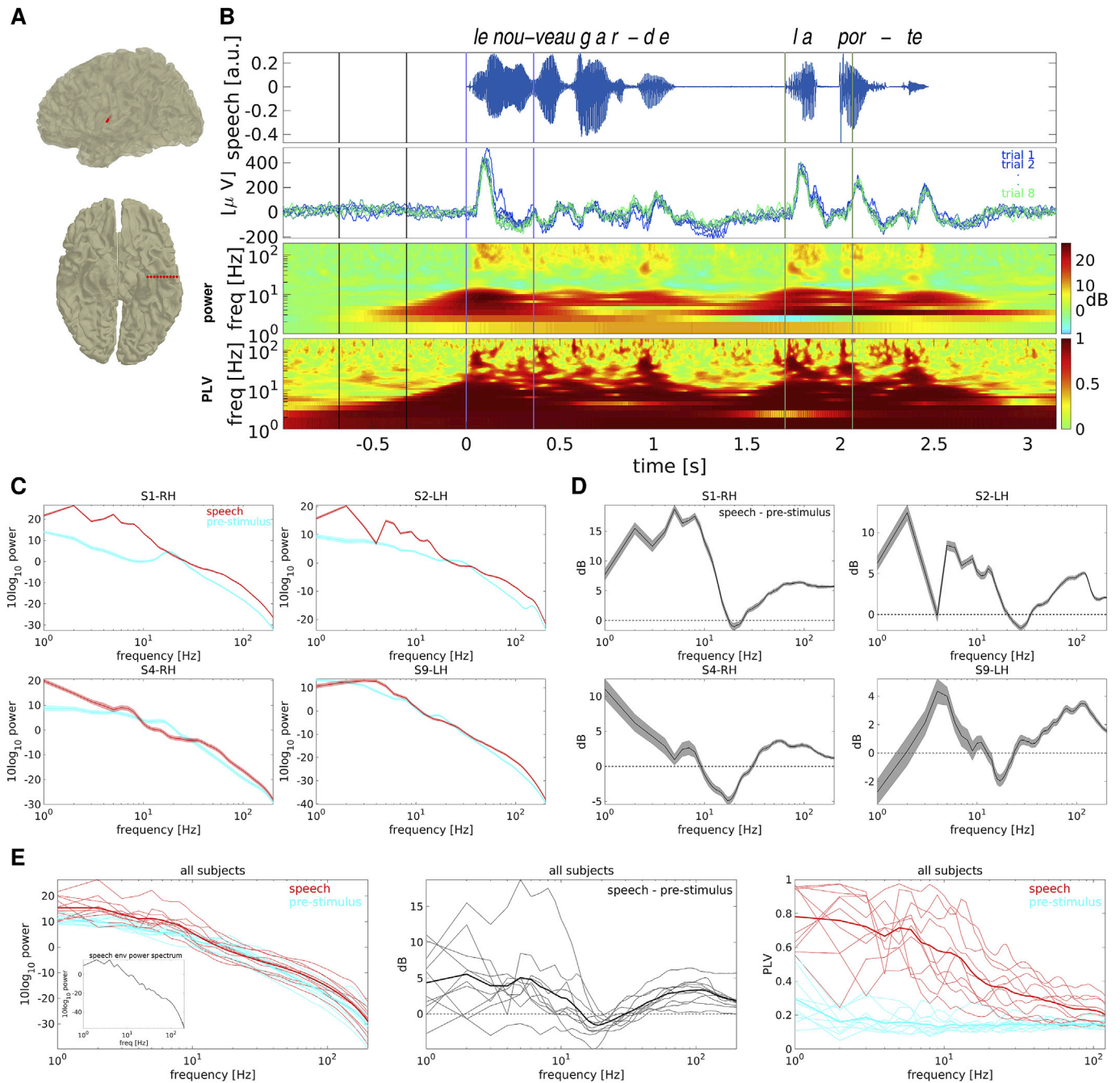


Fig. 1. Neural responses to speech in early cortical areas. A: Anatomical image showing the multi-contact electrode implantation in early auditory cortex in an example subject on a standard brain. B: Stimulus and neural responses recorded from a strongly speech responsive channel from subject S1. Top: input stimulus corresponding to the French sentence “le nouveau garde la porte”. Middle-top: voltage trajectories for the first 8 trials. Middle-bottom: trial-averaged wavelet power spectrogram. Differences in $10\log_{10}$ power with respect to a pre-stimulus baseline ([-1000,-700] ms) are shown for the ease of visualization. Bottom: PLV spectrogram. Vertical lines indicate the three 200 ms time windows used for the speech detection and discrimination analyses shown in subsection “Stimulus encoding occurs preferentially at delta-theta and gamma frequencies”, which corresponds to pre-stimulus activity (black), “le nou” (purple) and “la po” (green). Corresponding panels for all selected channels and for both phrases are shown in Fig. S1. C: Trial-averaged power spectra calculated separately for the pre-stimulus (cyan) and the speech (red) period for four representative subjects. Shaded areas indicate s.e.m. Corresponding panels for all selected channels are shown in Fig. S2. D: As in C), but the difference in spectral power between the speech and the pre-stimulus period is shown. Corresponding panels for all selected channels are shown in Fig. S3. E: Left, middle: As in C), D), for each selected channel (thin lines). Average quantities over the set of selected channels are shown with thick lines. Inset shows the power spectrum of the speech envelope signal. Right: wavelet PLV, calculated separately for the pre-stimulus (cyan) and the speech (red) period, for each selected channel (thin lines) and averaged over the set of selected channels (thick lines). Phase-locking values for each selected channel are shown separately in Fig. S4.

each pair of iEEG subjects. The similarity between the model and each iEEG subject was assessed by computing, for each subject and for each frequency value between 1 and 200 Hz, the normalized cross-correlation between the $10\log_{10}$ power trajectories for each pair of trials ($t_i^{model}, t_j^{S_k}$),

where t_i^{model} is a trial from the model simulation and $t_j^{S_k}$ is a trial from the iEEG subject S_k . The normalization ensures that the resulting values are included in [0,1], and that the corresponding auto-correlations at zero lag equal one. Then, cross-correlations were averaged across trial pairs.

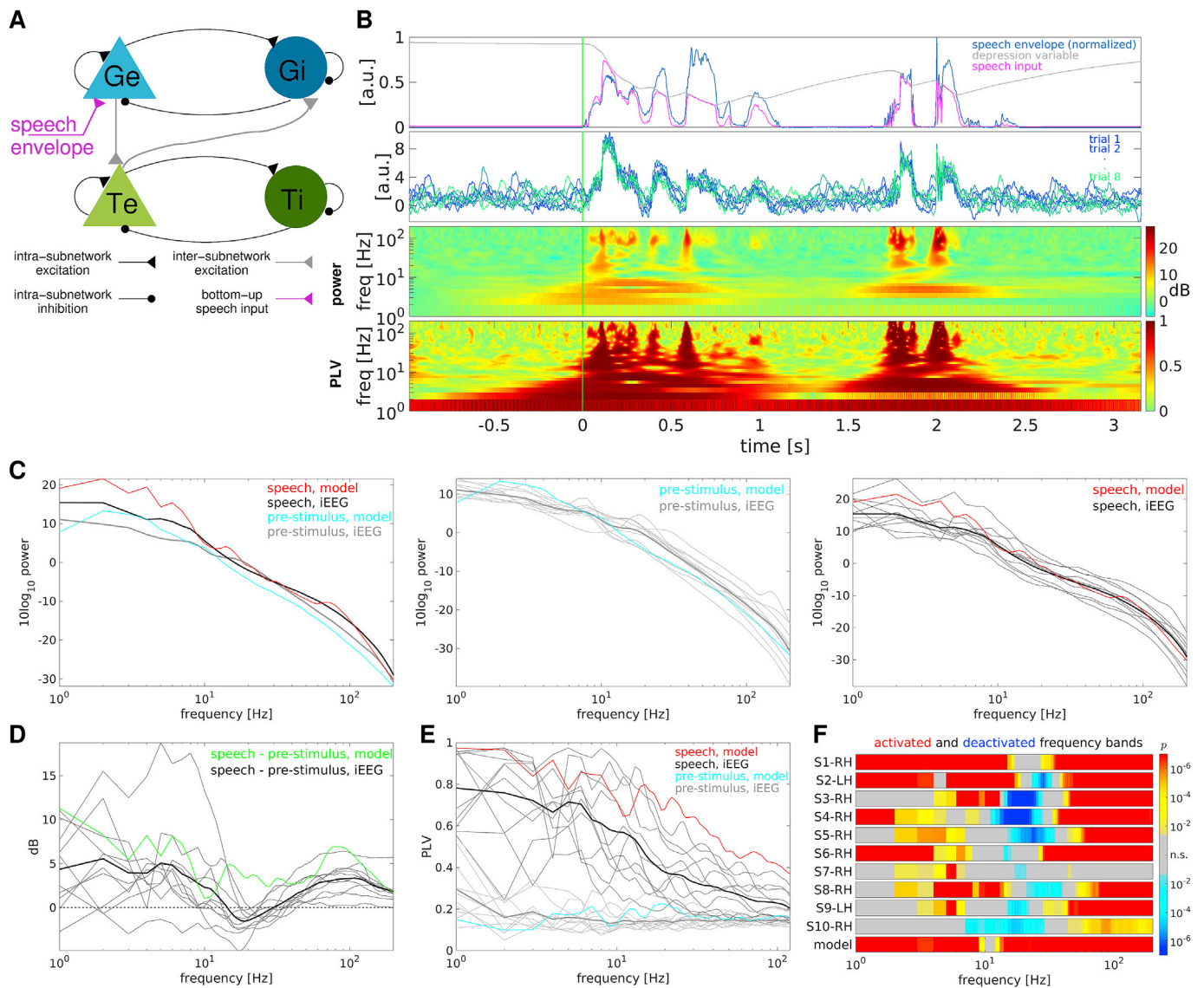


Fig. 2. Responses to speech in a population rate model. A: Model diagram. Shades of blue [green] indicate units belonging to the fast (gamma) [slow (theta)] subnetwork. Excitatory (inhibitory) units are indicated by triangles (circles). Excitatory (inhibitory) connections are shown as lines ending in triangles (circles). Black (gray) lines indicate intra- (inter-) subnetwork connections, while the magenta line indicates the bottom-up sensory input. B: The model responses to speech are shown in the same format as in Fig. 1B, with the exception of the top subpanel, which illustrates how synaptic short-term depression is implemented in the model, showing the scaled speech envelope (blue line), the dynamical variable representing the fraction of available neurotransmitters (green), and the signal injected to Ge units (magenta), resulting from the product of the first two quantities for each time point. Synaptic depression qualitatively reproduces the large responses to speech onsets and attenuated responses to subsequent syllables observed in the iEEG recordings. C: Trial-averaged power spectra calculated separately for the pre-stimulus and the speech period are shown in the left panel for the model (in cyan and red, respectively) and for the iEEG sample (average across selected channels; shown in gray and black, respectively). Power spectra corresponding to the pre-stimulus (speech) period only are shown for the model and each selected iEEG channel in the middle (right) panel. D: As in Fig. 1E, middle panel, with the addition of the corresponding model result (green line). E: As in Fig. 1E, right panel, with the addition of the corresponding model results (cyan for the pre-stimulus period, red for the speech period). F: Activated and de-activated frequency bands. For each subject and for each frequency value, the p-value from a one-tailed *t*-test between the speech and the baseline power spectra is shown (results from right-tail [left-tail] tests, which indicate higher power during speech [baseline], are shown in red [blue]). Only p-values smaller than p_{crit} (corresponding to a FDR *q* value of 0.05) are shown, and only if at least two adjacent frequency values show concordant effects.

Our goal here was not to assess whether the model tends to lead or lag the iEEG data, but rather to assess the similarity between model and iEEG activity at the single-trial level, and to relate it to the degree of similarity between iEEG subjects, in a spectrally-resolved fashion. Hence, we also considered the opposite ordering, that is, the trial-averaged cross-correlation between each iEEG subject and the model (obtained from the average cross-correlation between the model and each iEEG subject by reversing the time delay axis), and averaged across orderings. Then, the maximum cross-correlation was retained for each subject and frequency value, and was then averaged across subjects to obtain a spectrally-

resolved profile of similarity between model and iEEG (Fig. 3A).

We then assessed how this similarity relates to the average similarity across iEEG subjects (inter-subject variability) by running the same analysis considering every pair of iEEG subjects (Fig. 3B). These analyses revealed that both the model - iEEG similarity and the iEEG - iEEG similarity profiles exhibit a bimodal shape, with high values at low (delta-theta) and high (gamma - high-gamma) frequencies, and low values for intermediate frequencies (in the alpha-beta range) (Fig. 3C and D).

Finally, we assessed the within-subject intertrial similarity in the same way as described above for the inter-subject similarity, but

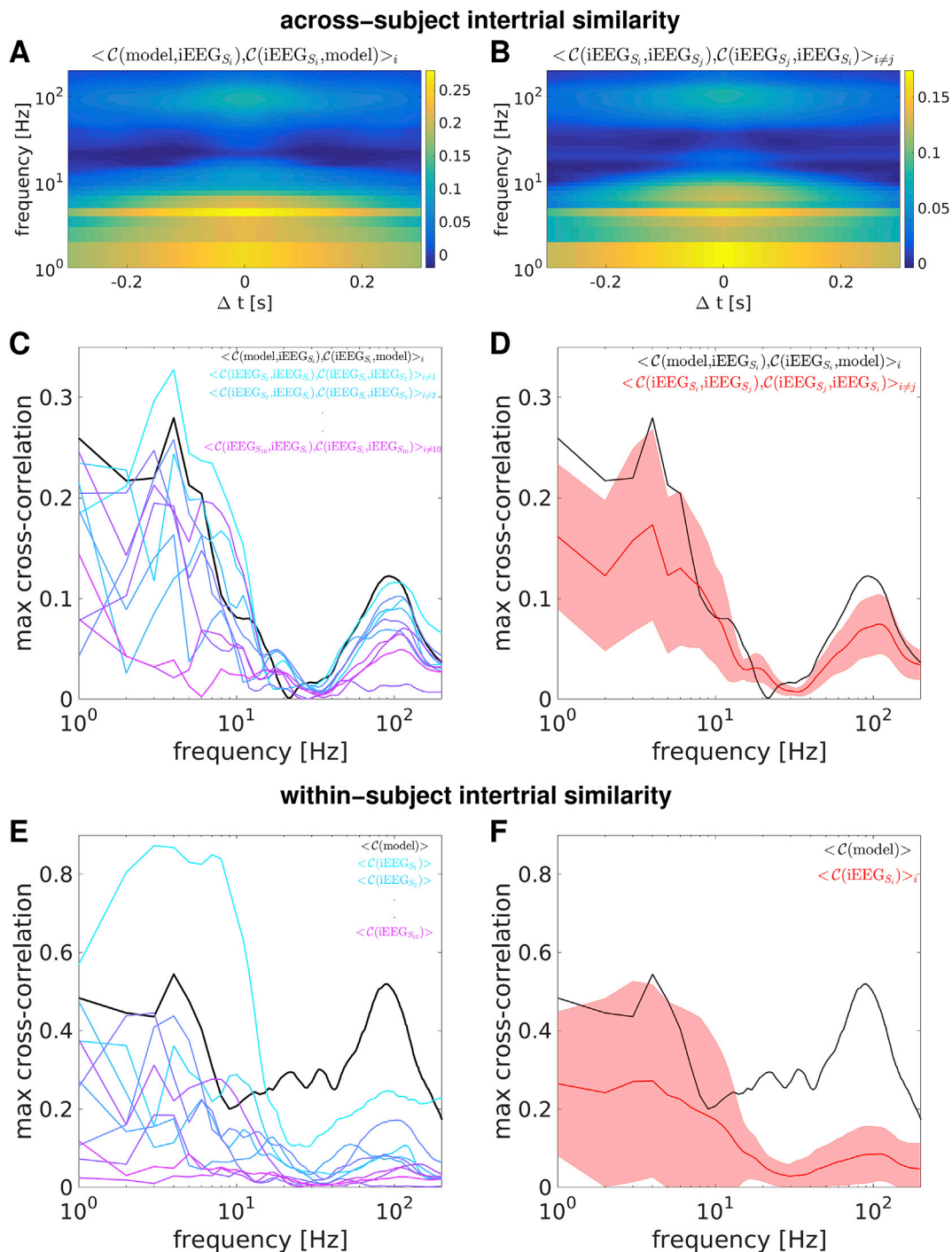


Fig. 3. The spectrally-resolved profile of similarity between the model and the iEEG dataset is comparable to inter-subject similarity. A: Average spectrally-resolved cross-correlation between the model and the iEEG dataset. B: Average spectrally-resolved cross-correlation between pairs of iEEG subjects. C: The spectrally-resolved similarity between the model and the iEEG dataset, obtained by taking the maximum cross-correlation for each frequency from the cross-correlation plots shown in A), is shown in black. Corresponding spectrally-resolved similarity curves between each iEEG subject and the remaining iEEG dataset are shown with colored lines. D: As in C), but the spectrally-resolved similarity curves between each iEEG subject and the remaining iEEG dataset have been averaged across subjects (red line). The red shading indicates the SD across subjects. Both the black and the red curves are bimodal, with high similarity in low (delta-theta) and high (gamma - high-gamma) frequencies, and low similarity for intermediate frequencies (alpha-beta). E: as in C), but lines indicate within-subject intertrial similarity for each subject (colored lines) and for the model (black line). F: as in D), for within-subject intertrial similarities.

computing the average cross-correlation between the $10 \log_{10}$ power trajectories for each pair of trials $(t_i^{S_k}, t_j^{S_k})$, where $i \neq j$, for each subject S_k as well as for the model (Fig. 3E and F). The spectral profile of within-subject intertrial similarity was qualitatively similar to the spectral profile of inter-subject similarity: it also exhibited the same bimodal shape,

albeit similarity values tended to be higher for the within-subject than for the inter-subject similarity, especially in the delta-theta range. Within-subject intertrial similarity in the gamma and high-gamma range was higher in the model than in the iEEG population, suggesting that the model does not accurately reproduce the iEEG trial-by-trial variability in this frequency range.

4.2. Stimulus encoding occurs preferentially at delta-theta and gamma frequencies

To assess whether the speech-induced band-limited power and phase modulations described in the previous section convey information about speech tokens, we performed multivariate classification analyses on the CWT representation of individual trial voltage trajectories, as described in subsection “Multivariate classification analyses”. In particular, we trained and tested multivariate classifiers using single-trial spectro-temporal representations of iEEG signals corresponding to either baseline activity, or activity corresponding to neural processing of a speech token.

We considered all possible within-phrase binary decoding, as well as the binary decoding analyses between corresponding speech tokens in P1 and P2. We further grouped decoding analyses into three categories: detection (speech token vs. baseline), easy discrimination (between different speech tokens, or the same speech token immediately following continuous speech or occurring after a pause), and hard discrimination (between the same speech token occurring at sentence onset in the two phrases).

When considering a broad range of frequency values (between 4 and 200 Hz) simultaneously as multidimensional inputs to the classifiers, every selected channel displayed high and strongly significant decoding

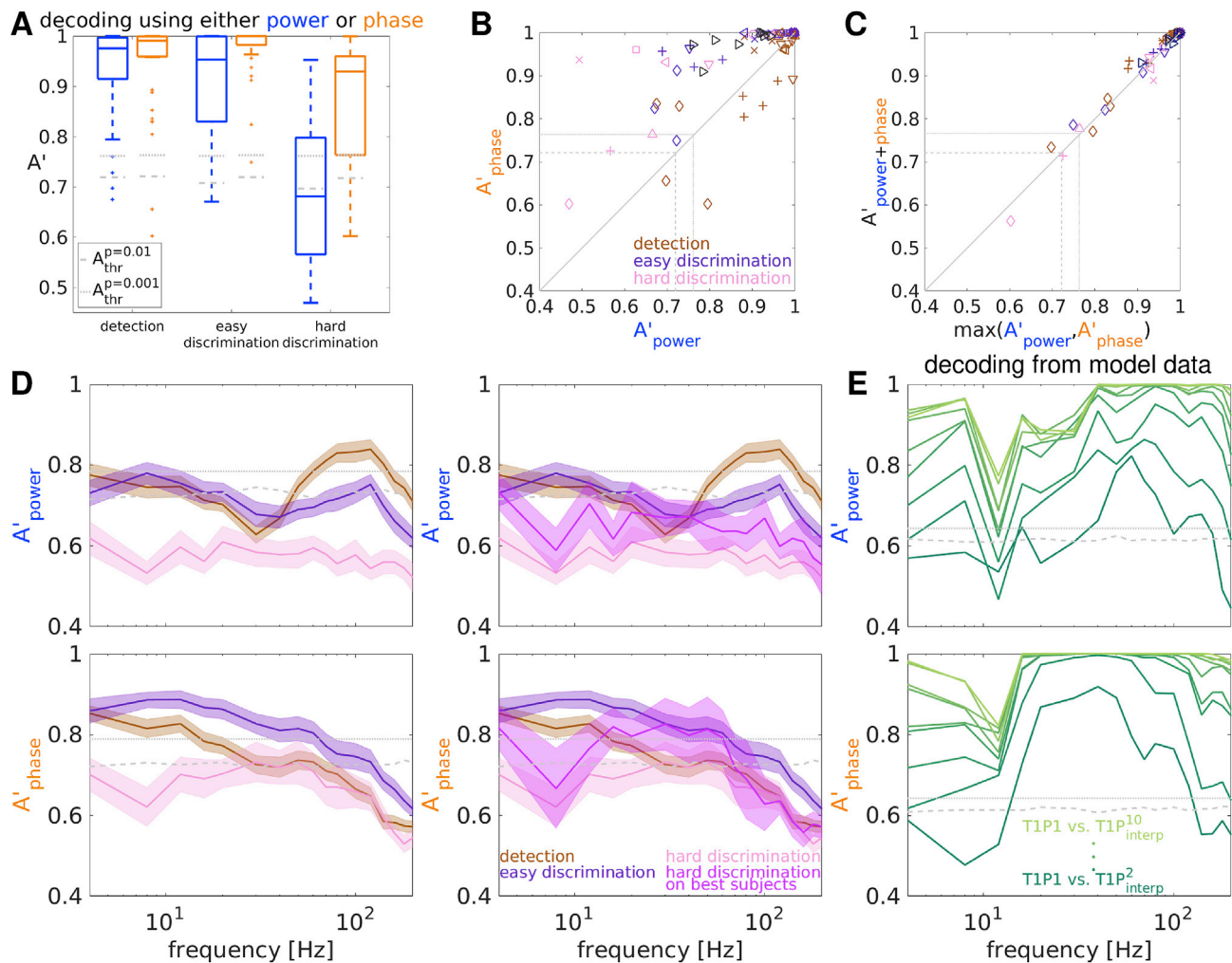


Fig. 4. Detection and discrimination of speech tokens from iEEG data. A: Decoding accuracies A' for three decoding categories using either power (blue bars) or phase (orange bars) values are shown as boxplots: central line indicates the median, box outlines show 25th and 75th percentiles, whiskers extend to the most extreme data points not considered outliers (i.e. up to $1.5 \times$ the interquartile range), and outliers are plotted individually. Decoding accuracies for each decoding analysis are shown for each selected channel separately in Fig. S5. B: Decoding accuracies using phase are plotted against the corresponding decoding accuracies using power for each subject and each classification considered. Colors code for decoding analysis type, as indicated in the legend, while symbol type codes for subject identity as indicated in Fig. S5. Gray dash and dotted lines indicate maximum (across subjects and decoding types) significance thresholds at $p = 0.01$ and 0.001 , respectively, obtained by permutation. Apart from some detection decoding analyses, phase is more informative than power. C: Decoding accuracies using power and phase combined are plotted against the corresponding maximum decoding accuracies using either power or phase separately, for each subject and each classification considered. The combination of power and phase does not seem to convey additional information beyond the one conveyed by the most informative of the two variables (power or phase). D: Spectrally-resolved decoding accuracies using either power (top) or phase (bottom) values, for the three decoding types considered. Power discriminability for detection and easy discrimination exhibits a bimodal profile, with high discriminability for low (delta-theta) and high (gamma - high-gamma) frequency values, and low discriminability for intermediate (alpha-beta) frequencies. Right panels also include an additional trace indicating the average hard discrimination decoding accuracy over the four subjects exhibiting significant hard discrimination decoding accuracy across time and frequency at $p < 0.001$. Phase discriminability for detection and easy discrimination exhibits a low-pass profile in the group average, but a bimodal profile for hard discrimination is unveiled when averaging only over the four subjects that exhibit highly significant hard discrimination decoding accuracy across time and frequency. Spectrally-resolved decoding accuracies for each decoding analysis are shown for each selected channel separately in Fig. S6. E: Spectrally-resolved speech discrimination decoding in the model. Decoding T1P1 vs. $T1P^i_{interop}$ using either patterns of power (top) or phase (bottom) values. Lighter shades of green indicate greater distance from P1. Gray dash and dotted lines indicate significance thresholds at $p = 0.01$ and 0.001 , respectively, obtained by permutation.

accuracies for most binary classifications, which in some case attained perfect discriminability (Fig. S5). Conversely, decoding accuracy in the “control” decoding analysis discriminating baseline activity in the two phrases was low and never significant, apart from one channel where decoding accuracy for phase was weakly significant (just above the $p = 0.01$ significance threshold, uncorrected, Fig. S5). For many of the selected channels, decoding accuracy was high and strongly significant also for the hard discrimination analysis, albeit generally lower than in detection or easy discrimination analyses.

When comparing decoding accuracies using power and phase values, we observed that phase decoding generally yields higher accuracy for every channel and in most of the decoding analyses considered (Fig. 4A–B, $p = 2.47 \cdot 10^{-07}$, two-sided Wilcoxon signed-rank test). The few exceptions comprised exclusively detection decoding analyses, that is, differentiations between speech tokens and baseline activity. In fact, phase and power did not yield statistically different decoding accuracies when the test was restricted to detection decoding analyses ($p = 0.35$), while phase outperformed power for both easy and hard discrimination decoding analyses ($p = 1.49 \cdot 10^{-08}$ and $p = 0.002$, respectively). The advantage of phase over power increased with discrimination difficulty, with the effect size (Cohen’s d) being 0.79 for easy discrimination and 1.21 for hard discrimination decoding. In particular, phase yielded strongly significant ($p < 0.001$) decoding results in 9/10 subjects for the hard discrimination, while power did so only in 4/10 subjects. Primacy of phase information over power information, especially for distinguishing between activity patterns corresponding to different sensory stimuli, is in accordance with several prior reports in auditory, visual and tactile modalities (Luo and Poeppel, 2007; Howard and Poeppel, 2010; Ng et al., 2013; Gross et al., 2013; Schyns et al., 2011; Ronconi et al., 2017; Wang et al., 2018; Baumgarten et al., 2015), in motor signals (Hammer et al., 2013), as well as in more complex paradigms involving working memory and decision making (Rizzuto et al., 2003; Lopour et al., 2013). In our dataset and for the binary decoding analyses we considered, the combination of power and phase did not convey additional information beyond the one conveyed by the most informative between power and phase (Fig. 4C).

Next, we proceeded to a spectrally-resolved evaluation of information content by assessing decoding accuracies for a set of frequency values between 4 and 200 Hz. When decoding using power values, the profile of decoding accuracies for detection and easy discrimination parallels our previous results on speech-induced power changes: we observed a timescale separation between high decoding accuracies at low (delta-theta) and high (gamma - high-gamma) frequencies, with low and non-significant decoding accuracies for intermediate (alpha-theta) frequency values (Fig. 4D, top left panel). Conversely, when decoding using phase values, the highest decoding accuracies were observed for low frequency values in the delta, theta and low-alpha range, and decreased for higher frequencies (Fig. 4D, bottom left panel). However, a bimodal profile with enhanced discriminability at low and high frequencies was also observed in some individual subject (Fig. S6). Interestingly, such a bimodal profile was also observed in the hard discrimination decoding when considering only the four subjects that exhibited significant hard discrimination decoding accuracy across time and frequency at $p < 0.001$. This suggests that, while a broad frequency range might be informative enough for easy discriminations, hard discriminations might unveil the bimodal spectral profile of information content also for discriminations based on phase values.

Variability across trials in our iEEG data is expected to arise from multiple sources: stochastic opening and closing of ionic channels, stochastic synaptic vesicle release, as well as fluctuations in global physiological state which could be related to variations in attention, arousal, mind wandering, and other factors. Our rate model, albeit endowed with a noise term (which represents mostly microscopic stochasticity in the local network), is not expected to fully capture the extent and complexity of the inter-trial variability in the iEEG dataset. In fact, the rate model yields perfect decoding accuracy in all the detection and discrimination

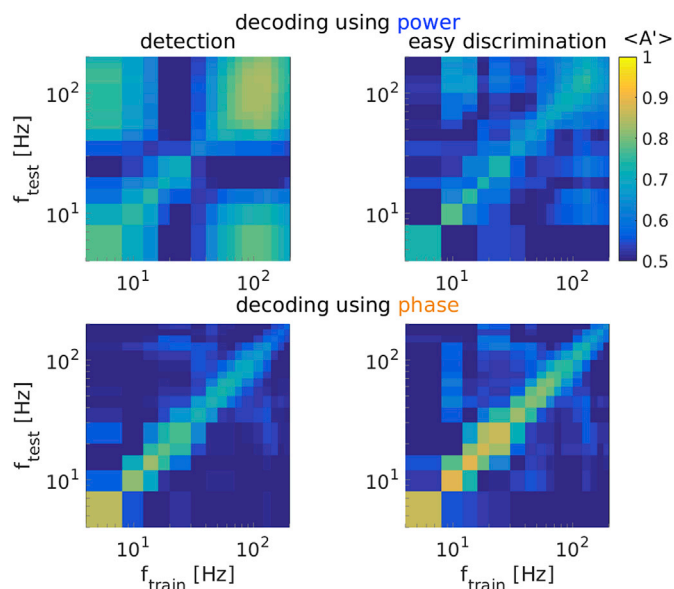


Fig. 5. Specificity vs. generalization in the information conveyed by different frequency bands A: Decoding accuracies A' averaged across subjects and decoding types when training with patterns of power values extracted from frequency f_{train} and testing with patterns of power values extracted from frequency f_{test} . B: As in A), but considering phase instead of power.

decoding analyses presented above.

However, when the model is tested with a challenging discrimination analysis (i.e., discriminating T1P1 vs. $T1P1_{\text{interp}}^i$, where the latter denotes a stimulus that has been obtained by interpolating between P1 and P2, see Methods for details), it also exhibits a bimodal profile of decoding accuracy as a function of frequency (Fig. 4E). Hence, this profile can be explained as a consequence of the synaptic interactions between excitatory and inhibitory neural populations at theta and gamma timescales, as captured in our model.

The analyses presented so far do not inform on whether the information conveyed by any frequency is specific to that frequency, or whether it generalizes to other frequencies. We estimated the degree of specificity vs. generalization in the read-out of the information conveyed by patterns of either power or phase values using a decoding approach where training and test patterns correspond to different frequency values. When testing speech detection, i.e. distinguishing between neural activity during speech vs. baseline, the information conveyed in patterns of power values largely generalizes across low and high frequency values, but not to intermediate frequency values within the alpha-beta range (Fig. 5A, left panel). This result is consistent with our characterization of spectral power changes in response to speech stimulation (Fig. 1), and suggests that speech-related brain activity can be distinguished from baseline simply on the basis of a power increase in relatively broad spectral ranges corresponding to low (delta-theta) or high (gamma - high-gamma) frequencies.

Conversely, the information conveyed by different frequency bands is more specific when considering speech discrimination analyses (distinguishing between neural activity corresponding to different speech segments) (Fig. 5A, right panel), suggesting that the patterns of power values at different frequency bands convey different information about the precise speech segment being heard.

When considering phase decoding, either for speech detection or discrimination, the matrices of decoding accuracies for pairs of ($f_{\text{train}}, f_{\text{test}}$) only exhibit high values for pairs of frequencies close to the diagonal: decoders tested on frequencies that are sufficiently different from those that are used for training perform poorly, which indicates that speech information is encoded in frequency-specific phase patterns (Fig. 5B).

To summarize, these analyses revealed that patterns of power values

convey information for speech *detection* that largely generalizes across frequencies. Conversely, speech *discrimination* relies on precise spectro-temporal patterns that are largely frequency-specific. The specificity of information across frequency values is especially conspicuous when considering phase as the relevant information-bearing spectral quantity, which we demonstrated to convey more information than power for speech discrimination (Fig. 4A and B).

4.3. Model-derived relationships between spectral features and stimulus encoding

The relationships between spectral features of neural activity and its stimulus encoding properties is a topic of great interest in sensory neuroscience (e.g. (Belitski et al., 2008; Belitski et al., 2010; Tsuchiya et al., 2008; Schyns et al., 2011; Gross et al., 2013; Ronconi et al., 2017)). The development of a neuronal network model that exhibits remarkable correspondence to iEEG data in terms of both spectral properties and stimulus encoding enabled us to formulate an approach for relating these two aspects of neural activity, which have hitherto been mostly investigated separately.

In particular, we generated families of models by varying some key parameters from their values in the canonical model shown in previous sections. The parameters that were varied are the bias current injected in either Ge or Te units ($I_{dc_{Ge}}$, $I_{dc_{Te}}$), and the strength of recurrent excitatory connections in either the fast or the slow subnetwork ($J_{Ge_{Ge}}$, $J_{Te_{Te}}$). A linear system analysis indicates that these parameters affect the network dynamics in different ways: an increase in the bias current tends to linearize the dynamics of the corresponding rate units, while an increase in the recurrent excitatory coupling tends to shift the eigenvalues of the linearized system towards the right, hence increasing the oscillatory character of the subnetwork at its natural frequency. We conducted two sets of simulations: in one set, we varied the bias current and the excitatory self-connection strength in the gamma subnetwork in a two-dimensional interval, while keeping all other parameters at their canonical values (as reported in Table S3); in the other set, we varied the bias current and the excitatory self-connection strength in the theta subnetwork in a two-dimensional interval, while keeping all other parameters at their canonical values. Each individual model has been evaluated in terms of both spectral features (in particular, spectral power in the delta-theta and gamma bands) and stimulus encoding properties (considering the challenging discrimination decoding analysis TIP1 vs. TIP^2_{interp}). This enabled us to assess the relationships between spectral features and stimulus encoding in a much larger population than our iEEG dataset. In addition, stimulus encoding could be estimated using a higher number of trials than those available in our iEEG datasets (100 per stimulus type as opposed to 30–50 in iEEG subjects), where trial numbers are limited by experimental and clinical constraints, hence yielding more accurate estimates.

Importantly, these simulations and analyses showed that the information conveyed by patterns of phase values is robustly higher and more stable in the face of parameter variations in comparison to that conveyed by patterns of power values (Fig. S7). They also showed that the strength of recurrent excitatory connections in either the fast or the slow subnetwork ($J_{Ge_{Ge}}$ or $J_{Te_{Te}}$) is the parameter that has the largest effect on the decoding performance. Very high values of the recurrent excitatory connection strength impair decoding accuracy, especially when coupled with strong bias currents to the excitatory units. In particular, when varying gamma subnetwork parameters, median decoding accuracy using phase is always above 0.975, except in the network with the highest J_{EE} and I_{dc_E} tested, where it decreases to 0.951. Likewise, when varying theta subnetwork parameters, median decoding accuracy using phase is always above 0.95, except in the network with the highest J_{EE} and I_{dc_E} tested, where it decreases to 0.908. For the highest J_{EE} and I_{dc_E} values tested, decoding using power yields a median A' below 0.79 in both cases. This impairment in coding properties likely derives from an

excess of baseline synchronization in the corresponding frequency, and, perhaps more critically, a strong oscillatory responsiveness to inputs impairing selective (hence informative) modulations by the speech input: similar speech segments elicit very similar oscillatory patterns (especially in terms of power, but also in terms of phase), close to saturation. It is worth remarking that this decrease in decoding accuracy is more pronounced when considering patterns of power values, suggesting that, during regimes of strong baseline oscillations and strong oscillatory responsiveness, power modulations induced by temporally patterned inputs are poorly reflected in the network output, whereas the precise phase of neural activity still bears substantial information about the stimulus.

Overall, this analysis shows that the model network encodes speech input with a consistent fidelity across a broad range of parameters, with phase consistently bearing more information than power. Stimulus encoding only deteriorates if the self-excitation strength and the depolarizing excitatory drive in a subnetwork are increased to very high levels, close to dynamic instability, yielding a regime that could be likened to a pathological synchronization, i.e., a seizure-like state (Fig. S7 and S8).

5. Discussion

We recorded intracortical responses to pure tones and short sentences from early auditory cortical areas. Based on a combination of anatomical (proximity to Heschl's gyri) and functional (maximum ERP amplitude in response to pure tones) criteria, we selected electrode contacts capturing the neural activity of the primary or secondary auditory cortex and characterized their activity in terms of spectral power and phase-locking during the pre-stimulus baseline period and during speech perception. Pre-stimulus activity was characterized by a prominent power peak in the beta range. Speech caused the disappearance of the beta peak and an increase in power in two separate frequency bands, one located in the low frequency range, around the classically defined delta and theta ranges (~1–10 Hz), and another one in a high frequency range, corresponding to gamma and high-gamma activity (50–150 Hz). Conversely, phase-locking exhibited a low-pass profile at the group level, even if a bimodal profile could be detected in some subjects (Fig. 1).

We assessed the information encoding properties of iEEG activity by performing decoding analyses probing two aspects of speech perception, detection (distinguishing between neural activity corresponding to a speech segment versus pre-stimulus baseline) and discrimination (distinguishing between neural activity corresponding to different speech tokens). We considered either patterns of power values or patterns of phase values, thereby evaluating the encoding properties of these two distinct aspects of rhythmic neural activity, as well as patterns obtained by concatenating power and phase values. In accordance with previous studies, we demonstrated that phase information generally outperforms power information, with the few exceptions belonging exclusively to detection decoding analyses. The advantage of phase over power increased with discrimination difficulty, with only phase yielding strongly significant decoding results in the group average for the hard discrimination. Combining power and phase information did not improve decoding accuracy beyond the level obtained by the most informative of the two variables.

Consistent with power spectral change analyses, decoding analyses showed that information about the presence of speech (detection) conveyed by patterns of power values exhibited a bimodal profile, with a peak in the low (delta-theta) frequency range and another peak in the high (gamma - high-gamma) frequency range. In addition, the information for speech discrimination conveyed by patterns of power values also exhibited such a bimodal profile (Fig. 4D), with the exception of the hard discrimination decoding in some subjects, which did not exceed the significance threshold at any frequency (Fig. S6). Conversely, the information conveyed by phase patterns exhibited a low-pass profile at the group level, even if a bimodal profile with enhanced discriminability at

low and high frequencies was also observed in some individual subject (Fig. S6).

We then addressed whether the neural dynamics reflected in iEEG data could be reproduced by a simple model that assumes that theta and gamma-scale neural activity is underpinned by two interconnected subnetworks, each producing pseudo-rhythmic behavior at distinct timescales. The choice of this architecture is motivated by our own data and is consistent with previous experimental work (Lakatos et al., 2005) and theoretical proposals (Giraud and Poeppel, 2012; Hyafil et al., 2015a, 2015b) suggesting that such subnetworks could colocalize in superficial layers of the auditory cortex (see also (Traub et al., 1996; White et al., 2000; Gloveli et al., 2005) for related work in the hippocampus). We developed a population rate model comprising two subnetworks (each composed of a pair of self and mutually connected excitatory and inhibitory units) with different timescales, one exhibiting fast activity in the gamma - high-gamma range (85.9 Hz), and the other slow activity in the delta-theta range (3.8 Hz), with an inter-subnetwork connectivity pattern implementing a negative feedback loop between the fast and the slow subnetwork.

This simple model exhibits a remarkable similarity to our iEEG channel sample, both in terms of spectral properties at baseline and during speech stimulation (Fig. 2), and in terms of spectrally-resolved information about speech discrimination as measured by decoding analyses (Fig. 4E). These features, and in particular the spectral speech discrimination accuracy profile, do not require fine-tuning of the model parameters, with speech discrimination deteriorating only when the self-excitation strength and the depolarizing excitatory drive in a subnetwork are increased to non-physiological levels, leading to a hyper-synchronized, seizure-like, state.

5.1. Dual timescale processing in auditory cortex

Altogether, our findings are consistent with a dual timescale view of sensory processing in auditory cortex (Giraud and Poeppel, 2012; Teng et al., 2016, 2017; Teng and Poeppel, 2019; Giroud et al., 2020) where stimulus encoding is not uniform across frequency bands, nor does it merely reflect the spectral content of the input, but it is instead realized in distinct frequency bands corresponding to low (typically, delta and theta) and high (gamma and high-gamma) frequency bands. Our approach went beyond classical univariate assessment of speech-induced power changes by investigating the information content that each frequency band conveys about the stimulus using spectrally-resolved decoding. We observed that information about the presence of a stimulus (detection decoding), as well as information about stimulus identity (discrimination decoding), were both predominantly conveyed in low (delta-theta) and high (gamma, high-gamma) frequency bands when we considered power and mostly by low frequencies when phase was taken into account, with high frequencies being also informative in some subjects. Interestingly, such a dual timescale profile of neural coding has also been reported in the visual system (Kayser and König, 2004; Belitski et al., 2008, 2010; Lewis et al., 2016b), suggesting that it might be a general organizational principle of sensory neural processing.

We acknowledge that our decoding results do not unambiguously imply a corresponding functional role in speech processing. In fact, while significant speech decoding can be obtained from large regions of the temporal and frontal lobes, these activity differences do not necessarily signal the neural underpinnings of speech comprehension, but might reflect collateral processes (Bouton et al., 2018, 2019). Causal contributions are best explored with interventional approaches (Panzeri et al., 2017), which are beyond the scope of this work. However, our analysis portrays the informational attributes of neural activity in the first cortical stage of the auditory pathway, which constrains the speech informational structure that is available to downstream regions more closely involved with specific linguistic processing.

Intermediate frequency bands (alpha and beta) also showed power responses to stimulation (the classical beta power reduction), but this

phenomenon was not accompanied by informational activity that could enable reliable stimulus detection or discrimination at the single-trial level using patterns of power values. In addition, activity in these frequency bands was highly variable across individuals, suggesting that activity in this frequency bands is not as strongly driven by sensory stimulation as activity in lower (delta-theta) and higher (gamma - high-gamma) frequencies. Hence, our results lend additional support to the hypothesis that neural activity in the alpha and beta bands might predominantly reflect processes that are not directly involved in the bottom-up encoding of sensory inputs. The low-beta frequency range is indeed thought to be involved in several cognitive aspects of top-down control (Fontolan et al., 2014; Bastos et al., 2015a,b; Park et al., 2015). In the context of speech perception, alpha-beta activity might be predominantly involved with attention (Strauß et al., 2014; Wöstmann et al., 2016, 2017; Dimitrijevic et al., 2017), working memory (Obleser et al., 2012; Wilsch et al., 2015; Wilsch and Obleser, 2016), listening effort (Obleser et al., 2012; Wöstmann et al., 2015; Miles et al., 2017), intelligibility (Obleser and Weisz, 2012; Becker et al., 2013; Pefkou et al., 2017), as well as semantic (Wöstmann et al., 2015), syntactic (Bastiaansen et al., 2010; Bastiaansen and Hagoort, 2015; Lewis et al., 2016a) or temporal predictions (Arnal and Giraud, 2012; Arnal et al., 2015; Morillon and Baillet, 2017; Tavano et al., 2017; Biau and Kotz, 2018); these cognitive aspects (experimentally dissociable only to a partial extent) are consistent with a mostly top-down role.

5.2. Circuit motifs

Substantial progress in our mechanistic understanding of neuronal network function, and on the relationship between structure and activity in brain networks, came from the exploration of circuit motifs and of how they can implement basic neural computations (Womelsdorf et al., 2014; Silver, 2010). In this work we focused on a circuit motif at the inter-subnetwork level, namely the negative feedback loop between the fast and the slow subnetworks implemented by the Ge to Te and the Te to Gi connections. We showed how this inter-subnetwork connectivity configuration results in a remarkable correspondence to iEEG data in terms of both spectral properties and stimulus encoding.

Additional circuit motifs are also likely to be realized in the neuronal circuitry of early auditory cortical areas, and might further contribute to shaping the spectral responses to stimuli as well as to their encoding. For example, feed-forward inhibition has been described in ascending projections from the thalamo-recipient layer L4 to L2/3 in the mouse primary auditory cortex (Li et al., 2014). In our simulations, we observed that feed-forward inhibition (implemented by conveying the bottom-up speech envelope input to the Gi unit as well, in addition to the Ge unit) can modulate the relative proportion of high-vs. low-frequency power responses to stimulation, with increased feed-forward inhibition resulting in lower low-vs. high-frequency power change ratio. However, our canonical model does not include feed-forward inhibition for the sake of parsimony, as an exhaustive characterization of the effect of feed-forward inhibition in our model is beyond the scope of this work.

5.3. Spectral biases and scale-free spectral profiles

Our iEEG recordings tended to exhibit a scale-free power spectrum profile across a fairly broad frequency range (spectral power decreased as $1/f^\alpha$), as commonly observed in recordings from the brain and other complex systems (e.g. (He et al., 2010)). Scale-free behavior was prevalent during periods without auditory stimulation (i.e. pre-stimulus periods, Fig. 1C,E, cyan traces). However, it is worth noting that approximate scale-freeness in the power spectrum profile can coexist with band-limited power changes, here expressed as speech-induced synchronization in the delta-theta and gamma ranges and desynchronization in the alpha-beta range. Rather than being realized by a “rotation” of the power spectrum profile as suggested in (Podvalny et al., 2015), our data indicate that opposing changes in average power across neighboring

frequency bands (in particular, between speech-induced beta desynchronization and speech-induced gamma synchronization) result from a speech-induced modification of the power spectrum profile towards an inverted double “S” shape in a broader frequency range, corresponding to delta-theta and gamma activation, and alpha-beta deactivation (Fig. 1C, E, red traces). Our decoding results further confirm the presence of a timescale separation in our dataset, with low (delta-theta) and high (gamma - high-gamma) frequencies yielding high and strongly significant decoding accuracies, while intermediate (alpha-beta) frequencies yielded low, often non-significant, decoding performance.

Interestingly, the frequency ranges that exhibited speech-induced activation and speech-induced deactivation varied across subjects, with the transition between the low-frequency activation and the mid-frequency deactivation varying between 7 and 22 Hz, and the transition between the mid-frequency deactivation and the high-frequency activation varying between 20 and 40 Hz (Fig. 2F and S3). Band-limited processes that occur with a consistent structure in all subjects, but at different frequency values, are diluted when data is pooled across subjects, resulting in spectral profiles that are closer to scale-free curves (compare the group-averaged power spectra shown in Fig. 2C, black and gray lines, with the single-subject power spectra shown in Fig. 1C and S2). Hence, we expect band-limited processes to be more readily detected with detailed analyses that independently consider individual channels and behavioral conditions, as enabled by high signal-to-noise ratio recording modalities such as intracranial electrophysiology. Conversely, scale-freeness is more likely to emerge when data is pooled across channels or behavioral conditions, or when a narrower frequency range is considered, e.g. as in (Miller et al., 2009; Podvalny et al., 2015).

5.4. Potential pre-cortical contributions to cortical oscillations

Oscillatory signatures recorded in the cortex partly originates in the thalamus or in thalamocortical interactions (Steriade et al., 1993; Hughes et al., 2004; Crunelli and Hughes, 2010; Saleem et al., 2017; Li et al., 2017). However, cortical microcircuitry also has the potential to generate rhythmic oscillations in multiple frequency bands and might do so, at least in some contextual and neuromodulatory conditions (Buzsáki and Chrobak, 1995; Buzsáki and Draguhn, 2004; Buzsáki, 2006; Wang, 2010). In particular, prominent delta, theta and gamma activity has been localized in superficial layers of the auditory cortex on the basis of current-source density analyses of multi-layered recordings in monkeys (Lakatos et al., 2005) and humans (Halgren et al., 2018).

While the current model implements cortical oscillations as locally generated and assumes a simple, non-oscillatory scheme for pre-cortical stages of auditory processing, the physiological interpretation of each of the model components could be modified to accommodate a role of pre-cortical structures in the generation of rhythmic activity as detected by intracranial EEG in auditory cortex, for example by assuming that theta units could reflect not only local cortical activity but rather the composite activity of neural populations distributed across the cortex and key subcortical structures of the auditory pathway such as the inferior colliculus and the auditory thalamus, and maybe the hippocampus, the cerebellum and the basal ganglia (Kotz and Schwartze, 2010).

5.5. Modeling considerations

Population rate models based on the Wilson-Cowan formalism have been successfully applied to characterize ECoG recordings. In particular, they constitute a simple and suitable description for the $1/f^\alpha$ spectral profile of human ECoG activity and its modulation by task engagement (Chaudhuri et al., 2017). Such models have also been successfully applied to the description of feedforward and feedback processing, and their spectral fingerprints, in a large-scale model of the primate visual system (Mejias et al., 2016).

In contrast to previous proposals (e.g. Loebel et al., 2007; Hyafil et al., 2015a; Stringer et al., 2016; Yarden and Nelken, 2017; Harper

et al., 2016; Rahman et al., 2019; Chambers et al., 2019)), our network model does not incorporate a tonotopic structure. This approach is motivated by the psychophysical observation that speech comprehension can be accomplished based mostly on temporal cues (Shannon et al., 1995). Our goal was not to implement a detailed model that incorporated all the known features of the auditory system, but rather to build a minimal model that could reproduce the key features observed in iEEG datasets, most prominently the presence of discontinuous timescales that characterizes cortical responses to auditory speech stimuli.

In contrast to most previously proposed model of auditory processing (Viemeister and Wakefield, 1991; Moore, 2003; Teng et al., 2016), ours is instantiated at a biophysical level, and relies on specific synaptic interactions between a fast, gamma-scale subnetwork, and a slow, theta-scale subnetwork. As our understanding of auditory processing evolves from phenomenological, abstract models to biophysically grounded implementations, we can develop theories that can be tested not only at the psychophysical level, but also at the level of neural activity patterns. In addition, biophysical models pave the way towards bridging between the neuronal microscale, the iEEG mesoscale and perceptual levels, and bear the potential to relate microscopic anomalies observed in conditions such as autism or dyslexia to corresponding mesoscale anomalies, and, ultimately, to the perceptual deficits in speech and/or reading comprehension that underlie some of their debilitating consequences (e.g. (Wang and Krystal, 2014)).

A recent study proposed a rate network model that bears a close similarity with the theta subnetwork considered here (Doelling et al., 2019). The model could reproduce MEG phase responses to periodic stimuli better than a competing model based on evoked responses, providing evidence for endogenous oscillations in auditory cortex. This work differs from the contribution of Doelling et al. along a few dimensions: *i*) by exploiting the high temporal resolution of intracranial EEG, we could also assess high-frequency activity up to 200 Hz, *ii*) we assessed brain activity both in the pre-stimulus baseline as in response to speech, and *iii*) we could propose a rate network model that captures the spectral profile of neural activity both in the pre-stimulus baseline period as in response to speech in a broad frequency range that extends from 1 to 200 Hz. Our model also highlights the importance of noise in yielding realistic spectral profiles of brain activity.

Biophysical models encapsulate, in an explicit and directly testable format, a set of hypotheses about the underlying neural mechanisms that are deemed sufficient for the emergence of a set of observed features of interest; here, the spectral power features observed at baseline and in response to speech in early auditory cortex, and their information encoding characteristics. From this perspective, the biological features that are not appropriately reproduced by the model are also of great interest, since they highlight the limitations of the current modeling approach and might indicate promising directions for future developments.

With respect to the spectral power profile at baseline and its modulation by speech stimulation, a key feature that our model does not reproduce is the presence of a power peak in the beta range, which is reduced by speech stimulation. In contrast to rhythmic activity at lower (theta, alpha) or higher (gamma) frequencies (Buzsáki and Chrobak, 1995; Wang, 2010), the biophysical mechanisms underlying beta rhythmicity are not well understood. A recent study based on *in vivo* data from humans, monkeys and mice suggested that beta activity could originate from near-simultaneous volleys of excitatory synaptic drive impinging on proximal and distal dendrites of pyramidal neurons (Sherman et al., 2016), which could conceivably originate from the lemniscal and the nonlemniscal thalamus, respectively. Conversely, previous *in vitro* work proposed a mechanism for the emerge of beta rhythmicity, and in particular its slower variant (with spectral peak ~ 15 Hz), based on the concatenation of gamma and fast beta across superficial and deep cortical layers, respectively (Kramer et al., 2008).

Regardless of the mechanistic details of beta activity generation, prominent activity in this frequency band has been associated to a status-

quo regime where changes in sensory/cognitive states are not expected (Engel and Fries, 2010) and local information processing and propagation are mechanistically inhibited (Sherman et al., 2016), maybe in favor of a more global operating mode (Kopell et al., 2000, 2011). Neural activity in the beta range exhibits a prominent top-down directional bias across the neural hierarchy (Buschman and Miller, 2007; van Kerkoerle et al., 2014; Fontolan et al., 2014; Bastos et al., 2015a,b; Michalareas et al., 2016), lending strong support to the hypothesis that beta oscillations as recorded in early auditory cortical areas might be largely inherited from higher cortical areas. From this perspective, it is not surprising that our computational model, designed to capture the essential neuronal microcircuitry of the superficial layers of a single cortical area, does not reproduce the beta power spectrum peak at baseline and its reduction upon speech presentation. Future modeling work will address the interplay of bottom-up sensory inputs and top-down expectations or linguistic knowledge by developing a hierarchical neuronal network with spiking neurons and realistic laminar organization (e.g. (Bastos et al., 2012; Lee et al., 2013; Lee et al., 2015)). Importantly, the development of neuronal network models that incorporate linguistic knowledge could pave the way towards the discernment and a mechanistic understanding of the micro- and meso-scale origins of the macroscopic features in brain oscillation (Obleser and Weisz, 2012; Becker et al., 2013; Pefkou et al., 2017), coherence with speech (Luo and Poeppel, 2007; Peelle and Davis, 2012; Doelling et al., 2014; Kösem and van Wassenhove, 2017), and scale-free amplitude dynamics (Borges et al., 2018) that selectively accompany speech comprehension.

Another feature of interest that our model only reproduces to a partial extent is inter-trial variability, a feature of neural activity that is directly related to information encoding. Variability across trials in our iEEG data is expected to arise from multiple sources: stochastic opening and closing of ionic channels, stochastic synaptic vesicle release, as well as fluctuations in global physiological state which could be related to variations in attention, arousal, mind wandering, and other factors. Our rate model, albeit endowed with a noise term (representing mostly microscopic stochasticity in the local network), is not expected to fully capture the extent and complexity of the inter-trial variability in the iEEG dataset. Notably, and in spite of its inability to quantitatively match iEEG inter-trial variability, our model network qualitatively reproduces the bimodal spectral profile of decoding accuracy for speech discrimination observed in our iEEG dataset (Fig. 4E), when discrimination is assessed by considering activity patterns corresponding to highly similar speech inputs. Further modeling work could incorporate additional terms with slowly varying dynamics accounting for specific aspects of brain physiology that are expected to affect neural variability, such as sensory adaptation and shifts in attention and arousal (maybe by applying similar strategies as employed in (Goris et al., 2014) for the characterization of variability in spiking activity), hence potentially elucidating the influence of each of these processes on inter-trial variability.

6. Conclusion

We showed that neural activity in early auditory cortex in humans, as directly recorded by intracranial EEG, displays separable timescales at low (delta-theta) and high (gamma - high-gamma) frequencies, where speech stimulation increases spectral power and where reproducibility and information for speech detection and discrimination is higher than in the intermediate alpha - beta range. We further showed that a simple rate model comprising two excitatory/inhibitory subnetworks with different timescales and interconnected with a negative feedback loop between the fast and the slow subnetwork is sufficient to account for a vast majority of the experimental observations. Interestingly, the model failed to reproduce the baseline beta peak and its speech-induced desynchronization, which is consistent with the recent proposals suggesting that beta activity might have different underpinnings, possibly involving dynamic coordination across the cortical hierarchy.

After acknowledging that the main spectral and informational

features of neural activity at baseline and in response to speech are well captured by a simple rate model with four neuronal populations, one might feel inclined to believe that the apparent complexity of brain activity could be reduced to a fairly low-dimensional dynamical system. We would suggest a different interpretation, not incompatible with but complementary to the former: if these simple models can capture broad features of the iEEG, these broad features might only convey a degraded version of the information presented at finer resolutions. While this could be enough for a rough reconstruction of the stimulus, it would not allow for re-creating the complex and highly informative phenomenological experience that speech and music can elicit in healthy, awake humans.

It is therefore of interest to assess the extent to which simple models, which embody well-specified hypotheses about the neurobiological substrates (in particular, the connectivity diagram between excitatory and inhibitory populations and the time constants that govern their interactions), can account for various spectral measures of iEEG activity, advancing from relatively gross features such as ERP amplitude, spectral power and phase-locking characteristics, to progressively finer measures related to variability and information encoding and transformation. Hence, we believe that the interaction of advanced data analysis techniques, addressing both dynamical and informational features of brain activity, and computational modeling at multiple levels of resolution can offer a promising toolkit for improving our understanding of the inner workings of the brain, and in particular for clarifying the spatial resolution that maximizes the causal effectiveness of neural activity (Hoel et al., 2013), arguably a necessary prerequisite for a mechanistic understanding of brain function.

Data and code availability

The data and code used in this study are available in the OSF (Open Science Framework) repository at <https://osf.io/kxu4m>.

CRediT authorship contribution statement

Fabiano Baroni: Conceptualization, Investigation, Methodology, Validation, Visualization, Writing - original draft, Writing - review & editing. **Benjamin Morillon:** Resources, Writing - review & editing. **Agnès Trébuchon:** Resources, Writing - review & editing. **Catherine Liégeois-Chauvel:** Resources, Writing - review & editing. **Itsaso Olsagasti:** Funding acquisition, Methodology, Writing - review & editing. **Anne-Lise Giraud:** Conceptualization, Methodology, Project administration, Supervision, Funding acquisition, Writing - review & editing.

Acknowledgments

We thank all subjects for their participation in the study. This work was funded by a project grant from the Swiss National Science Foundation to ALG (#320030-163040). The computations were performed at the University of Geneva on the Baobab cluster.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuroimage.2020.116882>.

References

- Arnal, L.H., Doelling, K.B., Poeppel, D., 2015. Delta-beta coupled oscillations underlie temporal prediction accuracy. *Cerebr. Cortex* 25, 3077–3085 (New York, N.Y. : 1991).
- Arnal, L.H., Giraud, A.L.L., 2012. Cortical oscillations and sensory predictions. *Trends Cognit. Sci.* 16, 390–398.
- Barbour, D.L., Callaway, E.M., 2008. Excitatory local connections of superficial neurons in rat auditory cortex. *J. Neurosci.* 28, 11174–11185. Publisher: Society for Neuroscience Section: Articles.

- Baroni, F., van Kempen, J., Kawasaki, H., Kovach, C.K., Oya, H., Howard, M.A., Adolphs, R., Tsuchiya, N., 2017. Intracranial markers of conscious face perception in humans. *Neuroimage* 162, 322–343.
- Bastiaansen, M., Hagoort, P., 2015. Frequency-based segregation of syntactic and semantic unification during online sentence level language comprehension. *J. Cognit. Neurosci.* 27, 2095–2107.
- Bastiaansen, M., Magyari, L., Hagoort, P., 2010. Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *J. Cognit. Neurosci.* 22, 1333–1347.
- Bastos, A.M., Litvak, V., Moran, R., Bosman, C.A., Fries, P., Friston, K.J., 2015a. A DCM study of spectral asymmetries in feedforward and feedback connections between visual areas V1 and V4 in the monkey. *Neuroimage* 108, 460–475.
- Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., Friston, K.J., 2012. Canonical microcircuits for predictive coding. *Neuron* 76, 695–711.
- Bastos, A.M.M., Vezoli, J., Bosman, C.A.A., Schoffelen, J.M.M., Oostenveld, R., Dowdall, J.R.R., De Weerd, P., Kennedy, H., Fries, P., 2015b. Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron* 85, 390–401.
- Baumgarten, T.J., Schnitzler, A., Lange, J., 2015. Beta oscillations define discrete perceptual cycles in the somatosensory domain. *Proc. Natl. Acad. Sci. U.S.A.* 112, 12187–12192.
- Becker, R., Pefkou, M., Michel, C.M., Hervais-Adelman, A.G., 2013. Left temporal alpha-band activity reflects single word intelligibility. *Front. Syst. Neurosci.* 7.
- Belitski, A., Gretton, A., Magri, C., Murayama, Y., Montemurro, M.A., Logothetis, N.K., Panzeri, S., 2008. Low-frequency local field potentials and spikes in primary visual cortex convey independent visual information. *J. Neurosci. : Off. J. Soc. Neurosci.* 28, 5696–5709.
- Belitski, A., Panzeri, S., Magri, C., Logothetis, N.K., Kayser, C., 2010. Sensory information in local field potentials and spikes from visual and auditory cortices: time scales and frequency bands. *J. Comput. Neurosci.* 29, 533–545.
- Benítez-Burraco, A., Murphy, E., 2016. The oscillopathic nature of language deficits in autism: from genes to language evolution. *Front. Hum. Neurosci.* 10.
- Besle, J., Schevon, C.A., Mehta, A.D., Lakatos, P., Goodman, R.R., McKhann, G.M., Emerson, R.G., Schroeder, C.E., 2011. Tuning of the human neocortex to the temporal dynamics of attended events. *J. Neurosci. : Off. J. Soc. Neurosci.* 31, 3176–3185.
- Bhalla, U.S., Iyengar, R., 1999. Emergent properties of networks of biological signaling pathways. *Science* 283, 381–387 (Publisher: American Association for the Advancement of Science Section: Report).
- Biau, E., Kotz, S.A., 2018. Lower beta: a central coordinator of temporal prediction in multimodal speech. *Front. Hum. Neurosci.* 12.
- Boemio, A., Fromm, S., Braun, A., Poeppel, D., 2005. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat. Neurosci.* 8, 389–395.
- Borges, A.F.T.F.T., Giraud, A.L.L., Mansvelder, H.D., Linkenkaer-Hansen, K., 2018. Scale-free amplitude modulation of neuronal oscillations tracks comprehension of accelerated speech. *J. Neurosci. : Off. J. Soc. Neurosci.* 38, 710–722.
- Bouton, S., Chambon, V., Golestani, N., Giraud, A.L., 2019. Integrating Hypothesis- and Data-Driven Approaches in Speech and Language Neuroscience Preprint. *PsyArXiv*.
- Bouton, S., Chambon, V., Tyrand, R., Guggisberg, A.G., Seeck, M., Karkar, S., van de Ville, D., Giraud, A.L., 2018. Focal versus distributed temporal cortex activity for speech sound category assignment. *Proc. Natl. Acad. Sci. Unit. States Am.* 115, E1299–E1308.
- Breska, A., Deouell, L.Y., 2017. Neural mechanisms of rhythm-based temporal prediction: delta phase-locking reflects temporal predictability but not rhythmic entrainment. *PLoS Biol.* 15.
- Brugge, J.F., Nourski, K.V., Oya, H., Reale, R.A., Kawasaki, H., Steinschneider, M., Howard, M.A., 2009. Coding of repetitive transients by auditory cortex on Heschl's gyrus. *J. Neurophysiol.* 102, 2358–2374.
- Buschman, T.J., Miller, E.K., 2007. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science (New York, N.Y.)* 315, 1860–1862.
- Buzsáki, G., Chrobak, J.J., 1995. Temporal structure in spatially organized neuronal ensembles: a role for interneuronal networks. *Curr. Opin. Neurobiol.* 5, 504–510.
- Buzsáki, G., 2006. *Rhythms of the Brain*, 1 edition. Oxford University Press.
- Buzsáki, G., Draguhn, A., 2004. Neuronal oscillations in cortical networks. *Science (New York, N.Y.)* 304, 1926–1929.
- Calderone, D.J., Lakatos, P., Butler, P.D., Castellanos, F.X., 2014. Entrainment of neural oscillations as a modifiable substrate of attention. *Trends Cognit. Sci.* 18, 300–309.
- Chambers, J.D., Elgueta, D., Fritz, J.B., Shamma, S.A., Burkitt, A.N., Grayden, D.B., 2019. Computational neural modeling of auditory cortical receptive fields. *Front. Comput. Neurosci.* 13, 28.
- Chaudhuri, R., He, B.J., Wang, X.J.J., 2017. Random recurrent networks near criticality capture the broadband power distribution of human ECoG dynamics. *Cerebr. Cortex* 1–13 (New York, N.Y. : 1991).
- Clayton, M.S., Yeung, N., Cohen Kadosh, R., 2015. The roles of cortical oscillations in sustained attention. *Trends Cognit. Sci.* 19, 188–195.
- Cravo, A.M., Rohenkohl, G., Wyart, V., Nobre, A.C., 2011. Endogenous modulation of low frequency oscillations by temporal expectations. *J. Neurophysiol.* 106, 2964–2972.
- Crunelli, V., Hughes, S.W., 2010. The slow (<1 Hz) rhythm of non-REM sleep: a dialogue between three cardinal oscillators. *Nat. Neurosci.* 13, 9–17.
- Dimitrijevic, A., Smith, M.L., Kadis, D.S., Moore, D.R., 2017. Cortical alpha oscillations predict speech intelligibility. *Front. Hum. Neurosci.* 11.
- Ding, N., Simon, J.Z., 2012. Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11854–11859.
- Doelling, K.B., Arnal, L.H., Ghitza, O., Poeppel, D., 2014. Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* 85 (Pt 2), 761–768.
- Doelling, K.B., Assaneo, M.F., Bevilacqua, D., Pesaran, B., Poeppel, D., 2019. An oscillator model better predicts cortical entrainment to music. *Proc. Natl. Acad. Sci. Unit. States Am.* 116, 10113–10121.
- Edelstein-Keshet, L., 2005. *Mathematical Models in Biology* Classics in Applied Mathematics. Society for Industrial and Applied Mathematics.
- Edwards, E., Chang, E.F., 2013. Syllabic (~2-5 Hz) and fluctuation (~1-10 Hz) ranges in speech and auditory processing. *Hear. Res.* 305, 113–134.
- Elhilali, M., Fritz, J.B., Klein, D.J., Simon, J.Z., Shamma, S.A., 2004. Dynamics of precise spike timing in primary auditory cortex. *J. Neurosci. : Off. J. Soc. Neurosci.* 24, 1159–1172.
- Engel, F., Fries, P., Engel, A.K., Fries, P., 2010. Beta-band oscillations—signalling the status quo? *Curr. Opin. Neurobiol.* 20, 156–165.
- Fenton, A.A., 2015. Excitation-inhibition discoordination in rodent models of mental disorders. *Biol. Psychiatr.* 77, 1079–1088.
- Fontolan, L., Morillon, B., Liegeois-Chauvel, C., Giraud, A.L.L., 2014. The contribution of frequency selectivity to hierarchical information processing in the human auditory cortex. *Nat. Commun.* 5.
- Fries, P., Reynolds, J.H., Rorie, A.E., Desimone, R., 2001. Modulation of oscillatory neuronal synchronization by selective visual attention. *Science (New York, N.Y.)* 291, 1560–1563.
- Gao, R., Penzes, P., 2015. Common mechanisms of excitatory and inhibitory imbalance in schizophrenia and autism spectrum disorders. *Curr. Mol. Med.* 15, 146–167.
- Giraud, A.L.L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517.
- Giroud, J., Trébuchon, A., Schön, D., Marquis, P., Liegeois-Chauvel, C., Poeppel, D., Morillon, B., 2020. Asymmetric sampling in human auditory cortex reveals spectral processing hierarchy. *PLoS Biol.* 18, e3000207 (Publisher: Public Library of Science).
- Gloveli, T., Dugladze, T., Saha, S., Monyer, H., Heinemann, U., Traub, R.D., Whittington, M.A., Buhl, E.H., 2005. Differential involvement of oriens/pyramidal interneurons in hippocampal network oscillations in vitro. *J. Physiol.* 562, 131–147.
- Gonzalez-Burgos, G., Lewis, D.A., 2008. GABA neurons and the mechanisms of network oscillations: implications for understanding cortical dysfunction in schizophrenia. *Schizophr. Bull.* 34, 944–961.
- Goris, R.L., Movshon, J.A., Simoncelli, E.P., 2014. Partitioning neuronal variability. *Nat. Neurosci.* 17, 858–865.
- Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., Garrod, S., 2013. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol.* 11.
- Halgren, M., Fabó, D., Ulbert, I., Madsen, J.R., Eröss, L., Doyle, W.K., Devinsky, O., Schomer, D., Cash, S.S., Halgren, E., 2018. Superficial slow rhythms integrate cortical processing in humans. *Sci. Rep.* 8.
- Hammer, J., Fischer, J., Ruescher, J., Schulze-Bonhage, A., Aertens, A., Ball, T., 2013. The role of ECoG magnitude and phase in decoding position, velocity, and acceleration during continuous motor behavior. *Front. Neurosci.* 7.
- Harper, N.S., Schoppe, O., Willmore, B.D., Cui, Z., Schnupp, J.W., King, A.J., 2016. Network receptive field modeling reveals extensive integration and multi-feature selectivity in auditory cortical neurons. *PLoS Comput. Biol.* 12.
- He, B.J., 2014. Scale-free brain activity: past, present, and future. *Trends Cognit. Sci.* 18, 480–487.
- He, B.J., Zempel, J.M., Snyder, A.Z., Raichle, M.E., 2010. The temporal structures and functional significance of scale-free brain activity. *Neuron* 66, 353–369.
- Hoel, E.P., Albantakis, L., Tononi, G., 2013. Quantifying causal emergence shows that macro can beat micro. *Proc. Natl. Acad. Sci. Unit. States Am.* 110, 19790–19795.
- Howard, M.F., Poeppel, D., 2010. Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *J. Neurophysiol.* 104, 2500–2511.
- Hughes, S.W., Lörincz, M., Cope, D.W., Blethyn, K.L., Kékesi, K.A., Parri, H.R., Juhász, G., Crunelli, V., 2004. Synchronized oscillations at alpha and theta frequencies in the lateral geniculate nucleus. *Neuron* 42, 253–268.
- Hyafil, A., Fontolan, L., Kabdebon, C., Gutkin, B., Giraud, A.L.L., 2015a. Speech encoding by coupled cortical theta and gamma oscillations. *eLife* 4.
- Hyafil, A., Giraud, A.L.L., Fontolan, L., Gutkin, B., 2015b. Neural cross-frequency coupling: connecting architectures, mechanisms, and functions. *Trends Neurosci.* 38, 725–740.
- Jerbi, K., Freyermuth, S., Dalal, S., Kahane, P., Bertrand, O., Berthoz, A., Lachaux, J.P.P., 2009. Saccade related gamma-band activity in intracerebral EEG: dissociating neural from ocular muscle activity. *Brain Topogr.* 22, 18–23.
- Jochaut, D., Lehongre, K., Saitovitch, A., Devauchelle, A.D.D., Olasagasti, I., Chabane, N., Zilbovicius, M., Giraud, A.L.L., 2015. Atypical coordination of cortical oscillations in response to speech in autism. *Front. Hum. Neurosci.* 9.
- Kayser, C., König, P., 2004. Stimulus locking and feature selectivity prevail in complementary frequency ranges of V1 local field potentials. *Eur. J. Neurosci.* 19, 485–489.
- King, J.R.R., Dehaene, S., 2014. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cognit. Sci.* 18, 203–210.
- Klimesch, W., 2012. α -band oscillations, attention, and controlled access to stored information. *Trends Cognit. Sci.* 16, 606–617.
- Kopell, N., Ermentrout, G.B., Whittington, M.A., Traub, R.D., 2000. Gamma rhythms and beta rhythms have different synchronization properties. *Proc. Natl. Acad. Sci. Unit. States Am.* 97, 1867–1872.
- Kopell, N., Whittington, M.A., Kramer, M.A., 2011. Neuronal assembly dynamics in the beta1 frequency range permits short-term memory. *Proc. Natl. Acad. Sci. Unit. States Am.* 108, 3779–3784.
- Kösem, A., van Wassenhove, V., 2017. Distinct contributions of low- and high-frequency neural oscillations to speech comprehension. *Lang. Cognit. Neurosci.* 32, 536–544.

- Kotz, S.A., Ravignani, A., Fitch, W.T., 2018. The evolution of rhythm processing. *Trends Cognit. Sci.* 22, 896–910.
- Kotz, S.A., Schwartz, M., 2010. Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends Cognit. Sci.* 14, 392–399.
- Kovach, C.K., Tsuchiya, N., Kawasaki, H., Oya, H., Howard, M.A., Adolphs, R., 2011. Manifestation of ocular-muscle EMG contamination in human intracranial recordings. *Neuroimage* 54, 213–233.
- Kramer, M.A., Roopun, A.K., Carracedo, L.M., Traub, R.D., Whittington, M.A., Kopell, N.J., 2008. Rhythm generation through period concatenation in rat somatosensory cortex. *PLoS Comput. Biol.* 4, e1000169+.
- Lachaux, J.P., Rodriguez, E., Martinerie, J., Varela, F.J., 1999. Measuring phase synchrony in brain signals. *Hum. Brain Mapp.* 8, 194–208.
- Lakatos, P., Shah, A.S., Knuth, K.H., Ulbert, I., Karmos, G., Schroeder, C.E., 2005. An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J. Neurophysiol.* 94, 1904–1911.
- Lee, J.H., Whittington, M.A., Kopell, N.J., 2013. Top-down beta rhythms support selective attention via interlaminar interaction: a model. *PLoS Comput. Biol.* 9, e1003164+.
- Lee, J.H., Whittington, M.A., Kopell, N.J., 2015. Potential mechanisms underlying inter cortical signal regulation via cholinergic neuromodulators. *J. Neurosci. : Off. J. Soc. Neurosci.* 35, 15000–15014.
- Lehongre, K., Morillon, B., Giraud, A.L.L., Ramus, F., 2013. Impaired auditory sampling in dyslexia: further evidence from combined fMRI and EEG. *Front. Hum. Neurosci.* 7.
- Lehongre, K., Ramus, F., Villiermet, N., Schwartz, D., Giraud, A.L.L., 2011. Altered low-frequency sampling in auditory cortex accounts for the three main facets of dyslexia. *Neuron* 72, 1080–1090.
- Lewis, A.G., Schoffelen, J.M.M., Schriefers, H., Bastiaansen, M., 2016a. A predictive coding perspective on beta oscillations during sentence-level language comprehension. *Front. Hum. Neurosci.* 10.
- Lewis, C.M., Bosman, C.A., Brunet, N.M., Lima, B., Roberts, M.J., Womelsdorf, T., de Weerd, P., Neuenschwander, S., Singer, W., Fries, P., 2016b. Two Frequency Bands Contain the Most Stimulus-Related Information in Visual Cortex. *bioRxiv*.
- Li, G., Henriquez, C.S., Fröhlich, F., 2017. Unified thalamic model generates multiple distinct oscillations with state-dependent entrainment by stimulation. *PLoS Comput. Biol.* 13.
- Li, LyY., Ji, XyY., Liang, F., Li, YtT., Xiao, Z., Tao, H.W., Zhang, L.I., 2014. A feedforward inhibitory circuit mediates lateral refinement of sensory representation in upper layer 2/3 of mouse primary auditory cortex. *J. Neurosci. : Off. J. Soc. Neurosci.* 34, 13670–13683.
- Liegeois-Chauvel, C., Musolino, A., Chauvel, P., 1991. Localization of the primary auditory area in man. *Brain : J. Neurol.* 114 (Pt 1A), 139–151.
- Loebel, A., Nelken, I., Tsodyks, M., 2007. Processing of sounds by population spikes in a model of primary auditory cortex. *Front. Neurosci.* 1, 197–209.
- Lopour, B.A., Tavassoli, A., Fried, I., Ringach, D.L., 2013. Coding of information in the phase of local field potentials within human medial temporal lobe. *Neuron* 79, 594–606.
- Lundqvist, M., Rose, J., Herman, P., Brincat, S.L., Buschman, T.J., Miller, E.K., 2016. Gamma and beta bursts underlie working memory. *Neuron* 90, 152–164.
- Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010.
- Luo, H., Poeppel, D., 2012. Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. *Front. Psychol.* 3.
- McGinley, M.J., Vinck, M., Reimer, J., Batista-Brito, R., Zaghera, E., Cadwell, C.R., Tolia, A.S., Cardin, J.A., McCormick, D.A., 2015. Waking state: rapid variations modulate neural and behavioral responses. *Neuron* 87, 1143–1161.
- Mejias, J.F., Murray, J.D., Kennedy, H., Wang, X.J., 2016. Feedforward and feedback frequency-dependent interactions in a large-scale laminar network of the primate cortex. *Sci. Adv.* 2, e1601335+.
- Michalareas, G., Vezoli, J., van Pelt, S., Schoffelen, J.M.M., Kennedy, H., Fries, P., 2016. Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron* 89, 384–397.
- Miles, K., McMahon, C., Boisvert, I., Ibrahim, R., de Lissa, P., Graham, P., Lyxell, B., 2017. Objective assessment of listening effort: coregistration of pupillometry and EEG. *Trends Hear.* 21.
- Miller, K.J., Sorensen, L.B., Ojemann, J.G., den Nijs, M., 2009. Power-law scaling in the brain surface electric potential. *PLoS Comput. Biol.* 5, e1000609+.
- Moore, B.C.J., 2003. Temporal integration and context effects in hearing. *J. Phonetics* 31, 563–574.
- Morillon, B., Baillet, S., 2017. Motor origin of temporal predictions in auditory attention. *Proc. Natl. Acad. Sci. U.S.A.* 114.
- Morillon, B., Hackett, T.A., Kajikawa, Y., Schroeder, C.E., 2015. Predictive motor control of sensory dynamics in auditory active sensing. *Curr. Opin. Neurobiol.* 31, 230–238.
- Morillon, B., Lehongre, K., Frackowiak, R.S., Ducours, A., Kleinschmidt, A., Poeppel, D., Giraud, A.L.L., 2010. Neurophysiological origin of human brain asymmetry for speech and language. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18688–18693.
- Morillon, B., Liégeois-Chauvel, C., Arnal, L.H., Bénar, C.G.G., Giraud, A.L.L., 2012. Asymmetric function of theta and gamma activity in syllable processing: an intracortical study. *Front. Psychol.* 3.
- Ng, B.S.W.S., Logothetis, N.K., Kayser, C., 2013. EEG phase patterns reflect the selectivity of neural firing. *Cerebr. Cortex* 23, 389–398 (New York, N.Y. : 1991).
- Nourski, K.V., Steinschneider, M., Rhone, A.E., Oya, H., Kawasaki, H., Howard, M.A., McMurray, B., 2015. Sound identification in human auditory cortex: differential contribution of local field potentials and high gamma power as revealed by direct intracranial recordings. *Brain Lang.* 148, 37–50.
- Obleser, J., Weisz, N., 2012. Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. *Cerebr. Cortex* 22, 2466–2477 (New York, N.Y. : 1991).
- Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., Maess, B., 2012. Adverse listening conditions and memory load drive a common α oscillatory network. *J. Neurosci. : Off. J. Soc. Neurosci.* 32, 12376–12383.
- Panzeri, S., Harvey, C.D., Piasini, E., Latham, P.E., Fellin, T., 2017. Cracking the neural code for sensory perception by combining statistics, intervention, and behavior. *Neuron* 93, 491–507.
- Park, H., Ince, R.A., Schyns, P.G., Thut, G., Gross, J., 2015. Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Curr. Biol. : CB* 25, 1649–1653.
- Peelle, J.E., Davis, M.H., 2012. Neural oscillations carry speech rhythm through to comprehension. *Front. Psychol.* 3.
- Pefkou, M., Arnal, L.H., Fontolan, L., Giraud, A.L.L., 2017. θ -Band and β -band neural activity reflects independent syllable tracking and comprehension of time-compressed speech. *J. Neurosci. : Off. J. Soc. Neurosci.* 37, 7930–7938.
- Pizzarelli, R., Cherubini, E., 2011. Alterations of GABAergic signaling in autism spectrum disorders. *Neural Plast.* 2011, 297153.
- Podvalny, E., Noy, N., Harel, M., Bickel, S., Chechik, G., Schroeder, C.E., Mehta, A.D., Tsodyks, M., Malach, R., 2015. A unifying principle underlying the extracellular field potential responses in the human cortex. *J. Neurophysiol.* 114, 505–519.
- Poeppel, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun.* 41, 245–255.
- Rahman, M., Willmore, B.D.B., King, A.J., Harper, N.S., 2019. A dynamic network model of temporal receptive fields in primary auditory cortex. *PLoS Comput. Biol.* 15, e1006618.
- Ramamoorthi, K., Lin, Y., 2011. The contribution of GABAergic dysfunction to neurodevelopmental disorders. *Trends Mol. Med.* 17, 452–462.
- Ravignani, A., Dalla Bella, S., Falk, S., Kello, C.T., Noriega, F., Kotz, S.A., 2019. Rhythm in speech and animal vocalizations: a cross-species perspective. *Ann. N. Y. Acad. Sci.* 1453, 79–98.
- Rene, T., D'Ari, R., 1990. Biological Feedback, 1 edition edition. CRC Press, Boca Raton.
- Rifkin, R., Yeo, G., Poggio, T., 2003. Regularized least-squares classification, 190, 131–154. *Nato Science Series Sub Series III Computer and Systems Sciences*.
- Rizzuto, D.S., Madsen, J.R., Bromfield, E.B., Schulze-Bonhage, A., Seelig, D., Aschenbrenner-Scheibe, R., Kahana, M.J., 2003. Reset of human neocortical oscillations during a working memory task. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7931–7936.
- Rohenkohl, G., Nobre, A.C., 2011. α oscillations related to anticipatory attention follow temporal expectations. *J. Neurosci. : Off. J. Soc. Neurosci.* 31, 14076–14084.
- Ronconi, L., Oosterhof, N.N., Bonmassar, C., Melcher, D., 2017. Multiple oscillatory rhythms determine the temporal organization of perception. *Proc. Natl. Acad. Sci. Unit. States Am.* 114, 13435–13440.
- Roopun, A.K., Cunningham, M.O., Racca, C., Alter, K., Traub, R.D., Whittington, M.A., 2008. Region-specific changes in gamma and beta2 rhythms in NMDA receptor dysfunction models of schizophrenia. *Schizophr. Bull.* 34, 962–973.
- Ross, B., Miyazaki, T., Thompson, J., Jamali, S., Fujioka, T., 2014. Human cortical responses to slow and fast binaural beats reveal multiple mechanisms of binaural hearing. *J. Neurophysiol.* 112, 1871–1884.
- Saleem, A.B., Lien, A.D., Krumin, M., Haider, B., Rosón, M.R.R., Ayaz, A., Reinhold, K., Busse, L., Carandini, M., Harris, K.D., 2017. Subcortical source and modulation of the narrowband gamma oscillation in mouse visual cortex. *Neuron* 93, 315–322.
- Saleh, M., Reimer, J., Penn, R., Ojakangas, C.L., Hatsopoulos, N.G., 2010. Fast and slow oscillations in human primary motor cortex predict oncoming behaviorally relevant cues. *Neuron* 65, 461–471.
- Schyns, P.G., Thut, G., Gross, J., 2011. Cracking the code of oscillatory activity. *PLoS Biol.* 9, e1001064+.
- Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. *Science (New York, N.Y.)* 270, 303–304.
- Sherman, M.A., Lee, S., Law, R., Haegens, S., Thorn, C.A., Hämaläinen, M.S., Moore, C.I., Jones, S.R., 2016. Neural mechanisms of transient neocortical beta rhythms: converging evidence from humans, computational modeling, monkeys, and mice. *Proc. Natl. Acad. Sci. U.S.A.* 113, E4885–E4894.
- Silver, R.A., 2010. Neuronal arithmetic. *Nat. Rev. Neurosci.* 11, 474–489.
- Simon, D.M., Wallace, M.T., 2016. Dysfunction of sensory oscillations in autism spectrum disorder. *Neurosci. Biobehav. Rev.* 68, 848–861.
- Soltész, F., Szűcs, D., Leong, V., White, S., Goswami, U., 2013. Differential entrainment of neuroelectric delta oscillations in developmental dyslexia. *PLoS One* 8.
- Stefanics, G., Hangya, B., Hernádi, I., Winkler, I., Lakatos, P., Ulbert, I., 2010. Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *J. Neurosci. : Off. J. Soc. Neurosci.* 30, 13578–13585.
- Steriade, M., McCormick, D.A., Sejnowski, T.J., 1993. Thalamocortical oscillations in the sleeping and aroused brain. *Science (New York, N.Y.)* 262, 679–685.
- Strauß, A., Wöstmann, M., Obleser, J., 2014. Cortical alpha oscillations as a tool for auditory selective inhibition. *Front. Hum. Neurosci.* 8.
- Stringer, C., Pachitariu, M., Steinmetz, N.A., Okun, M., Bartho, P., Harris, K.D., Sahani, M., Lesica, N.A., 2016. Inhibitory control of correlated intrinsic variability in cortical networks. *eLife* 5.
- Sun, L., Castellanos, N., Grützner, C., Koethe, D., Rivolta, D., Wibrall, M., Kranaster, L., Singer, W., Leweke, M.F., Uhlhaas, P.J., 2013. Evidence for dysregulated high-frequency oscillations during sensory processing in medication-naïve, first episode schizophrenia. *Schizophr. Res.* 150, 519–525.
- Szikla, G., Bouvier, G., Hori, T., Petrov, V., 1977. *Angiography of the Human Brain Cortex*. Springer, Berlin.
- Talairach, J., Tournoux, P., 1988. *Co-planar Stereotaxic Atlas of the Human Brain: 3-dimensional Proportional System*. Thieme Medical Publishers, New York.

- Tallon-Baudry, C., Bertrand, O., Delpuech, C., Pernier, J., 1996. Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. *J. Neurosci.* 16, 4240–4249.
- Tavano, A., Schröger, E., Kotz, S.A., 2017. Low Beta Power Reflects Perceived Temporal Probabilities for Uncertain Future Events. *bioRxiv*, p. 234070.
- Teng, X., Poeppel, D., 2019. Theta and gamma bands encode acoustic dynamics over wide-ranging timescales. *Cerebr. Cortex* bhz263.
- Teng, X., Tian, X., Poeppel, D., 2016. Testing multi-scale processing in the auditory system. *Sci. Rep.* 6.
- Teng, X., Tian, X., Rowland, J., Poeppel, D., 2017. Concurrent temporal channels for auditory processing: oscillatory neural entrainment reveals segregation of function at different scales. *PLoS Biol.* 15.
- Thomson, A.M., Deuchars, J., 1994. Temporal and spatial properties of local circuits in neocortex. *Trends Neurosci.* 17, 119–126.
- Thomson, A.M., Bannister, A.P., 2003. Interlaminar connections in the neocortex. *Cerebr. Cortex* 13, 5–14 (Publisher: Oxford Academic).
- Traub, R.D., Whittington, M.A., Colling, S.B., Buzsáki, G., Jefferys, J.G., 1996. Analysis of gamma rhythms in the rat hippocampus in vitro and in vivo. *J. Physiol.* 493 (Pt 2), 471–484.
- Tsodyks, M., Pawelzik, K., Markram, H., 1998. Neural networks with dynamic synapses. *Neural Comput.* 10, 821–835.
- Tsuchiya, N., Kawasaki, H., Oya, H., Howard, M.A., Adolphs, R., 2008. Decoding face information in time, frequency and space from direct intracranial recordings of the human brain. *PLoS One* 3, e3892+.
- Uhlhaas, P.J., Roux, F., Rodriguez, E., Rotarska-Jagiela, A., Singer, W., 2010. Neural synchrony and the development of cortical networks. *Trends Cognit. Sci.* 14, 72–80.
- Uhlhaas, P.J., Singer, W., 2012. Neuronal dynamics and neuropsychiatric disorders: toward a translational paradigm for dysfunctional large-scale networks. *Neuron* 75, 963–980.
- van Kerkoerle, T., Self, M.W., Dagnino, B., Gariel-Mathis, M.A., Poort, J., van der Togt, C., Roelfsema, P.R., 2014. Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc. Natl. Acad. Sci. Unit. States Am.* 111, 14332–14341.
- Viemeister, N.F., Wakefield, G.H., 1991. Temporal integration and multiple looks. *J. Acoust. Soc. Am.* 90, 858–865.
- Voytek, B., Knight, R.T., 2015. Dynamic network communication as a unifying neural basis for cognition, development, aging, and disease. *Biol. Psychiatr.* 77, 1089–1097.
- Wang, X.J.J., 2010. Neurophysiological and computational principles of cortical rhythms in cognition. *Physiol. Rev.* 90, 1195–1268.
- Wang, X.J.J., Krystal, J.H., 2014. Computational psychiatry. *Neuron* 84, 638–654.
- Wang, Y., Wang, P., Yu, Y., 2018. Decoding English alphabet letters using EEG phase information. *Front. Neurosci.* 12.
- White, J.A., Banks, M.I., Pearce, R.A., Kopell, N.J., 2000. Networks of interneurons with fast and slow gamma-aminobutyric acid type A (GABAA) kinetics provide substrate for mixed gamma-theta rhythm. *Proc. Natl. Acad. Sci. U.S.A.* 97, 8128–8133.
- Wilsch, A., Henry, M.J., Herrmann, B., Maess, B., Obleser, J., 2015. Alpha oscillatory dynamics index temporal expectation benefits in working memory. *Cerebr. Cortex* 25, 1938–1946 (New York, N.Y. : 1991).
- Wilsch, A., Obleser, J., 2016. What works in auditory working memory? A neural oscillations perspective. *Brain Res.* 1640, 193–207.
- Wilson, H.R., Cowan, J.D., 1972. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* 12, 1–24.
- Womelsdorf, T., Valiante, T.A., Sahin, N.T., Miller, K.J., Tiesinga, P., 2014. Dynamic circuit motifs underlying rhythmic gain control, gating and integration. *Nat. Neurosci.* 17, 1031–1039.
- Wöstmann, M., Herrmann, B., Maess, B., Obleser, J., 2016. Spatiotemporal dynamics of auditory attention synchronize with speech. *Proc. Natl. Acad. Sci. U.S.A.* 113, 3873–3878.
- Wöstmann, M., Herrmann, B., Wilsch, A., Obleser, J., 2015. Neural alpha dynamics in younger and older listeners reflect acoustic challenges and predictive benefits. *J. Neurosci. : Off. J. Soc. Neurosci.* 35, 1458–1467.
- Wöstmann, M., Lim, S.J.J., Obleser, J., 2017. The human neural alpha response to speech is a proxy of attentional control. *Cerebr. Cortex* 27, 3307–3317 (New York, N.Y. : 1991).
- Yarden, T.S., Nelken, I., 2017. Stimulus-specific adaptation in a recurrent network model of primary auditory cortex. *PLoS Comput. Biol.* 13.
- Zion Golumbic, E.M., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., McKhann, G.M., Goodman, R.R., Emerson, R., Mehta, A.D., Simon, J.Z., Poeppel, D., Schroeder, C.E., 2013. Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron* 77, 980–991.