



DOCTORAT EN NEUROSCIENCES  
des Universités de Genève  
et de Lausanne



UNIVERSITÉ DE GENÈVE

FACULTÉ DES SCIENCES

Professeur Anne-Lise Giraud, directrice de thèse  
Professeur Boris Gutkin, co-directeur de thèse

## ENCODING SPEECH THROUGH BRAIN RHYTHMS

THÈSE  
Présentée à la  
Faculté des Sciences

de l'Université de Genève

pour obtenir le grade de  
Docteur en Neurosciences

par

**Lorenzo FONTOLAN**

de Milan, Italie

Thèse N° 148

Université de Genève

Doctoral Thesis

ENCODING SPEECH THROUGH BRAIN RHYTHMS:  
A COMPUTATIONAL MODEL OF AUDITORY CORTEX

Lorenzo Fontolan  
*Université de Genève*

## **Publications References**

- Fontolan, L., Krupa, M., Hyafil, A., & Gutkin, B. (2013). Analytical insights on theta-gamma coupled neural oscillators. *Journal of Mathematical Neuroscience*, 3(1), 16. doi:10.1186/2190-8567-3-16
- Fontolan, L., Morillon, B., Liegeois-Chauvel, C., & Giraud, A.-L. (2014). The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. *Nature Communications*, 5(May), 4694. doi:10.1038/ncomms5694
- Hyafil, A., Fontolan, L., Kabdebon, C., Gutkin, B., & Giraud, A.-L. (2015). Speech demultiplexing and encoding by coupled cortical theta and gamma oscillations. *eLife* (in press).

## **Abstract**

Brain oscillations are dynamic entities, rapidly varying in time and frequency that are extensively recorded in the mammalian brain. The aim of this thesis is to investigate the role of cortical rhythms in human auditory cortex during speech perception, both with computational and experimental methods. The first part of the doctoral work consisted in developing a neural microcircuit model of nested oscillations for early auditory processing, involving a fast Gamma rhythm (30-100 Hz) coupled to a slow Theta rhythm (3-8 Hz). The model is capable of parsing speech into its constituents (i.e. syllables), and extracting the syllabic information for latter categorization. In fact, Theta oscillations flexibly track the quasi-periodic syllabic content of speech, and temporally arrange Gamma spikes so that phonemic information can be efficiently encoded. We further employed a set of advanced tools from dynamical system theory in order to uncover the mathematical description of the neural circuit we have implemented in the simulations. The second part of this dissertation relates to the transmission of information across the auditory hierarchy in the brain, using two segregated frequency channels. We analysed EEG signals from intracranial recordings in humans using cross-frequency coupling and Granger causality. We found that Bottom-up information is dominated by Gamma oscillations, while Top-Down flow is conveyed through Theta oscillations. We have also shown that both flows fluctuate over time at a rate of about  $\sim 1\text{-}3$  Hz, which suggests that sensory information is conveyed using distinct frequencies and via discrete time windows.

## Résumé

Les oscillations cérébrales sont des entités dynamiques, rapidement variables dans le temps et la fréquence, qui sont enregistrées largement dans le cerveau des mammifères. L'objectif de cette thèse est d'étudier le rôle des rythmes corticaux dans le cortex auditif humain lors de la perception de la parole, à la fois avec les méthodes computationnelles et expérimentales. La première partie de la thèse de doctorat a consisté en l'élaboration d'un modèle de microcircuit de neurones d'oscillations mêlées pour le premiers stades du traitement auditif, composé d'un rythme rapide Gamma (30-100 Hz) couplé à un rythme lent Theta (3-8 Hz). Le modèle est capable de décomposer la forme d'onde vocale en ses constituants (i.e. syllabes), et extraire les informations syllabique pour les catégoriser. En fait, les oscillations Theta suivent le contenu syllabique quasi-périodique de la parole de manière flexible, et organisent temporellement les spikes Gamma afin que l'information phonémique peut être efficacement codé. Nous avons également utilisé un ensemble d'outils avancés à partir de la théorie des systèmes dynamiques, afin de découvrir la description mathématique du circuit neuronal que nous avons mis en place dans les simulations. La deuxième partie de cette thèse concerne la transmission d'informations à travers la hiérarchie auditif du cerveau, en utilisant deux canaux de fréquences distincts. Nous avons analysé les signaux EEG à partir d'enregistrements intracrâniens chez les humains, en utilisant le couplage entre fréquences et la causalité de Granger. Nous avons découvert que l'information Bottom-up est dominé par des oscillations Gamma, tandis que les flux Top-Down est transporté à travers des oscillations Theta. Nous avons également montré que les suppléants de l'inférence causale dans le temps entre les deux flux, à un taux de ~ 1-3 Hz, ce qui suggère que la transmission de l'information procèdent par des fenêtres de temps discrets à des fréquences distinctes. Nous avons également montré que les deux flux fluctuent dans le temps à un taux d'environ ~ 1-3 Hz, ce qui suggère que l'information sensorielle est transmise en utilisant des fréquences distinctes et à travers des fenêtres de temps discrets.

## Contents

<b>Publications References .....</b>	<b>iii</b>
<b>Abstract .....</b>	<b>iv</b>
<b>Résumé .....</b>	<b>v</b>
<b>Contents .....</b>	<b>vi</b>
<b>1 Introduction overview.....</b>	<b>1</b>
<b>2 Brain rhythms .....</b>	<b>2</b>
<b>2.1 Historical background and recent developments .....</b>	<b>2</b>
<b>2.2 Oscillations and brain functions .....</b>	<b>3</b>
2.2.1 1/f background and the meaning of spectral peaks .....	5
2.2.2 Synchrony and cross-frequency coupling.....	6
2.2.3 Phase-amplitude coupling and causality .....	9
<b>3 Computational models of brain oscillations .....</b>	<b>12</b>
<b>3.1 Single neuron models.....</b>	<b>12</b>
<b>3.2 Neuronal networks: circuitry for oscillations .....</b>	<b>14</b>
<b>3.3 Nested brain rhythms .....</b>	<b>17</b>
<b>4 Neural substrates of speech processing .....</b>	<b>19</b>
<b>4.1 The auditory pathway in a nutshell .....</b>	<b>20</b>
<b>4.2 Physiology of the auditory system.....</b>	<b>21</b>
<b>4.3 Computational models of the auditory system .....</b>	<b>22</b>
<b>4.4 Speech representation and processing in the brain.....</b>	<b>23</b>
<b>4.5 Multiplexing and oscillations .....</b>	<b>25</b>

<b>5</b>	<b>Information flows in sensory cortex.....</b>	<b>26</b>
5.1	The connectionist approach .....	27
5.2	Predictive coding.....	28
5.3	Oscillations and information passing.....	29
5.4	Summary of hypotheses .....	30
<b>6</b>	<b>Description of the articles .....</b>	<b>31</b>
6.1	Article 1: Speech encoding by coupled cortical theta and gamma oscillations	31
6.1.1	Summary of results .....	31
6.2	Article 2: Analytical Insights on Theta-Gamma Coupled Neural Oscillators	32
6.2.1	Summary of results .....	32
6.3	Article 3: The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex .....	33
6.3.1	Summary of results .....	33
<b>7</b>	<b>Discussion.....</b>	<b>33</b>
7.1	Foreword.....	33
7.2	Modeling speech processing in auditory cortex .....	34
7.3	Predictions and perception .....	37
<b>8</b>	<b>Conclusions and future directions .....</b>	<b>40</b>
<b>9</b>	<b>Appendices .....</b>	<b>42</b>
A.	Circular-to-linear correlation.....	42
B.	Single neuron models.....	42
B.1.	Integrate-and-fire neuron and the canonical model.....	42
B.2.	The effect of inhibition and the birth of a bifurcation .....	43
B.3.	Phase response curves .....	44

<b>10</b>	<b>Bibliography.....</b>	<b>44</b>
<b>11</b>	<b>Articles .....</b>	<b>58</b>
<b>11.1</b>	<b>Speech encoding by coupled cortical theta and gamma oscillations</b>	<b>58</b>
<b>11.2</b>	<b>Analytical insights on Theta-Gamma coupled neural oscillators....</b>	<b>104</b>
<b>11.3</b>	<b>The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex .....</b>	<b>125</b>

## 1 Introduction overview

Temporal dynamics is often both a blessing and curse for physicists. Whether it concerns the evolution of the wave function of some subatomic particle, the motion of a comet crossing our solar system or the rate of growth of a population in biology, it is hardly easy to deal with the flow of time in equations. This dissertation has a lot to do with time and synchronization, and how these two aspects might be important for the brain and, in particular, for the understanding of speech.

In the first Chapter of this dissertation we will go over almost a century of research into brain rhythms, the synchronized waves of activity that are ubiquitously observed in brain recordings. Although their presence has been linked to an astounding amount of cognitive functions and mechanisms, their role remains unclear and many neuroscientists still believe that they are an epiphenomenon. In this regard, computational and analytical models can be extremely useful in tying down experimental results to precise predictions and verify the consistency of a hypothesis. Theoretical analyses do not only help the understanding of observed phenomena, but also, most importantly, can highlight the actual advantages for the brain to use a particular organization strategy, such as the use of brain rhythms, to perform its tasks. In particular, simple models are often the most appealing because, as easy to understand, they can be easily manipulated and understood. To introduce our theoretical work on coupled oscillations (Fontolan et al., 2013), we present the most relevant mathematical models of neuronal oscillation in Chapter 2, starting with a tutorial on the basic models of single neurons and finishing with the state-of-the-art models of coupled oscillations in the brain. In Chapter 3 we will see how the processing of speech in auditory cortex is a privileged system to investigate the potential active role of oscillations, given the remarkable matching between timescales in speech and in cortical rhythms recorded locally. We will review the experimental evidence that motivated our computational work on speech parsing by means of Theta (3-8 Hz) and Gamma (30-100 Hz) frequency oscillations (Hyafil et al., 2015). Finally, in Chapter 4, we will see how the role of oscillations is not limited to the decoding of sensory stimuli, such as speech waveforms, but appears to be linked to a more global organizational principle of the brain, i.e. the transmission of information from one brain area to another. We found, by analyzing data recorded intracranially in humans, that two distinct frequency channels are employed by the brain: Gamma band for stimulus-driven, bottom-up communication, and Beta (13-30 Hz) band for anticipatory, top-down processes (Fontolan et al., 2014). A thorough reflection on the theoretical and experimental results of this dissertation is deferred to the Discussion. For the moment, let us start with an historical overview of research on brain rhythms.

## 2 Brain rhythms

*Pacha:* What happened?

*Old Man:* Well, I threw off the Emperor's groove.

*Pacha:* What?

*Old Man:* His groove! The rhythm in which he lives his life. His pattern of behavior. I threw it off. And the Emperor had me thrown out the window.

*Pacha:* Oh, really? I'm supposed to see him today.

*Old Man:* Don't throw off his groove!

*Pacha:* Oh, okay.

*Old Man:* Bewaaare, the grooove.

*Pacha:* Hey, are you gonna be all right?

*Old Man:* Grooove...

*The Emperor's New Groove*, Disney (2000)

From whatever angle we look at it, the human brain strikes us as being an incredibly complex system. Inside approximately 1300 grams of dense, bulky biological matter, there are about a hundred families of neurons, the basic units of computation, for a total of 86 billions (Azevedo et al., 2009), and more than  $10^{14}$  estimated synapses, the functional contacts between brain cells, bearing a total of 100 different kind of neurotransmitters each with its own dynamics. Neuronal spikes travel along about 176000 km of myelinated axons (average length 10 cm) to reach other neurons (Marner et al., 2003) at the speed of ~0.1-100 meters per second (de Callataÿ, 1992). Taken together, these facts indicate a great variety of delays and transmission times, which lead to the kind of noise-like neuronal activity that is usually observed (Dale and Kandel, 1993; Hawkins et al., 1993; Murphy, 1997). And yet, in spite of this intricate complexity (Koch, 1999), neurons are not at all careless about each other: their activity becomes coordinated, giving rise to periods of synchronized firing followed by quasi-silent epochs. These alternations are ubiquitous in the brain and can be observed using several different neural data recordings techniques, provided that the latter have a good temporal resolution (Ward, 2003; Buzsáki and Draguhn, 2004).

In this chapter I will review the key experimental facts relevant to the study of brain oscillations, followed by the possible computational mechanism that could generate such rhythmic activity and its potential function(s) in cognitive processing.

### 2.1 Historical background and recent developments

Rhythmic neuronal discharges have been documented in the early 20th century by Hans Berger, who is the inventor of the electroencephalography (EEG) technique and was the first to detect electrical oscillations in the human brain (Berger, 1929). Thereafter, brain oscillations have been studied across many different species and through a miscellany of invasive and noninvasive techniques, starting from the aforementioned EEG, at the level of the scalp (Tallon-Baudry et al., 1999) and at the subcranial level (Kahana et al., 1999), but they have also been found in magnetoencephalography signals (MEG) (Siegel et al., 2012), local field potentials (LFP) (Bragin et al., 1995; Jutras et al., 2013), single unit activity (SUA) and multi unit activity (MUA) both *in vivo* (Sanchez-Vives and

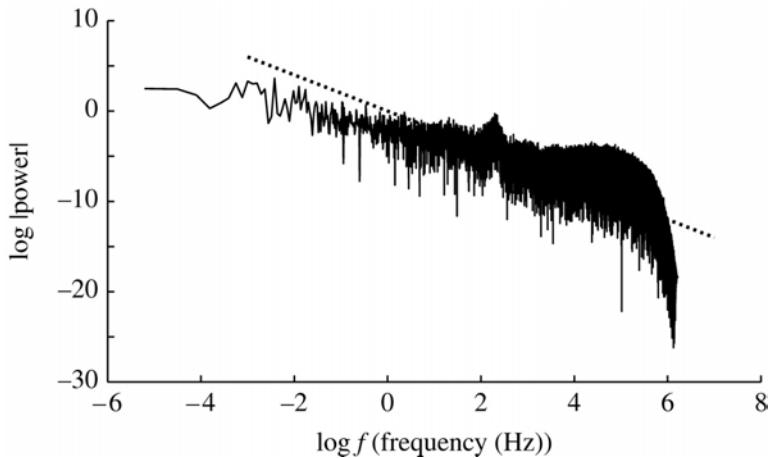
McCormick, 2000; Lee et al., 2005) and in vitro (Carracedo et al., 2013). Countless clinical and fundamental studies have examined oscillatory patterns under the most disparate brain states, ranging from conscious processing of complex stimuli (Henry et al., 2014) to states of deep coma (Schabus et al., 2011).

Initially the amplitude of rhythmic oscillations was found to be much larger during rest, sleep, unconscious states and under anesthesia. It is for this reason that the cognitive and functional role of these patterns have been overlooked for many years, and only recently researchers have turned their attention to the potential role of oscillations during awake brain states. As a matter of fact, even some of the patterns previously thought to be a signature of a lack of consciousness have now been related to brain activity in the vigilant state, such as, for example, the replay of spatial memory sequences in the hippocampus (Louie and Wilson, 2001). Furthermore, with the help of pharmacological tools, brain waves can now be produced in vitro and observed in vivo at the level of single cells, allowing for the investigation of causal relations between rhythms and behavior. Combined together, these factors prompted the comeback of interest towards studying oscillations and their potential functions during cortical and subcortical operations (Buzsáki and Draguhn, 2004). They seem to be likely candidates for the facilitation of information selection and transmission through input filtering (Hutcheon and Yarom, 2000; Akam and Kullmann, 2010), coordination of distant areas (Womelsdorf et al., 2007) and binding of perceptual information (Engel et al., 2001; Roopun et al., 2008a; Singer, 2013). Similarly, the execution of sensory-motor functions (Schoffelen et al., 2005), the formation and persistence of memories (Fell et al., 2001; Jensen and Lisman, 2005; Axmacher et al., 2006; Steriade, 2006), and synaptic plasticity dynamics (Huerta and Lisman, 1995; Bukalo et al., 2013) also involve oscillations of some sort. Nonetheless, the task of going beyond simple correlations between rhythms in the brain and behavior, and thus demonstrating causal relationships between oscillations and cognitive functions, proved to be impervious. This reflects in the fact that, until recently, many prominent neuroscientists were skeptical and considered them an epiphenomenon (Koch, 1993; Frégnac et al., 1994; Pareti and De Palma, 2004). It also reflects a lack of appropriate methodological tools to properly assess the causal role of oscillations: experimental techniques did not allow to selectively manipulate and perturb rhythms in vivo, until the very recent emergence of advanced optogenetics techniques. This fantastic tool relies on a set of genetically modifications that change the molecular composition of the membrane in specific neurons, so that, when targeted with light at a given wavelength, the firing activity of these neurons can be enhanced or suppressed (Boyden et al., 2005). Future optogenetics studies might have enormous potential implications in revealing the actual role of brain waves, if any. In addition, a number of methodological precautions must be considered, i.e. to avoid the effect of volume conduction from distant areas (Sirota et al., 2008) or to minimize false positives when filtering (Quiroga et al., 2001). It becomes crucial then to track down the neural mechanisms that generate brain waves, and to tie them to the function they perform in the brain, but before doing that let us shortly recapitulate how oscillations are measured in brain recordings.

## 2.2 Oscillations and brain functions

In order to be captured by the macroscopic electrodes of EEG, brain waves like the ones

observed by Berger must arise from time activity of somewhat synchronized big neuronal ensembles. Intuitively, in fact, the asynchronous activity of thousands of neurons would produce a bunch of waves with random phases, which, when summed together, would on average result in the cancellation of their amplitudes and thus one should not be able to measure an electrical signal. Indeed, it is almost impossible to avoid the presence of brain oscillations when recording from almost any region in the brain, at least on a large spatial scale, but, in addition to the average signal produced by the collective dynamics of large cell assemblies, oscillations can as well be detected in single neuron activity (Llinás, 1988). The neuronal membrane often displays coherent fluctuations in its voltage that can be measured intracellularly or extracellularly in the LFP, which contains contributions from all ionic processes occurring in the cell, at all levels: soma, axon, dendrites and even synapses (Buzsáki et al., 2012). The number of neurons that can be recorded and the spatial resolution of the signal heavily depends on the size of the inserted electrode, ranging from the few tens of micrometers of microarrays used in animal research (about 1000 neurons with a radius of  $\sim$ 50-120  $\mu\text{m}$ ) to the few millimeters of stereotactic depth electrodes implanted in epileptic patients (about 10000 neurons with a  $\sim$ 0.25-0.5 mm radius). The latter technique, called stereotactic electroencephalography (SEEG) is used to record local fields in human subjects, as for example in epileptic patients to locate the seizure focus. Intracranial EEG (*iEEG*) methods (stereotactic EEG, SEEG, and electrocorticography, *ECoG*) have both a higher signal-to-noise ratio and a better spatial resolution than any non-invasive technique like EEG or MEG (Kahana et al., 2001). Thus, they are well placed in order to detect oscillations at all frequencies, even high frequency oscillations (40-120 Hz) that are very hard to see in scalp recordings (Lachaux et al., 2012).



**Figure 1. Power laws in brain recordings.**

Power spectrum of a single electrode in human EEG recordings, exhibiting the characteristic power-law decay. The log-log scale plot transforms the  $1/f^\alpha$  curve into a straight line, whose slope is the exponent  $\alpha$  (in this figure  $\alpha \sim -2$ ). Subjects were recorded with their eyes closed during resting-state behavior. Adapted from (Jirsa, 2009).

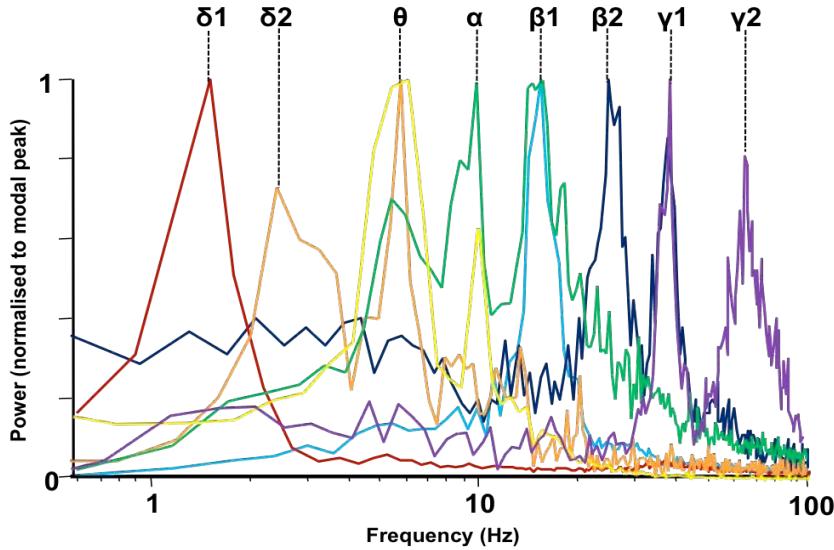
### 2.2.1 1/f background and the meaning of spectral peaks

When recording from multiple units, regardless of the technique used, the so-called 1/f background noise is an omnipresent feature of the frequency spectrum. Figure 1 shows the average power spectrum  $P(f)$  of EEG recordings in humans as a function of frequency  $f$ , computed using complex Morlet wavelets (Gouillaud et al., 1984). By fitting the power spectrum to the function  $P(f) = f^{-\alpha}$  one typically obtains an exponent in the range  $\alpha = [0,3]$ , where the extreme values might be approached only asymptotically. A distribution on this kind is called a *power law* and it has been reported in studies regarding the most disparate physical (Bak and Paczuski, 1995), biological (Musha and Yamamoto, 1997) and social systems (Baillie, 1996). This class of distributions has several peculiar features that make it a rather special class. First of all, for large values of  $f$ , the probability  $P(f)$  remains finite and much higher than in the Gaussian distribution. Intuitively, this happens because large frequency values result in a bigger denominator in  $\frac{1}{f^\alpha}$ , while in power laws they cause a (much more dramatic) change in the exponential ( $P_{Gauss}(f) = e^{-f}$ ); thus, extreme events are much more likely to occur if the underlying process is explained by a power law than the usual Gaussian distribution. For this reason power laws are often called *fat-* or *heavy-tailed* distributions. The second key property is scale invariance: power laws described by the same exponent  $\alpha$  are scaled versions of each other, so that there is no typical scale or size for the variable  $f$ . This peculiarity suggests a connection with fractals and with critical phenomena in physics that could potentially revolutionize our understanding of the brain as a complex system poised in the vicinity of a critical point (Kello et al., 2010). Such systems are more efficient in adjusting their internal representations to be good proxies of reality (Hidalgo et al., 2013, 2014), a feature that would obviously be of utmost interest to neuroscientists. In fact, power spectra estimated from scalp or intracranial EEG data indeed show a  $\frac{1}{f^\alpha}$  dependency at least within about two orders of magnitude (Freeman et al., 2003) and could be generated thanks to activity-dependent time-varying synaptic strengths (Levina et al., 2007). However, the identification of underlying processes giving rise to power laws in the brain is still controversial: a line of work pursued for example by Destexhe and collaborators has lead to the hypothesis that  $\frac{1}{f^\alpha}$  distribution in brain signals is a consequence of spatial filtering due to the structure of extracellular media (Bédard et al., 2006, 2010; Bédard and Destexhe, 2009) and thus should be treated as noise instead of a proof for criticality. From a theoretical perspective there are, borrowing the words of Rudyard Kipling<sup>1</sup>, nine and sixty ways of constructing power laws, and every single one of them is right: i) from the spectrum of some classes of diffusion processes (Milotti, 2002), ii) from a central limit theorem for multiplicative growth processes (Ijiri and Simon, 1977; Reed and Hughes, 2002), iii) from a superposition of exponentially relaxing processes (Dutta and Horn, 1981), and iv) even from windowed sampling of white noise (Kasdin, 1995). At the moment no consensus has been reached regarding the explicit mechanisms that originate power laws in brain signals, although an interesting view links these distributions to the activity of coupled neural oscillators in the vicinity of one or more attractors of the dynamics (Reed and

---

<sup>1</sup> 'In the Neolithic Age', *The Seven Seas* (Nabu Press, 2010).

Hughes, 2002; Teramae and Tanaka, 2004; Deco and Jirsa, 2012). In fact, a system of brain oscillators, almost uniformly distributed on a logarithmic scale along the frequency spectrum as in Figure 2, could give rise to such a scenario (Buzsáki and Draguhn, 2004). Brain rhythms are usually visible in the frequency spectrum of short term EEG or MEG as ‘bumps’, standing out from the  $1/f$  shaped background activity. If different frequency bands can be present at rest, they can be modulated by cortical input or neuromodulators during the performance of an active cognitive or motor task. Also, distinct rhythms can coexist and sometimes interact, giving rise to very interesting effects that will be reviewed below.



**Figure 2. Brain rhythms in the frequency spectrum.**

Normalized amplitude of different brain rhythms generated in cortical slices of rodents. The spectral peaks, that identify the different rhythms (frequencies may vary across species), are well-separated in frequency and are generated in different layers of the cortical sheet. Adapted from (Roopun et al., 2008a).

### 2.2.2 Synchrony and cross-frequency coupling

Theoretical studies had already pointed out that grouping together neurons which are carrying similar (to achieve redundancy and therefore robustness) or complementary (so as to bind together the various pieces of information related to a common object) information must be a fundamental task of the brain (Tsukada et al., 1996). This claim has gain much more attention recently, as it has been actually shown that neural assemblies in sensory cortex do display a dynamical and rhythmic time course, particularly those who fire in the Gamma (30-100 Hz) frequency band (Shadlen and Movshon, 1999; Fries et al., 2002; Buzsáki and Wang, 2012), grouping together the activity of cells on a small spatial scale. The coordination of these neural populations in time and space would also be fundamental in order to enhance (or inhibit) communication between distant areas, a goal that can be achieved by tuning the two populations so that they spike in phase (or antiphase), according to the “communication through coherence” hypothesis proposed by Fries and collaborators (Fries, 2005).

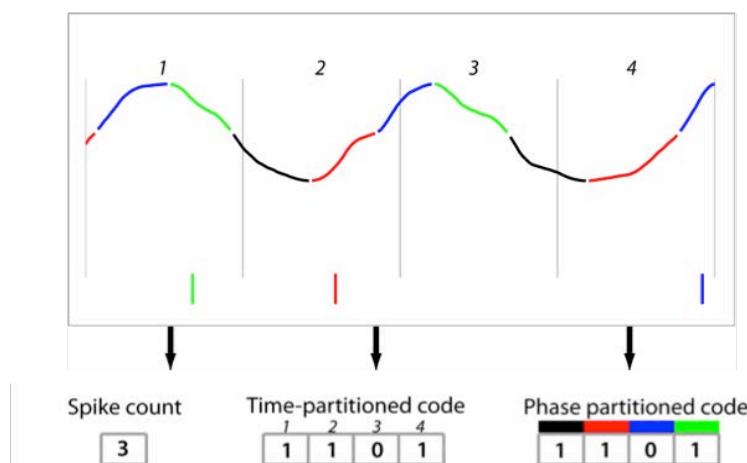
Further theoretical work by Akam and Kullman investigated the potential advantages of spatial and temporal oscillations in selectively transmitting population-coded information, even in the presence of asynchronous background noise from other neural sources (Akam and Kullmann, 2010). What is the mechanism that underlies the control of Gamma neural assemblies then? Several lines of evidence, both from computational and experimental studies, converge towards low frequency oscillations (1-20 Hz) as being well suited for the coordination of neurons on larger spatial scales (Buzsáki and Draguhn, 2004) using long-range connections, while fast Gamma waves act prevalently on short-range monosynaptic synapses (von Stein and Sarnthein, 2000). The interactions between low and high frequency rhythms, named cross-frequency coupling (CFC), have been extensively observed both with *in vitro* and *in vivo* experiments (Jensen and Colgin, 2007; Young and Eggermont, 2009), and can take complex forms. CFC will be the object of a review article we currently writing (Hyafil, Giraud, Fontolan, & Gutkin, *in preparation*). In fact, an oscillation is defined through three main features (frequency, amplitude and phase) that can interact and generate up to nine distinct coupling combinations. However, only four of all potential pairs are actually meaningful: *phase-frequency coupling*, i.e. when the phase of a slow oscillation (SO) interacts with the frequency of a fast rhythm (FO); *phase-phase coupling* (i.e. phases of SO and FO); *phase-amplitude coupling* (i.e. phase of SO and amplitude of FO); and, last, *amplitude-amplitude coupling* (i.e. amplitudes of SO and FO). The latter is the easiest to measure although its functional relevance has not been understood yet (Canolty and Knight, 2010): its presence, either measured between two areas or within a single region, has been linked to increased membrane excitability, possibly via a common input source (Young and Eggermont, 2009). Conversely, phase-phase couplings are hard to reveal experimentally due to the smallness of the effect (Hunter and Milton, 2003) and the fact that they are typically found in conjunction with phase-amplitude coupling. Recently, with the appropriate methods to measure phase of non-harmonic oscillators, phase-phase coupling were revealed in animal models (Akam et al., 2012; Belluscio et al., 2012). Phase-frequency couplings are extremely hard to disentangle from other forms of coupling. The most interesting and prominent of the cross-frequency combinations is indeed phase-amplitude coupling (PAC, not to be mistaken with primary auditory cortex). PAC has been reported between many distinct frequency bands, using a myriad of recording techniques on several different organisms and brain areas (Canolty et al., 2006; Schroeder and Lakatos, 2009; Belluscio et al., 2012; van der Meij et al., 2012). Perhaps the most famous example of PAC has been found in the hippocampus, involving Theta (3-8 Hz) and Gamma (30-100 Hz) frequencies, both in rodents (Bragin et al., 1995; Tort et al., 2007a; Wulff et al., 2009; Lisman and Jensen, 2013) and humans (Axmacher et al., 2006; Staudigl and Hanslmayr, 2013), the interpretation being that the phase of slow Theta cycles correlates with the power of fast Gamma rhythm to represent sequences of related objects (features). This idea, originally formulated by Lisman (Lisman and Idiart, 1995; Jensen and Lisman, 2005), was initially applied to the encoding of multiple items in working memory by the hippocampus. More recently, it has been extended to other neural systems, frequencies and computational functions, suggesting that the PAC scheme could be a fundamental configuration of mammalian brain (Canolty and Knight, 2010). For example, while Theta-Gamma PAC in the hippocampus

could underpin non-visual working memory, Alpha-Gamma PAC over parietal-occipital networks (where Alpha has a frequency range of 9-13 Hz) is thought to mediate the visual part of working memory (Roux and Uhlhaas, 2013). Similarly, Theta-Gamma coupling has been linked to spatial navigation in rodents, where, specifically, Theta sequences reflect the capacity of the hippocampus to bind together and store information about spatial locations at a rate of one location per Gamma cycle (Lisman, 2005; Lisman and Buzsáki, 2008). Alpha-Gamma coupling in humans and in monkeys was instead proposed to be involved in the spatial encoding of the visual field in primary visual cortex whereby, in this instance, the order of activation of neural representations within one Alpha cycle would depend on the saliency of the corresponding stimuli (VanRullen and Koch, 2003; Jensen et al., 2014). Even more importantly, a series of recent works has highlighted the presence of a whole hierarchy of *nested* oscillators, i.e. a battery of rhythms mutually linked through PACs, within visual (Lakatos et al., 2008; Schyns et al., 2011) and auditory (Lakatos et al., 2005) regions of primates. The authors measured the LFP/EEG signal across different layers and regions, and noticed that distinct rhythms are tied to each other through PACs. Such a system of interacting neural populations, firing at distinct frequencies, is a powerful tool to regulate neuronal excitability inside a particular region and within discrete time windows. Furthermore, this interplay brings a major contribution to the understanding of how neurons might encode stimulus features.

A number of influential articles have put forward the idea that neurons firing in the Gamma band are the readout units of the brain, not only in the hippocampus to help spatial navigation, but also in prefrontal and sensory cortices (Fries et al., 2007; Lisman and Jensen, 2013). Per contra, the actual neural code is unknown at the moment: despite the countless number of publications on this subject, for many years neuroscientists have limited their interpretation to the firing rate in order to explain their data, i.e. the average number of spike emitted by a neuron per unit time (often further averaged across trials). Alternative strategies, such as taking into account the exact temporal order of spikes, are appealing given that they are more informative (Diesmann et al., 1999; Salinas and Sejnowski, 2001), albeit much more fragile, since they are much more vulnerable to trial-by-trial variability of neural signal (Ferster and Spruston, 1995; Softky, 1995). In order to reduce the influence of noise, the time windows within which firing rates are computed can be discretized into smaller bins, and spikes can be labeled according to the particular bin inside which they occur (Optican and Richmond, 1987). Tagging action potentials enhances the information decoded from recorded neurons (Panzeri et al., 2010), demonstrating that spike trains are remarkably reliable across trials, at least within an optimal time bin. However, this binning process is artificial, meaning that the brain has no access to the binning scheme imposed by experimenters. So, does the brain have access to some kind of internal time discretization strategy? Low frequency oscillations, like those in the Theta range, can be the answer to this question, by providing the internal time reference necessary to enhance the encoding of information in the network. Being coupled to neurons participating in the Gamma rhythm, their time course relative to Gamma spikes is quite consistent and provides a powerful tool to define the decoding window for a much more accurate readout (Brasselet et al., 2012; Kayser et al., 2012; Panzeri et al., 2014). In the case of PAC, a

straightforward encoding scheme is constructed by combining the temporal sequence of Gamma neurons spikes with a *phase-code* (Hopfield, 1995; Varela et al., 2001; Mehta and Lee, 2002; Kayser et al., 2009) obtained by binning the phase of slow frequency bands into  $n$  intervals, and then assigning a tag to each spike according to the interval in which they fall (Figure 3). Hence, cross-frequency coupling, and especially phase-amplitude coupling, may be a rather efficient method to organize the collective behavior of neurons. With the help of computational models we show that this combinatorial code not only maximizes the encoded auditory stimulus information, but also adjusts the timing of sensory output for upstream decoding neurons (Hyafil et al., 2015).

Given the pervasiveness of CFC, the question is whether or not this scheme constitutes a fundamental blueprint for communication in the brain. It becomes therefore necessary to rule out the possibility that it might reflect an epiphenomenon, and to provide even more compelling evidence to bind together cross-frequency couplings and brain functions, with the help of methodological advancements. We are currently writing a review paper on cross-frequency coupling, with the aim of illustrating the link between the mechanistic functions and the underlying neural circuits generating the couplings (*Trends in Neurosciences*, review proposal accepted).



**Figure 3. Phase information code.**

Schematic illustration of different neuronal codes. In the *spike count* (or *firing rate*) code we simply sum the number of spikes occurring within a window of fixed length (1, 2, 3 or 4). In the *time-partitioned* code, the window is divided into four subintervals, and each spike is assigned to a subinterval depending on its timing. In the *phase-partitioned* code, we assign an additional label to each spike, depending on the phase of low frequency LFP at which a spike occurs. The latter code is much more informative (Hyafil et al., 2015) and rises from an endogenous reference mechanism that could be accessible to neurons, in contrast with the other codes that are somehow artificially determined. Adapted from (Kayser et al., 2009).

### 2.2.3 Phase-amplitude coupling and causality

Having established that hyperpolarization-depolarization cycles within a given layer or area give rise to oscillatory patterns in the LFP, and that multiple rhythms coexist and intertwine, it becomes then crucial to establish whether the coupling is just apparent, as it would result if the two rhythms were driven by a common input, or is caused by

directed influence from one rhythm to the other. Currently, the most popular family of measures used to assess the causal interactions between time series is that of Granger causality (GC). First introduced by Clive Granger in 1969 (Granger, 1969) to better understand economic data, this measure lies at the core of many causal inference methods such as information theory based estimators (e.g. transfer entropy) and classical linear estimators based on correlation or coherence (partial directed coherence, directed transfer function). In its simplest implementation, based on multivariate autoregressive models (MVARs), has already been successfully applied to large-scale brain data (Bressler and Seth, 2010; Brovelli, 2012; Saalmann et al., 2012). This approach relies on a few simple steps and works for as many variables as desired. Here, we illustrate the computations for the two variables case. First, the current value of a time series is modeled as the linear sum of its past values weighted by some vector, plus a term incorporating the unpredicted stochastic variations:

$$X_t = \sum_{i=1}^n a_i^x X_{t-i} + \varepsilon_t^x, \quad (2.1)$$

where  $n$  is the order of the autoregressive model. Then we construct a second model, where the past of a second variable  $Y$  is added to Equation 2.1:

$$X_t = \sum_{i=1}^n a_i^{xx} X_{t-i} + \sum_{i=1}^n a_i^{xy} Y_{t-i} + \eta_t^x, \quad (2.2)$$

and a similar equation for  $X_t$ . Now, if the presence of the second term in Equation 2.2 helps to reduce the error term  $\eta_t$  with respect to  $\varepsilon_t$ , then  $Y$  is said to *cause*  $X$ , i.e. its past reduces the uncertainty over the present of  $X$ . The Granger causality coefficient is easily defined as:

$$GC_{Y \rightarrow X} = \log \left( \frac{\text{var}(\varepsilon_t^x)}{\text{var}(\eta_t^x)} \right). \quad (2.3)$$

An equivalent measure in the frequency domain has been introduced by Geweke (Geweke, 1982), based essentially on applying a Fourier transform on the MVAR so that:

$$\begin{pmatrix} A^{xx}(\omega) & A^{xy}(\omega) \\ A^{yx}(\omega) & A^{yy}(\omega) \end{pmatrix} \begin{pmatrix} X(\omega) \\ Y(\omega) \end{pmatrix} = \begin{pmatrix} \eta^x(\omega) \\ \eta^y(\omega) \end{pmatrix}, \quad (2.4)$$

where  $A^{xx}(\omega) = 1 - \sum_k^\infty a_k^{xx} e^{-i\omega k}$ ,  $A^{xy}(\omega) = 1 - \sum_k^\infty a_k^{xy} e^{-i\omega k}$ , etc.

If matrix  $\mathbf{A}(\omega)$  is invertible, by applying  $\mathbf{A}^{-1}(\omega)$  to both sides of Equation 2.4 we can write

$$\begin{pmatrix} X(\omega) \\ Y(\omega) \end{pmatrix} = \begin{pmatrix} H^{xx}(\omega) & H^{xy}(\omega) \\ H^{yx}(\omega) & H^{yy}(\omega) \end{pmatrix} \begin{pmatrix} \eta^x(\omega) \\ \eta^y(\omega) \end{pmatrix}, \quad (2.5)$$

where  $\mathbf{H}(\omega)$  is called *transfer matrix*. The *spectral matrix*  $\mathbf{S}(\omega)$  of the process can be factored in terms of the transfer matrix  $\mathbf{H}(\omega)$ , its complex conjugate  $\mathbf{H}^*(\omega)$  and the covariance matrix  $\mathbf{\Sigma} = \begin{pmatrix} \text{var}(\eta_t^x) & \text{cov}(\eta_t^x, \eta_t^y) \\ \text{cov}(\eta_t^y, \eta_t^x) & \text{var}(\eta_t^y) \end{pmatrix}$ , according to the spectral factorization theorem (see Gevers and Anderson, 1981, for further details):

$$\mathbf{S}(\omega) = \mathbf{H}(\omega) \mathbf{\Sigma} \mathbf{H}^*(\omega). \quad (2.6)$$

The two diagonal elements of the spectral matrix,  $S_{xx}(\omega)$  and  $S_{yy}(\omega)$  simply represent the spectra of variables  $X$  and  $Y$ , while off-diagonal terms,  $S_{xy}(\omega) = S_{yx}^*(\omega)$ , contain the cross-spectrum of the two variables. After proper normalization ((Chen et al., 2006), the causal interdependence in the frequency domain can be expressed as the ratio:

$$GC(\omega)_{Y \rightarrow X} = \log \left( \frac{S_{xx}(\omega)}{\Sigma_{xx}|H^{xx}(\omega)|^2} \right), \quad (2.7)$$

where  $S_{xx}(\omega) = \Sigma_{xx}|H^{xx}(\omega)|^2 + \Sigma_{yy}|H^{xy}(\omega)|^2$ . The first term, equal to the denominator, represents the portion of the spectrum due to  $X_t$ , whereas the second term reproduces the power of  $X_t$  contributed by  $Y_t$ . Thus, if  $Y_t$  has a direct impact on the spectrum of  $X_t$  at a given frequency  $\omega$ , then  $GC(\omega)_{Y \rightarrow X} \neq 0$  and  $Y_t$  is said to *cause*

$X_t$  at frequency  $\omega$ . Although the MVAR approach is simple and powerful, it requires a few conditions to be met in order to work correctly: i) the integral of  $GC(\omega)_{Y \rightarrow X}$  over all frequencies must be equal to the time-domain GC; ii) the time series  $X_t$  and  $Y_t$  must have zero mean and their covariance must be stationary. Unfortunately, time series recorded from the brain often do not meet these criteria, plus it has been shown that the use of bandpass filters may destroy the correct causal relations (Seth, 2010), making the MVAR formulation useless.

An alternative, more powerful approach to compute GC on brain data has been proposed more recently by Dhamala and collaborators (Dhamala et al., 2008). Their formulation is *nonparametric*, meaning that matrices  $\mathbf{S}$ ,  $\mathbf{H}$  and  $\boldsymbol{\Sigma}$  are not derived from the MVAR parameters but are inferred directly from a time-frequency decomposition of the time series. For brain recordings, this means that the spectral matrix  $\mathbf{S}$  can be easily obtained from any spectral representation of the data (Fourier decomposition, wavelet transform, etc.), avoiding the flaws of MVARs, such as having to use very high order models to account for nonstationarity, thereby increasing the estimation error and potentially introducing spurious causal effects (Dettori et al., 2012). Transfer matrix  $\mathbf{H}$  and covariance matrix  $\boldsymbol{\Sigma}$  are computed using Wilson's spectral factorization algorithm (essentially a series expansion in nonnegative powers of the exponential function  $f(\omega) = e^{i2\pi\omega}$ ). In the case of Morlet wavelet decomposition (Gouplaud et al., 1984), i.e. the approach we used in (Fontolan et al., 2014), this allows to restrict the validity of the aforementioned criteria to the length of the wavelet, a condition much easier to fulfill. In this form, Granger causality has been successfully employed to test directional influences between two or more brain areas at a given frequency (Bosman et al., 2012; Bastos et al., 2014; Fontolan et al., 2014; van Kerkoerle et al., 2014), but it cannot be used to assess causal interactions across frequencies as for example in a PAC scheme. It does not exist, at the moment, any linear measure to assess causal interactions between two separate frequencies. The only available measure is dynamical causal modeling (DCM, see C. C. Chen, Henson, Stephan, Kilner, & Friston, 2009, and Friston, Moran, & Seth, 2013), which is nonlinear and requires a set of somewhat precise hypotheses, being model-based. Indirect causal interactions can be revealed to a substantial extent using a combination of GC methods and standard PAC measures such as the modulation index or circular-to-linear correlation, as it has been envisaged for the first time in our article (Fontolan et al., 2014, see Appendix A and Penny, Duzel, Miller, & Ojemann, 2008, for an exhaustive review of PAC methods).

Throughout this Section we have seen how synchronized assemblies of neurons can give rise to nontrivial patterns of alternating excitations and inhibition that can be revealed with almost all brain recording techniques. In the next Section we will turn our attention to a brief but complete review of recent literature on computational models of neuronal oscillations, and examine the possible different network architectures that

produce such rhythmic patterns in cortical circuits.

### 3 Computational models of brain oscillations

*The best material model of a cat is another, or preferably the same cat.*

Norbert Wiener, Philosophy of Science

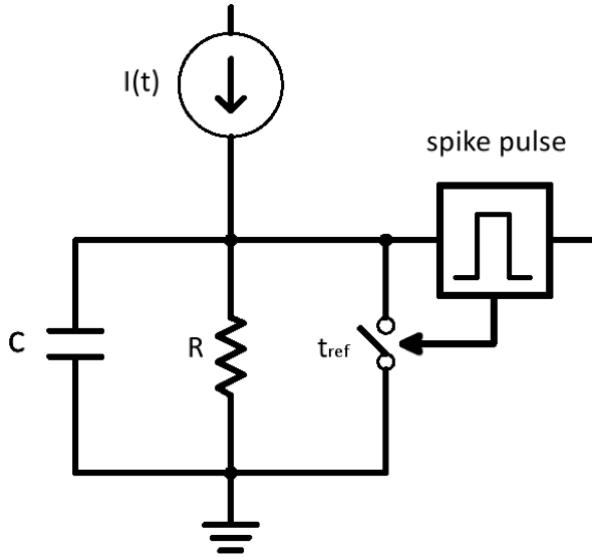
Since the very first recordings of neural activity, scientists have sought to formalize their observations into a mathematical framework, for a better understanding of the implications and potential consequences of the experimental results. In the context of brain rhythms, theoretical analysis and computational modeling are key tools for understanding how oscillations are generated, both at the cellular and at the circuit level, to establish why they work in that particular way and find out which function (if any) they subserve. A theoretical framework allows to tie together these aspects into a compact, efficient formal representation, providing testable predictions and, ideally, a set of unifying principles that underlie cortical rhythms. We will now review the biophysical mechanisms giving rise to these rhythms in individual neurons and in neuronal populations.

#### 3.1 Single neuron models

It is well known that the membrane potential of individual neurons displays oscillatory patterns at different frequencies (Llinás and Yarom, 1986; Llinás, 1988, 2014; Wang, 2010), suggesting that the simplest source of brain oscillations is the neuron itself. We may, in fact, consider all neurons as oscillators, as long as they exhibit periodic or quasi-periodic firing in response to an external constant current drive  $I_e$ . Generating a spike is, as a matter of fact, a dynamical process involving several internal (i.e. ion channels, membrane properties, neuronal topology) and external (i.e. chemical and electrical synaptic input from coupled surrounding neurons, LFP oscillations) variables across a myriad time scales. The most basic essential ingredients to obtain the nonsinusoidal wave patterns that are typical of biological oscillators (Kruse and Jülicher, 2005), irrespective of the particular system under consideration, are a rapid excitation followed by a relatively slow inhibition. This type of nonlinear dynamical system is an example of *relaxation oscillator*, commonly found in biological processes (Hill, 1933; Miroollo and Strogatz, 1990; Wang, 2001). During the generation of a spike, neurons behave as relaxation oscillators whose two crucial components are, in this particular case, voltage-gated activation of sodium current and subsequent activation of potassium channels. In the phase space, such system normally lies in a global attractor state when at rest or when in presence of weak (subthreshold) inputs, but it gets rapidly activated once a strong enough stimulus pushes the dynamics to make a wide excursion above threshold. Finally, the inactivation of sodium channels and the opening of potassium pumps bring the system back to the stable resting state. One might ask whether the fundamental dynamical patterns of a topologically complex real neuron, potentially implicating more than a hundred coupled variables (Yamada et al., 1989), might actually

be grasped by using only an exiguous number of components. The answer is, as we will see, yes in most cases, at least as far as synchronization properties are concerned. This mathematical reduction is extremely useful since low dimensional systems are certainly more tractable and sometimes even analytically solvable. The famous Hodgkin-Huxley (H-H) model (Hodgkin and Huxley, 1990) does a pretty good job mimicking a pyramidal neuron by modeling its dynamics with a set of four nonlinear differential equations. The most important of these equations defines the current-voltage relation, while the remaining three equations describe the activation of ion channels during spike generation. The H-H equation set accounts for many dynamical patterns observed in real data, like sub and suprathreshold behaviors, adaptation, bursting. At the price of being able to account for only one of these behaviors at a time, the H-H model can be further simplified by reducing the number of variables. Although the details of ionic dynamics are important for spike generation mechanisms and to compute the time it takes to repolarize the membrane potential, reduced models are useful to simulating large networks, due to their lower computational weight. Examples of simplified model include, in descending order of complexity, the Fitzhugh-Nagumo model (Fitzhugh, 1961; Nagumo et al., 1962), Izhikevich's neuron model (Izhikevich, 2004), and the integrate-and-fire (IF) neuron (Gerstein and Mandelbrot, 1964; Abbott and van Vreeswijk, 1993) in all its forms (linear, quadratic or exponential). The latter is particularly interesting in order to simulate large-scale neuronal oscillations (Burkitt, 2006), especially in its quadratic configuration (see Appendix B). For the moment let us introduce the *leaky integrate-and-fire* neuron with the help of Figure 4: it consists of a point-like physical entity whose voltage is described by a circuit containing one capacitance  $C$ , a single resistor  $R$  to account for membrane current *leakage*, both driven by an external current  $I(t)$  symbolizing synaptic input from other neurons. Whenever the voltage  $V$  reaches a threshold value  $V_{th}$  the neuron emits an instantaneous spike and the switch closes the circuit. When the system is short-circuited it means that the integrate-and-fire neuron is insensitive to further inputs for a fixed amount of time  $t_{ref}$ . Such simple circuit only captures the very basic features of a biological neuron, although adding a voltage-dependent capacitance that allows to model the rapid opening of sodium channels during spike generation can of course enrich it. When the capacitance dependency is quadratic the model is called *quadratic integrate-and-fire* (QIF, Ermentrout and Kopell, 1986). The QIF is an example of *type I* neuron, whose canonical form is called the *theta model*: for their importance in computational neuroscience, the two models are reviewed in Appendix B. Neuron models are classified into type I or II depending on the particular class of *bifurcations* they display: type I neurons are associated with *saddle-node* (SN) bifurcations, and their oscillation frequency approaches zero when reducing the input drive; instead, the onset of oscillations in type II cells (e.g. the H-H model) occurs at a finite frequency. Pyramidal excitatory neurons seem to behave like type I or II models depending on the area and the cortical layer considered (Tateno and Robinson, 2007; Tsubo et al., 2007), while inhibitory interneurons' behavior appears to be closer to type II (Tateno et al., 2004; Mancilla et al., 2007). This hence means that the QIF model is better suited to simulate a pyramidal cell in granular/infragranular layers (Tsubo et al., 2007), and that its spiking frequency can be made as low as desired, by varying the bifurcation parameter, i.e. the ratio of excitation

and inhibition (see Appendix B). In the next paragraph we will see what happens if we start connecting neurons to each other, using the computational models we have just introduced here.



**Figure 4. Electrical circuit of the neuronal integrator.**

Schematic diagram of electrical circuit describing the integrate-and-fire model. The external current  $I(t)$  charges the  $RC$  circuit. The voltage  $V(t)$  across the capacitance is compared to a threshold  $\theta$ . If  $V(t) > \theta$  at time  $t = t^*$  a spike pulse  $\delta(t - t^*)$  is emitted. The system is then short-circuited for a time  $t_{ref}$  by closing the switch on the right branch.

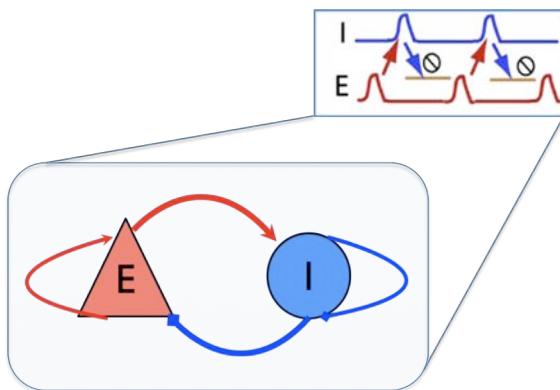
### 3.2 Neuronal networks: circuitry for oscillations

Even if we have just seen that a single neuron can produce a well-defined oscillation by spiking regularly in time, we know that cortical neurons are all highly interconnected through many excitatory and inhibitory synapses. Many physical and biological systems can be viewed as an ensemble of local subsystems, coupled to each other to produce a collective behavior. In others words, borrowed from physics, a system made of several, often countless, degrees of freedom can be quickly brought down to a system of just a few degrees of freedom, thanks to the mutual interactions between its oscillating elements. In physics this phenomenon is called *slaving principle* (Haken and Sauermann, 1963). In the case of cortical circuits, neural cells might fire at very different frequencies when isolated, but, when connected, they are attracted toward a single common spiking frequency. This scenario gives rise to a class of periodic or quasi-periodic phenomena named *strong rhythms*. When, instead, the system exhibits a single *average* collective mode of oscillation but its elements are still oscillating at distinct frequency, the rhythm is said to be *weak* or *sparse*. The two main parameters that determine the strength of the considered rhythm are obviously the potency of the coupling and, when present, the intensity of any external driving input. Coupling intensities (i.e. synaptic weights) can be quite heterogeneous in the brain, giving rise to both types of rhythms serving different purposes (Brunel, 2000; Brunel and Wang, 2003; Kopell et al., 2010; Buzsáki

and Wang, 2012). Also, computational studies have shown that neuronal oscillations can rise in presence of both irregular (Brunel and Hakim, 1999; Mongillo and Amit, 2001) and synchronized input (Börgers and Kopell, 2005; Kopell et al., 2010), depending on the considered parameter set. An important tool to study synchronization in presence of a small rapid inputs is the *Phase Response Curve* (PRC) defined as the time (or phase) shift in the oscillation cycle induced by an infinitesimal perturbation (Kuramoto, 1984; Hansel et al., 1995; Ermentrout, 1996). Let us consider a neuron which exhibits intrinsic firing with constant period  $T$ , a PRC measures the advancement (or delay) provoked by a small perturbation occurring at time  $t$ , or, equivalently, at phase  $\varphi$  (see Appendix B). PRCs have been measured experimentally in single neurons by stimulating the soma and measuring the changes time to successive spikes, and computed theoretically for the different classes of neural models. Previous works (Ermentrout, 1995; Brown et al., 2004) have thus pointed at the important qualitative difference between type I and type II models: the PRCs of type I models are always positive when stimulated with an excitatory input, while those of type II neurons, rather counterintuitively, present a negative portion, i.e. a region in the phase space where an excitatory stimulus will induce a delay in the generation of the next spike. Intuitively, the shape of the PRC is particularly relevant to understand the synchronization properties of interconnected cells (Smeal et al., 2010), especially since it can be extended to stronger and longer interactions (Gutkin et al., 2005). The synchronization properties of brain networks, in fact, can be predicted by computing the long term phase relationship of two (or more) mutually interacting neurons in the PRC formalism, provided that the effect of the perturbation is lost after maximum two cycles (Oprisan and Canavier, 2006). As a result, previous studies have found that both excitation and inhibition can lead to synchronize patterns of activity (Van Vreeswijk et al., 1994; Terman et al., 1998), although inhibition is more suitable than excitation (Hansel et al., 1995) in order to create stable synchronization. It is still debated whether reduced models combined with the PRC formalism is actually a useful method to simulate and study biological neural networks, given the degree of complexity of real brain circuits. For the purpose of this thesis we adopt the opinion expressed by the authors of (Smeal et al., 2010), i.e. that even if several quantitative and a few qualitative details are lost, the most important features of those circuits are still represented in the models mentioned above. The interested reader will find an exhaustive review on this topic in the article of Smeal and collaborators (Smeal et al., 2010).

Nevertheless, even if PRCs are useful in many situations and allow for a qualitative understanding of even rather strongly coupled systems, it becomes sometimes necessary to use numerical simulations in order to grasp the quantitative details and make the correct predictions for future experimental manipulations. This is the case for Gamma rhythms in hippocampus and neocortex. Indeed, we know from *in vitro* experimental works that there exist two main mechanisms of inhibition-based strong rhythms (Whittington et al., 2000; Tiesinga and Sejnowski, 2009; Pietersen et al., 2014): the *interneuron network Gamma* (ING) where a population of inhibitory interneurons are firing together thanks to self-inhibition, and the *pyramidal interneuron network Gamma* (PING) where synchrony is achieved by means of coupled excitation and inhibition. Whether it will be possible or not to distinguish between these two

schemes with specific experiments is a still debated issue, nonetheless they are thought to be behind the genesis of brain oscillations in neocortex and in the hippocampus, at least in the 20-80 Hz range (Isaacson and Scanziani, 2011), and they can both be modeled using networks of integrate-and-fire neurons and exponential synapses (Börgers and Kopell, 2005) (see Appendix B). Strong, slowly decaying inhibition plays a central role in both models, making them difficult to study with the sole PRC method. Indeed, during the last ten years, the PING framework has been increasingly employed to model Gamma oscillations in sensory areas (Börgers and Kopell, 2005, 2008; Kilpatrick and Ermentrout, 2011) and the hippocampus (Kopell et al., 2010; Börgers and Walker, 2013), due to its computational simplicity and the biological plausibility of its parameters. A population of pyramidal excitatory cells (PE), mainly found in middle layers (IV, V) of sensory cortex and in the hippocampus, is mutually coupled to a smaller population of fast spiking inhibitory interneurons (FS). An external driving current, that may represent input from a distant brain area (or an adjacent column) to layer IV (Douglas and Martin, 2004; Binzegger et al., 2009; Atencio and Schreiner, 2010), causes the PE to start firing, sparsely since their are not synchronized *a priori*. When the proper set of parameters is chosen, PE spikes reach FS cells and the latters also start firing, providing a window of rather strong inhibition that prevents further PE spikes, at least until inhibitory synaptic input has decayed. At that point almost all PE neurons are ready to fire, so that the subsequent volley is quite synchronized, quickly prompting one more inhibitory volley, and so on (Figure 5).



**Figure 5. PING network.**

Schematics of the pyramidal interneuron network Gamma model (PING). The excitatory network of pyramidal neurons (red circle) starts firing, eliciting a volley of spikes from the interneuron population (blue circle), which silences the network until inhibition fades and the excitatory population is ready to fire again.

Kopell and collaborators have extensively studied the different behaviors of PING networks across a wide range of parameter values, with both numerical and analytical tools. Gamma rhythm has been observed extensively in granular and superficial layers in auditory cortex of macaque monkeys (Lakatos et al., 2005), and the interplay between PE and FS giving rise to Gamma bursts has been observed *in vivo* and *in vitro* (Bragin et al., 1995; Tiesinga and Sejnowski, 2009). We will therefore use the PING model as our reference model for generating Gamma band oscillations in

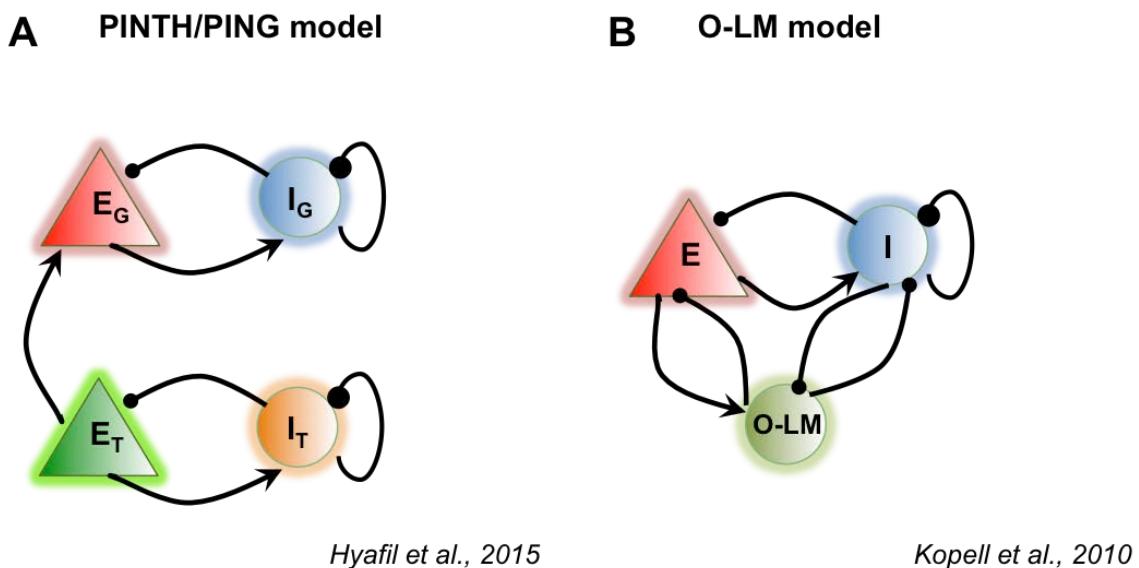
neocortex.

After exploring in this Section the mechanisms that are responsible for the generation of cortical rhythms within a single frequency band, let us now turn our attention to models dealing with cross-frequency coupling (as described in Section 2.2.2), as they will turn out to be useful to understand the work of this thesis.

### 3.3 Nested brain rhythms

Starting from the results of previous Sections, we briefly review the most interesting mechanistic models that produce CFC between two brain rhythms, phase-amplitude coupling models in particular. From a computational standpoint, frequency couplings can be classified according to: i) whether the cross-frequency perturbation is weak or strong, relatively to the intrinsic coupling strength; ii) whether the influence is temporally continuous or discrete and what is the size of the difference in time-scale between the two rhythms; iii) whether the interaction is one-sided or mutual. As for point i), the already mentioned plethora of synaptic connections provides room for both weak and strong interactions. The works proposed in (Fontolan et al., 2013; Hyafil et al., 2015) aim to build an analytical and computational model of Theta-Gamma interplay in auditory cortex using neural networks models. During speech recognition, Theta oscillations track the amplitude fluctuations in upcoming stimuli and reset their phase in correspondence of syllabic boundaries. Theta fluctuations provide a strong and continuous modulation to neurons firing in the Gamma range, whose frequency matches the phonemic time-scale (see Section 0). A weak Theta input would not be sufficient to reset Gamma, whereas it is crucial to be able to realign the cycle of Gamma neurons at the onset of a linguistically pertinent constituent, in order to maximize the information encoded in Gamma spike patterns. However, the great majority of theoretical studies focus on weak coupling, which is much easier to model using the tools of perturbation theory (Ermentrout and Kopell, 1986; Mancilla et al., 2007; Ledoux and Brunel, 2011). But what happens when the coupling is strong, i.e. when the amplitude of the synaptic input coming from the slowest rhythm is comparable with the membrane threshold value? Unfortunately, strong PAC has been analytically studied only in the (quite limited) case of discrete pulsatile coupling (Bressloff and Coombes, 2000; Tort et al., 2007b; Vierling-Claassen and Kopell, 2009). A very interesting article by Kopell and colleagues (Kopell et al., 2010) explores a Theta-Gamma PAC by modeling a PING coupled with stellate cells in entorhinal cortex (which they identify as equivalent to oriens lacunosum-molecular cells in the hippocampus — we will thus use the abbreviation O-LM from now on). In fact, these cells might be involved in producing Theta resonances due to the characteristic intrinsic currents they possess (White et al., 2000; Rotstein et al., 2005). Using the PRC formalism extended to three cells networks, Kopell and collaborators were able to draw the synchronization properties of two O-LM cells mutually coupled to a fast spiking inhibitory basket cell, and showed that the network fires in the Theta frequency range when loaded with biophysically compatible parameters (see Figure 6). For this scenario to occur, the two O-LM cells must fire first and provide long (around 20 ms) lasting inhibition to the FS basket cell, which is the last neuron to emit a spike within a network cycle and progressively synchronizes the two O-LM neurons (for the full mathematical derivation see Kopell et al., 2010). Note that a single O-LM neuron can actually generate a Theta rhythm, but two O-LM neurons

would not synchronize unless they receive inhibition from the FS cell. Although their model is indeed appealing and their analytical understanding both inspiring and intriguing, the existence of connections between O-LM and FS cells has still not been reported anatomically or physiologically, as stated by the authors themselves. Instead, to generate Theta oscillations in auditory cortex, as described in (Hyafil et al., 2015), we opted for a modified version of the PING (Figure 6), where PE neurons are coupled to slowly decaying inhibitory neurons like Martinotti cells, which are known to be connected to the distal dendrites of pyramidal neurons throughout neocortex (Silberberg and Markram, 2007a). We have also tested both models by comparing their ability to reset their phase in correspondence of a syllabic boundary, i.e. a steep rise in the amplitude of an external input, and we found that the Kopell model was much less suited for this purpose than our proposed model. For what concerns the time-scales at which the two oscillations operate, the PACs measured in auditory cortex involves a slow and a high frequency rhythm whose periods are separated by roughly one order of magnitude (e.g. for Theta-Gamma coupling in auditory cortex:  $T_\theta \sim 300 - 100$  ms,  $T_\gamma \sim 30 - 10$  ms, see Giraud and Poeppel, 2012 and Poeppel, 2014).



**Figure 6. Theta oscillations networks.**

Computational models used to simulate Theta oscillations in the brain. **A)** Modified Pyramidal Interneuron Gamma model used in (Hyafil et al., 2015), showing two separate modules each for generating Gamma (top, red and blue) and Theta (bottom, green and orange) oscillations. Within each of the two modules, a population of pyramidal cells ( $E$ ) is connected to a population of fast spiking inhibitory ( $I$ ). When  $E$  cells start firing, they trigger inhibitory spikes in the  $I$  population that prevents further firing in the  $E$  population until inhibition has faded away. Excitatory connections from Theta module excitatory neurons to Gamma module pyramidal cells link the two networks producing nesting. **B)** Model of hippocampal Theta oscillations introduced in (Kopell et al., 2010). Theta oscillations are generated in the O-LM/ $I$  inhibitory loop, where  $I$  cells are crucial for the synchronization of O-LM cells. The interaction with excitatory  $E$  cells further lead to the origination of the Gamma rhythm.

In addition to the computational results showed in the aforementioned article,

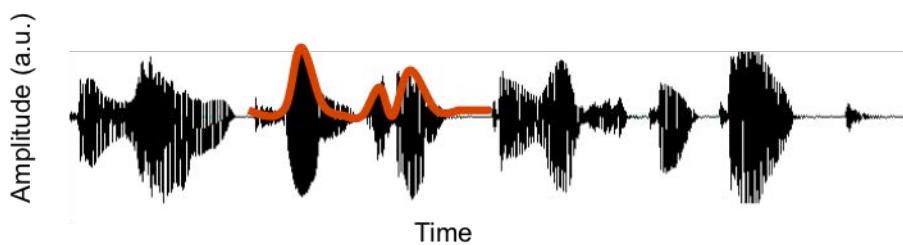
we explored the analytical properties of the network by considering a simplest version of the model: a QIF excitatory neuron coupled to an instantaneous inhibitory synapse (representing the FS inhibitory interneuron) and receiving slow sinusoidal input in the Theta frequency range (Fontolan et al., 2013). The intriguing results of both models will be discussed thoroughly later in this thesis.

## 4 Neural substrates of speech processing

*They were, perhaps, the first people to understand that the Tower was chaos, that order was chaos, and that language – the gift of tongues which Jahweh breathed into the mouth of Adam – has a rebellious and wayward vitality compared to which the foundations of the Pyramids are as dust.*

Bruce Chatwin, *The Songlines*

Language is perhaps the most unique of all the distinctive features of human beings. It still represents a puzzle of astonishing complexity both for neuroscientists and linguists, at both a high (e.g. grammar, syntax) and low level of description (e.g. transforming incoming sounds into neuronal spikes for subsequent auditory perception). The work of this thesis concerns the early, yet critical, low-level steps in the understanding of how language is processed and represented in the human brain: the chunking of speech into its primary constituents, i.e. syllables and phonemes. In physical terms, an uttered sentence reaching the outer ear is nothing but a mechanical pressure wave, a flux of energy traveling through the air medium. If we plot the amplitude of a speech waveform we get a representation like that of Figure 7, where we can immediately spot two concurring features: exceedingly fast fluctuations superimposed with slow global modulations. How can the neural circuits extract useful information from such an apparently simple source? To answer this question, in the next Section we will present a brief overview of the auditory system and the computational models that can help us in the understanding of its functioning.

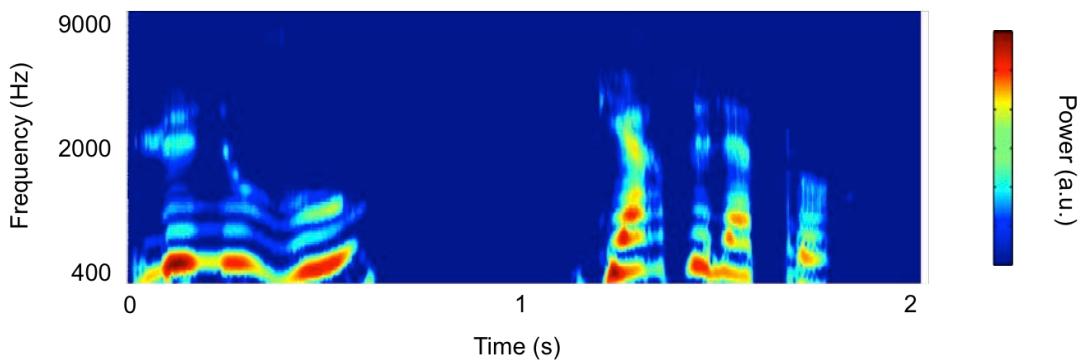


**Figure 7. Speech waveform.**

The amplitude of a speech waveform is plotted as a function of time. The envelope (i.e. low frequency power fluctuations) is shown in red. Syllables are visible as strong bursts separated by silence periods. Smaller scale constituents, such as phonemes, are almost indistinguishable.

## 4.1 The auditory pathway in a nutshell

When a sound wave reaches the outer ear, it first passes through the ear canal and then reaches the tympanic membrane. Once there, it gets amplified and subsequently transmitted to the cochlea by three tiny ossicles, the malleus, the incus and the stapes, which globally act both as a bandpass filter and an amplifier. As a result, humans can hear sounds in the 20 Hz to 20 kHz frequency range but are most sensitive in the 100-10,000 Hz range with a peak around 3 kHz (Rosen and Howell, 2011). Unsurprisingly, the frequency spectrum of human speech is comprised between 100 and 3000 Hz, i.e. near the highest sensitivity peak of the human ear. Once in the cochlea, the speech pressure wave travels in a biological fluid and provokes the motion of the basilar membrane, whose changes are promptly sensed by a few thousands hair cells. The basilar membrane does not move uniformly each time a sound wave is received; instead, distinct wavelengths elicit perturbations at different locations of the membrane. This anisotropy critically causes short wavelengths (i.e. high frequencies) to be more effectual in moving the membrane at its apical end, while long wavelengths (short frequencies) are progressively more effectual towards the basal end. This simple topographical organization of frequency representation in the cochlea is called *tonotopy*, and is preserved, at least partially, across the entire auditory pathway from thalamus to primary auditory cortex (Kanold et al., 2014).



**Figure 8. Auditory spectrogram of speech.**

Morlet wavelet time-frequency decomposition (spectrogram) of a spoken sentence in French. The obtained two-dimensional representation mimics the filtering performed by the human ear. The color code represents power at a given time-frequency points.

Essentially, the cochlea decomposes an incoming sound wave in frequency bands, and, by doing so, it expands the one-dimensional representation of Figure 7 into a two-dimensional one, where the instantaneous amplitude is now computed at a given frequency band (Figure 8). The time-frequency decomposed speech waveform, named *spectrogram*, is further transmitted from hair cells to the auditory nerve in form of a series of electric potentials. These voltage fluctuations activate the release of neurotransmitters in the synaptic clefts, which separate the hair cells and the auditory nerve fibers. These long fibers finally pass the auditory information, now transduced

into electrical impulses, to the cochlear nucleus, inside the brain stem. Cells in the cochlear nucleus are connected to multiple nuclei belonging to the pons, which, in turn, project to the inferior colliculus. The ascending auditory pathway culminates in the medial geniculate body (MGB) of the thalamus, from which the information is finally relayed to primary auditory cortex (A1, Broadmann area 41).

## 4.2 Physiology of the auditory system

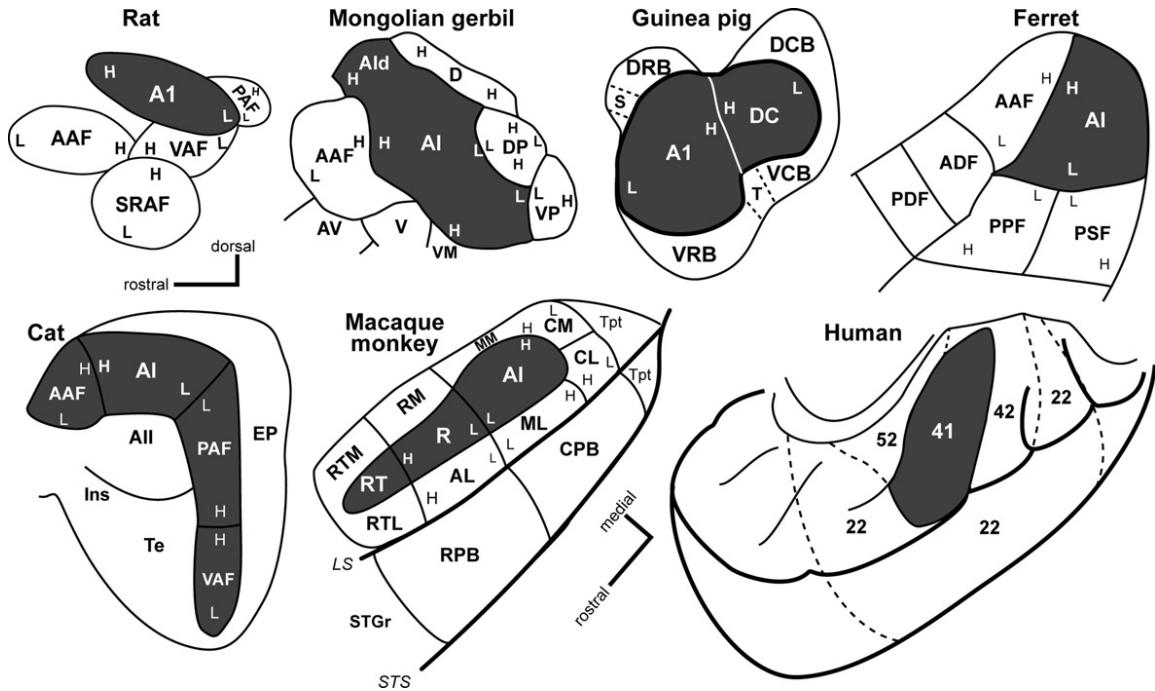
Neocortex is the most recently evolved portion of cerebral cortex in mammals, encompassing the areas assigned to the highest cognitive functions such as sensory perception, consciousness, language production and comprehension, planning, preparation of motor commands. Neocortex looks like a folded sheet made of six layers of various thickness, density and cellular conformation. The relatively constant number of neurons per millimeter found in neocortex (about 147000 in 1 mm<sup>2</sup>) lead many researchers to the conviction that the distinct sensory systems could have evolved to having similar organizational structures (Rockel et al., 1980; Carlo and Stevens, 2013). However, although neocortical regions do appear to share a number of basic principles in the spatial arrangement of cells and in connectivity patterns, areas corresponding to different sensory cortices are indeed quite dissimilar with respect to both physiology and connectivity (Herculano-Houzel et al., 2008), auditory cortex in particular (Read et al., 2002). In non-human primates, A1 is located in the superior temporal region of the brain and is part of the *core* region, a group of three adjacent areas that receive independent inputs from the MGB. These areas then project to a set of seven areas forming the *belt* region, which in turn sends information out to areas in the *parabelt* region (Kaas et al., 1999). Both the belt and the parabelt regions also receive feebler input from the thalamic MGB. The common view at the moment is that the spectrotemporal features are extracted in the core region and then integrated in the belt and parabelt regions to give rise to the perceived representations of auditory objects. Detailed functional analysis in humans has not been performed yet, due experimental limitations. However, a primary auditory region has been identified along the Heschl's gyrus (HG), surrounded by a group of non-primary areas that could potentially be analogous to the belt and parabelt regions in non-human primates (Figure 9). Since our work aims at modeling the extraction of speech features, we will limit the analysis of this Section to A1; we start by looking at the laminar organization of A1 as it has been investigated in various mammals: rodents, cats and non-human primates.

We know that the auditory input signal is mainly transmitted from the thalamic MGB to pyramidal cells inside layers III and IV (named *granular layer* because of the characteristic appearance of layer IV in visual cortex, as a consequence of the high density of pyramidal neurons) of primary auditory cortex, preserving the spatial tonotopy observed in the cochlea and along the thalamic pathway (Hackett, 2011). Other layers receive non-tonotopic input from other thalamic nuclei, being involved in the representation of extra dimensions of sound (e.g. spatial localization of sources). Layer III cells project to A1 in the opposite hemisphere (Code and Winer, 1985), to supragranular (I and II) and infragranular (V and VI) layers with slightly different delays (Atencio and Schreiner, 2010). The prevalent type of neurons in Although many complex interactions arise between layers within one brain area or across different areas, a simplified pictures shows that infragranular layers of hierarchically higher areas

are mainly responsible for feedback projections to supragranular layers of lower areas, while the latter convey feedforward information to granular layers of higher areas (Read et al., 2002; Hackett, 2011). One may also wonder how far do these connections extend horizontally in space, within one layer (these connections are termed *lateral*) or across different layers. More than fifty years ago, the group of Vernon Mountcastle found that neocortex appears to be organized into cylindrical, vertically arranged cortical mini-columns of about 30-90  $\mu\text{m}$  of diameter and 0.5-2 mm of height (Mountcastle, 1997). Mini-columns have been suggested to be a basic computational unit of neocortex, with repeated connections patterns both across layers and between different columns. The size and shape do vary significantly depending on the area, reflecting distinct functional purposes and anatomical changes. In visual cortex, neurons that lie inside the same minicolumn, i.e. that share a similar orientation preference, tend to synchronize at Gamma frequency (Singer, 2013), making an interesting link to brain rhythms. Mini-columns are further arranged into bigger columns, i.e. ensembles of mini-columns that are activated at the same time to perform a given processing step.

### 4.3 Computational models of the auditory system

Every single stage of auditory processing, as described in the two previous Sections, has been the object of detailed computational studies (Meddis, 2010). Some of those models require an extremely high level of complexity given the intricate structure of those areas, the cochlear nucleolus above all. A different kind of approach, more interesting in our opinion, was used in (Chi et al., 2005), where the authors avoided the overwhelming complexity of each single stage model and, instead, based their model on a careful balance of psychoacoustic results and essential neurophysiological findings. The computational model of Chi and collaborators can account for the three main dimensions that make up a sound, which are detected at the early and central stages of the auditory pathway: frequency, temporal and spectro-temporal modulations. The model is made of two parts, a subcortical portion addressing the pathway going from the cochlea to the thalamus, and an early cortical portion simulating A1 neurons. The first tract reconstructs the cochlear time-frequency decomposition plus the aforementioned bandpass filter and selective increase of relevant frequencies. Once the auditory spectrogram is formed in the subcortical part, the cortical segment detects the three modulations modes by modeling neurons according to their spectro-temporal receptive field (STRF), i.e. the two dimensional maps showing the firing rate of a particular neuron as a function of frequency and time occurrence (Meddis, 2010). In the end, the output of the model by Chi and colleagues can be seen as a complex wavelet time-frequency decomposition of auditory inputs (Chi et al., 2005). In our effort to build a model of speech parsing we used a slightly modified version of this model in order to shape the early and central stage auditory pathway, whose output serves as the input to the cortical network of nested rhythms we built (see Hyafil et al., 2015).



**Figure 9. Organization of auditory cortex across different mammals species.**

Identification of auditory cortex regions across six species of mammals. Primary auditory fields are shaded in dark gray, secondary regions appear in white. In humans, primary auditory cortex covers the entire Broadmann area 41 and part of Broadmann area 42. In species where they have been measured, tonotopic gradients are also visible (letter *H* for high frequencies, letter *L* for low frequencies). From (Hackett, 2011).

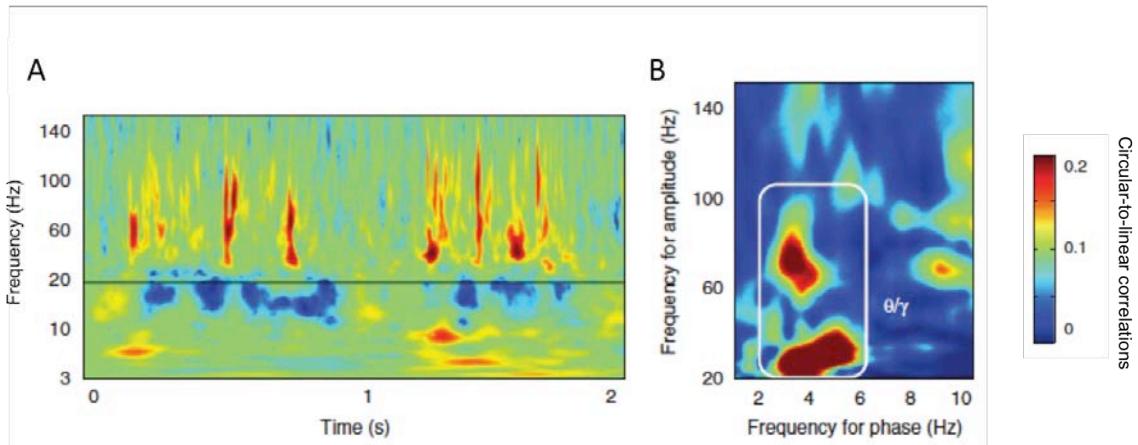
#### 4.4 Speech representation and processing in the brain

Although it would be of absolute interest to investigate the nature of speech representation in the human brain (whether for example the auditory system became progressively tuned to match the speech utterances produced in the vocal tract, or the motor system adapted to the frequency range of human hearing), speech is indeed an enormously complex topic; we will hence try to restrict our analysis to the essentials. First of all, we know that the fundamental units of speech, phonemes, are encoded in relatively sparse cortical neural populations along the superior temporal gyrus (STG) (Mesgarani et al., 2014), and that they are spatially clustered depending on their spectrotemporal features. Furthermore, perceived speech can be reconstructed from the filtered spiking activity of these populations, recorded intracranially in humans (Pasley et al., 2012). However, while the categorization of speech constituents has been recorded with a rather high level of detail, how constituents are actually identified and processed by neural networks is not clear at all. Traditionally, the regions involved in speech production and perception were thought to be limited to Broca's and Wernicke's area respectively, and the left hemisphere was considered to be strongly dominant in language and speech processing. Studies performed in the last fifteen years have progressively overturned this century-old viewpoint, by showing that production and perception of language are much more dispersed functions, implicating several areas of

both cerebral hemispheres (Schirmer et al., 2012; Poeppel, 2014).

The great surge of interest in brain oscillations and their remarkable correspondence to speech timescales lead to what is perhaps the most interesting novelty of auditory research in recent years: the specific role of brain rhythm during the processing of speech (Giraud and Poeppel, 2012). Temporal fluctuations at several distinct timescales are indeed a distinctive feature of speech waveforms (Rosen, 1992), with the information about intonation occurring in the 500-2000 ms range, words in the 300-500 ms range, syllables 120-300 ms and phonemes 20-80 ms. How the human brain can detect and identify these attributes is still not understood, but evidence has emerged that coherent brain oscillations seem related to the processing of speech characteristics (Poeppel, 2003, 2014; Giraud and Poeppel, 2012). Delta (<3 Hz) Theta (3-8 Hz) and Gamma (30-100 Hz) waves have been now recorded extensively during speech recognition, both with intracranial and scalp recording techniques. In particular, slow Delta oscillations are associated with prosody or any rhythmic information contained in perceived sounds (Ding and Simon, 2014). Instead, Theta oscillations have been shown to track the fluctuations in the speech envelope (i.e. the low (3-10 Hz) frequency variations contained in the speech waveform) by resetting their phase upon the presence of syllabic edges (Luo and Poeppel, 2007, 2012; Howard and Poeppel, 2012; Gross et al., 2013; Henry et al., 2014). To further support this claim, the steepness of the rise in power at syllabic edges has been found to positively correlate with the size of the brain entrainment to the envelope fluctuations (Doelling et al., 2014). In fact, the envelope frequency range (1-10 Hz) appears to be crucial for the comprehension of speech: experiments demonstrated that the intelligibility of a sentence is greatly reduced (if not destroyed) when 1-7 Hz fluctuations are filtered out of speech waveforms (Elliott and Theunissen, 2009), and, at the same time, phase-locking within the same band is improved if speech is properly understood (Peelle et al., 2013). Work from Ahissar and collaborators (Ahissar et al., 2001) also showed that speech comprehension correlates with the brain capacity to track the low frequency envelope (0-20 Hz), and that both drop when speech is time-compressed by a factor bigger than three, which would correspond to a syllabic rate higher than about 9 Hz. Syllabic rhythmicity is so important that the human brain is capable of understanding speech even in very much degraded conditions, if syllabic periodicity is not fully destroyed (Ghitza and Greenberg, 2009). With this amount of evidences, the entrainment of intrinsic Theta waves can be a viable and effective mechanism to detect the envelope dynamics, and hence to chunk incoming speech into subunits of about 110-300 ms, i.e. the average syllabic temporal duration (Shamir et al., 2009; Ghitza, 2011). Concurrently, slow fluctuations in speech also overlap with the slow temporal dynamics of Gamma power (Nourski et al., 2009; Pasley et al., 2012; Kubanek et al., 2013; Zion Golumbic et al., 2013), suggesting that Gamma activity might reflect the encoding of fine-grained phonemic features enclosed within the syllabic boundaries. Furthermore, both intracortical and scalp human recordings in auditory regions of subjects that were listening to speech (Nourski et al., 2009; Giraud and Poeppel, 2012; Morillon et al., 2012; Gross et al., 2013) reported significant cross-frequency coupling between Theta and Gamma bands (Figure 10). Based on these experimental facts, Ghitza and collaborators proposed a conceptual model to illustrate a potential role of brain rhythms in the parsing of speech (Ghitza,

2011). While previous models proceed by first recognizing and categorizing phonemes by template matching (Stevens, 2005), the claim in Ghitza's model is radically different: syllables are decoded as long as they are enclosed inside a time frame defined through an internal periodic clock running in the Theta frequency range (Figure 11). A clock ticking in the Beta frequency range ( $\sim 20$  Hz), coupled to the Theta rhythm so that its frequency is a multiple of Theta frequency, tracks dyadic<sup>2</sup> content, and an additional faster-paced clock potentially parses phonemic spectrotemporal features within the Gamma frequency band. Template matching is performed at each stage, from syllables to phonemes. A key requirement for this model to be able to explain observed behavioral results is that the Theta clock has to align to the beginning of syllables, at least to some extent. When the alignment is lost, a Theta time frame might fall in the middle of a syllable, making the decoding much harder if not impossible.



**Figure 10. Theta-Gamma frequency coupling in auditory cortex.**

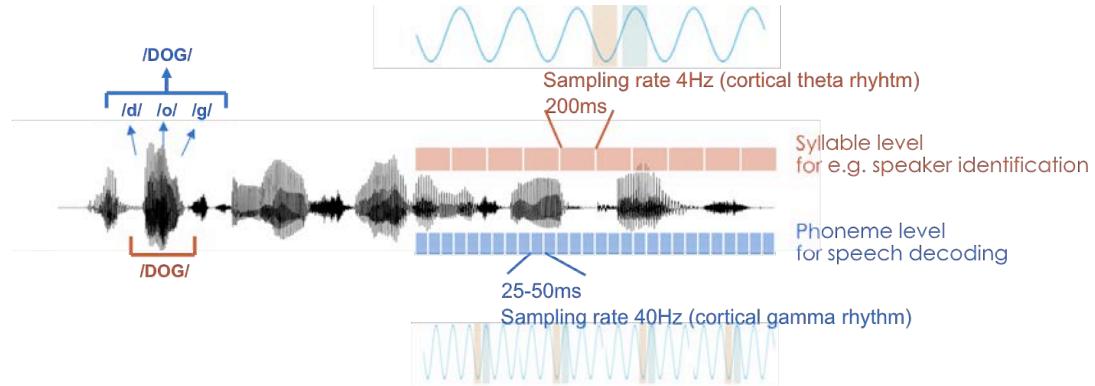
Intracranial EEG data from epileptic human patients. **A)** Spectrogram of brain activity in human primary auditory cortex during listening of speech. Time-frequency decomposition had been computed using complex Morlet wavelets and a baseline correction had been applied using a portion of the prestimulus period. Activation in the Theta (3-8 Hz) and Gamma (30-120 Hz) frequency bands is visible in red. **B)** Phase-amplitude coupling between low and high frequency rhythms in human primary auditory cortex, measured using circular-to-linear correlation. Significant coupling occurs between the phase of 2-6 Hz Theta oscillations and the amplitude of 20-80 Hz Gamma oscillations.

#### 4.5 Multiplexing and oscillations

The results discussed in the previous Section lead to the hypothesis that acoustic features can be encoded using distinct, although coupled, cortical rhythms, allowing the auditory hierarchy to achieve some kind of *multiplexing*, i.e. to make different areas communicate with each other by using a number of independent frequency channels

<sup>2</sup> Dyadic units were first defined by (Peterson et al., 1958) as i) containing parts of two phonemes with their overlapping spectrotemporal features somewhere in the middle; ii) having its beginning and ending at the phonetically most stable position of each phoneme.

but only a single transmission line (i.e., in neural terms, a single readout population). Multiplexing is an appealing strategy for the brain: it would favor interareal communication and it could optimize the encoding/decoding of features pertaining to different stimuli (Akam and Kullmann, 2010, 2014; King and Walker, 2012). Recently, it has been proposed that the auditory system could use a multiplexed transmission line in order to segment speech, formed by a hierarchy of oscillators coupled together (Ghitza, 2011; Giraud and Poeppel, 2012; Gross et al., 2013). As a matter of fact, a potential explanation of the observed Theta-Gamma CFC in auditory cortex (see Figures 10 and 11) would be that Theta oscillations are able to entrain to the syllabic fluctuations of incoming speech, thereby providing a window of opportunity for Gamma neurons to fire and transmit sensory information to areas higher in the auditory hierarchy, in a bottom-up fashion (Giraud and Poeppel, 2012). Such a scenario would imply that stimulus information is passed to upper regions in a multiplexed manner, at least at the syllabic and phonemic timescales. From these hypotheses, we were the first to construct a computational implementation of Theta-Gamma CFC in auditory cortex using neuronal networks, in order to verify the advantages of such neural architecture and provide a set of useful and testable predictions for future experiments (Hyafil et al., 2015). The important implications of our work are discussed later in this thesis. Let us now review a set of fundamental concepts regarding the interactions between distinct areas of the brain and the potential functions underpinning those mechanisms, since they will be important to introduce our work on frequency-specific communication channels (Fontolan et al., 2014).



**Figure 11. Speech waveform and its chunking by cortical rhythms.**

Syllabic (red) and phonemic (blue) components are both represented in the speech waveform. Cortical oscillations in the Theta and Gamma frequency bands respectively match the timescales of syllables and phonemes, hence constituting a potential mechanism to parse the speech signal into its constituents.

## 5 Information flows in sensory cortex

*In one of his gloomier moments Pascal said that all man's unhappiness steamed from a*

*single cause, his inability to remain quietly in a single room. "Notre nature", he wrote, "est dans le mouvement... La seule chose qui nous console de nos misères est le divertissement". Diversion. Distraction. Fantasy. Change of fashion, food, love and landscape. We need them as the air we breathe. Without change our brains and bodies rot. The man who sits quietly in a shattered room is likely to be mad, tortured by hallucinations and introspection.*

Bruce Chatwin, *Anatomy Of Restlessness*

At the dawn of theoretical neuroscience, scientists used to think of the brain as a passive machine, a collection of sophisticated filters capable to extract and interpret the information coming from the surrounding world, with the primary aim of prolonging survival and ensuring abundant and healthy offspring. Although this vision has proven extremely useful to explain some of the animal and human behaviors and the corresponding brain responses, it does not appear to be the most parsimonious and flexible way of operating for a device whose target is to link the perception of sensations with actions (Friston, 2010). There is evidence that the brain anticipates future events such as upcoming sensory stimuli, and adapts its reactions accordingly (Bar, 2007). To use the words of Daniel Dennett: "*A mind is fundamentally an anticipator, an expectation generator*"<sup>3</sup>. Intuitively, being able to predict stimuli and contexts from the environment constitutes an invaluable advantage for survival: it allows for optimizing actions with respect to the complexity of the environment, anticipating dangerous situations and maximizing the responses to relevant, unexpected events. Hence, the brain needs to carefully balance and integrate together the information flowing from peripheral sensory systems to cognitive areas (*bottom-up* flow) and the predictions traveling from cognitive areas to sensory cortices (*top-down* flow).

One of the most popular and interesting theories on prediction has been advanced by David Mumford about two decades ago (Mumford, 1992). The theory, that has been successively repurposed by other researchers (Rao and Ballard, 1999; Friston, 2002), is called *predictive coding* and revolves around two principles: i) it allows to link cognitive perception with passive sensing, by inferring the most probable causes of incoming stimuli (as first proposed by Helmholtz<sup>4</sup>) and ii) it actively maintains a model of the world, based on previous experience and future expectations, that must be plastic enough to be constantly updated by new experiences (in healthy individuals). In this chapter we will review previous models of predictions, then introduce the predictive coding theory and highlight the role of oscillations in relation to top-down vs. bottom-up communication.

## 5.1 The connectionist approach

A connectionist model typically consists in a network of interconnected minimal processing units, often displaying binary values [0,1]. The units are interpreted as simplified neurons, mutually connected by adjustable weighted links (Amari, 1972). The

---

<sup>3</sup> Dennet, D., 1996, *Kinds of minds: toward an understanding of consciousness*, Basic books.

<sup>4</sup> Helmholtz, H., 1860/1962, *Handbuch der physiologischen optik* (ed. J. P. C. Southall), vol. 3. New York: Dover.

output of such basic neuron is computed as some sort of weighted sum of its inputs, usually thresholded or modulated by a nonlinear function (e.g. a sigmoid). The value of the output after thresholding is called the *state* (or activation) of the unit, which, in turn, might constitute part of the weighted input to other units. Surprisingly, networks built out of these elementary units are able to capture a remarkable number of features of high order cognitive processes, such as memory (Hopfield, 1982), language (Norris, 1994) and vision (Feldman, 1985). Although these artificial networks are generally employed to describe bottom-up processes, they have been also applied to try to understand and explain prediction or anticipation. To be able to form any kind of useful prediction, the brain must be able to *learn* from previous experience, which in the connectionist paradigm translates into allowing processed stimuli to actively modify the distribution of the weights to minimize the mismatch between the network prediction and the actual stimulus. Irrespective of the specific error minimization principle used in the different instances of connectionist models, learning in these networks has to be supervised or semi-supervised. Learning in a supervised fashion means that an external source is needed to train the network, by knowing in advance the desired target output (i.e. by guiding the weights modifications towards the correct input-output associations using a subset of all input patterns). Unfortunately this approach has two main limitations in relation to predictions: first, these networks must know in advance many of the causes of incoming sensory input in order to be able to learn a set of appropriate input-output mappings; second, they often need a vast training corpus to reach reasonable performances, making them too slow or fragile to be useful to model predictive processes in the brain.

## 5.2 Predictive coding

To overcome the drawbacks of connectionist models David Mumford proposed a different approach based on what is known as the predictive coding model (Mumford, 1992). In this framework, the error that is minimized is not that between the network output and the target pattern, but the one between the actual observed input and the prediction from the network. This way, in predictive coding, the network is able to learn the generating causes behind the observed sensory patterns (i.e. updates its internal model of the world) by i) extracting the statistical regularities of the inputs; ii) comparing the predictions formed on the basis of previous experiences with the upcoming input signal. To translate this scheme into a potential cortical implementation, Mumford assumed that the “abstract” information regarding the world causes was retained in some central brain area, which in turn sent the information about the most expected signal to some lower area (top-down flow). The local circuits within each area computed the difference between the sensory input and predictions coming from upper areas (the *prediction-error*) and transmitted this information back to upper stages (bottom-up flow), where the prediction signal is updated accordingly. In the Mumford model not only the predictive step and the update step occur at the same time, in contrast to the connectionist approach where a training and a test period are needed, but the learning process is also completely unsupervised (Friston, 2002). After Mumford’s proposal, this approach has been used to successfully model extra-classical receptive fields in visual cortex (Rao and Ballard, 1999) and extended to a whole hierarchy of cortical stages, each one passing the predictive information to the next one

in a top-down fashion, and getting back the prediction-error signal in a bottom-up fashion (Friston, 2005; Kiebel et al., 2008). Multiple studies have then found evidence for predictive coding during cortical processing of information, under paradigms as diverse as mismatch negativity (Waengne et al., 2011; Winkler and Czigler, 2011), spoken word recognition (Gagnepain et al., 2012), priming (Recasens et al., 2014), visual adaptation to novel stimuli (Hosoya et al., 2005) and repetition suppression (Costa-Faidella et al., 2011). We refer the interested reader to (Clark, 2013) for a comprehensive and updated review on the subject.

Beyond the general idea of predictive coding, many implementations are being considered by researchers (Friston, 2003, 2008). A promising approach, partially supported by evidence from cortical connectivity studies (Markov and Kennedy, 2013), is that of Bayesian probabilistic inference and generative models (Dayan et al., 1995; Friston, 2002, 2003). A generative model is a way of reconstructing the sensory input from the representation of its causes that is stored in the network. In practice, a generative model performs the inverse of inferring the causes from sensory inputs: it generates a model of the observed sensory input starting from the most probable underlying cause. Consider as an example the shadow produced by an animal crossing the street under the lights of a car: to infer to which animal belongs the perceived shadow, a generative model would generate a copy of the shadow caused by the animal we encountered more often in our past experience. The predicted shadow is then compared to the observed shadow, and the corresponding prediction-error signal is produced, depending on whether the two patterns match or not. Cortical circuits might build similar predictions based on generative models and update the parameters of the model by means of Bayesian probabilistic inference (Friston, 2005; Bastos et al., 2012). Indeed, it is worth noting that a purely feedforward network cannot perform predictive coding: in fact backward connections from lower to upper stages are crucial for the update of the internal model (Friston, 2002).

Bastos and collaborators also suggest a possible spatial scale within which the processes of predictive coding might be integrated: cortical macrocolumns (see Section 4.2). These columns could be the fundamental units of cortical microcircuits for predictive coding: in fact, each column is computationally segregated from other columns belonging to the same brain area (i.e. is capable to accomplish the full computational duties assigned to its stage), but interacts with columns in other brain areas. In particular, although the physiological organization of cortical is all but simple, previous computational models (Mumford, 1992; Rao and Ballard, 1999; Bastos et al., 2012) putatively assigned deep layers of the column to the computation of predictions, then transmitted via top-down pathways to lower areas. Similarly, cortical superficial layers are associated with bottom-up transmission of prediction-error, i.e. the difference between the observed and the predicted input. Bastos and colleagues also suggest that neurons in deep layers should display lower powers at high frequencies, due to the longer window of integration needed to form predictions (Bastos et al., 2012). This could indeed explain the spectral differences observed across layers (see next Section), and, together with the mapping of cortical layers to communication mechanisms, suggests a strong link with cortical oscillations, as we shall see here below.

### 5.3 Oscillations and information passing

As we have seen in Section 3, brain rhythms are a consequence of the interplay among neurons, whose dynamics allows for the creation of transient synchronized assemblies. Different rhythms are created by distinct neuronal networks: not only their neurons have distinct electrical properties, but they are also located in different layers within the cortical sheet: Gamma oscillations were shown to be generated in superficial layers (II/III) of neocortex, while slower rhythms, like Beta (13-30 Hz), originate in deep layers (IV/V) (Roopun et al., 2008b; Wang, 2010; Buffalo et al., 2011; Cannon et al., 2014). In addition, several studies have demonstrated that the spectral cortical response to sensory inputs is strongly affected by top-down cognitive control, such as attention and predictions (Wang, 2010; Bosman et al., 2012). Among these studies, those on speech perception are particularly relevant, since frequency bands appear to be particularly affected by top-down mechanisms during the listening of speech (Arnal and Giraud, 2012; Ding and Simon, 2012; Gagnepain et al., 2012; Zion Golumbic et al., 2013; Arnal et al., 2014).

These two facts, the cortical separation of rhythm generators and the evidence that alterations in the frequency spectrum seem to correlate with top-down cognitive tasks, suggest a potential fundamental role for oscillations as powerful communication channels to convey information across different brain areas (Buschman and Miller, 2007; Wang, 2010; Arnal and Giraud, 2012). On the one hand it is well established that Gamma oscillations are strongly related to sensory processing (Buzsáki and Wang, 2012), even though the directionality of the flow has been established only recently (Bastos et al., 2014; Fontolan et al., 2014; van Kerkoerle et al., 2014). On the other hand, Beta oscillations, although classically associated with sensorimotor planning and execution, have been recently associated with top-down modulations and predictions in sensory cortices, thanks to longer conduction delays that could maintain synchrony over larger spatial scales and thus provide a long integration window, as needed to form predictions (Engel and Fries, 2010). According to this hypothesis, Beta could signal the maintenance of the current internal representation (the status quo) that is generated in high-level cortical areas. Consequently, Beta power would be suppressed in presence of a novel, engaging input capable of modifying the status quo, as it is actually observed experimentally (see Engel and Fries, 2010). An interesting study by Arnal and collaborators (Arnal et al., 2011), which has investigated the audiovisual integration of speech with MEG recordings, highlighted the significant alterations in the frequency spectrum provoked by mismatching audio and visual inputs. Essentially, the authors observed stronger cross-frequency coupling between Beta and Gamma rhythms in the superior temporal sulcus, together with enhanced Gamma power in auditory and visual cortex, when the visual input was incongruent to the auditory input. Hence, higher Gamma amplitude could underlie the computation of prediction-error, whereas higher Beta-Gamma PAC might signal a change in the status quo and the subsequent construction of a “new” prediction, that would solve the incongruity in favor of either the auditory or the visual sensory channels (Arnal and Giraud, 2012).

#### 5.4 Summary of hypotheses

Although their functional role has not been elucidated yet, brain oscillations appear to be ubiquitous in the mammalian brain. Gamma oscillations (30-100 Hz) in neocortex have been associated with the processing of sensory inputs, while slower rhythms like

Theta (3-8 Hz) have been related to various aspects of cognitive processes like working memory or the modulation of sensory and motor signals. Experimental evidence showed that oscillations at separate frequency bands are *nested*, i.e. they are linked through phase-amplitude coupling, in many sensory areas. This is particularly true in the auditory cortex during speech perception. In fact, to capture the many different relevant components of speech (syllables, phonemes), the brain must be able to parse the speech signal over these different time-scales at the same time. A network of nested theta-gamma rhythms could accomplish this task, given the matching of theta and gamma frequency with syllabic and phonemic time-scale respectively (Giraud and Poeppel, 2012). These hypotheses were tested in Articles 1 (Hyafil et al., 2015) and 2 (Fontolan et al., 2013).

In addition to the processing of bottom-up sensory information, the brain is also trying to predict and anticipate future events, according to a recent theory of brain functioning called *predictive coding*. In fact, partial evidence shows that the central neural system minimizes its reaction to environmental stimuli by predicting likely events, and inferring their most probable causes (Friston, 2005). This mechanism ensures that reactions are appropriate, i.e. maximal for unexpected events and minimal to frequent ones. Functionally, predictive coding is a possible realization of the anticipatory function of the brain. Practically, it states that the difference between top-down predictions and bottom-up sensory information is assessed at each processing stage (i.e. brain area), possibly within a each single cortical column scale, so that only the error signal (i.e. the magnitude of the mismatch) is further propagated up the cortical hierarchy. It has been then conjectured that top-down and bottom-up information could be transmitted throughout the sensory hierarchy via distinct frequency channels: in the Beta and Gamma band, respectively (Arnal and Giraud, 2012). In Article 3 (Fontolan et al., 2014) we tested this theory using a set of iEEG recordings in human subjects.

## 6 Description of the articles

### 6.1 Article 1: Speech encoding by coupled cortical theta and gamma oscillations

Alexandre Hyafil, Lorenzo Fontolan, Claire Kabdebon, Boris Gutkin and Anne-Lise Giraud

*eLife, in press*

**Contribution to the article:** Conception and design of the model; Analysis and interpretation of data and results; Drafted and revised the manuscript.

#### 6.1.1 Summary of results

In this study we created a neural network model of the microcircuits that, putatively, generate Theta and Gamma oscillations inside auditory cortex. Each rhythm arises from a network of mutually interconnected excitatory and inhibitory quadratic integrate-and-fire neurons (PING). The excitatory neuronal population, which simulates pyramidal

cortical neurons, starts firing upon receiving the spectrogram of a spoken sentence, processed by a state-of-the-art cochlear model. Through a number of excitatory synapses, spikes from pyramidal neurons activate the population of inhibitory interneurons. Inhibition then reaches pyramidal neurons, creating a refractory period during which inhibition prevents excitatory neurons from reaching the spiking threshold. The two PING-like networks are then connected to reproduce phase-amplitude coupling. We tested the model capability to chunk and extract speech constituents using many different measures on the standard Texas Instruments & Massachusetts Institute of Technology (TIMIT) speech corpus. We showed that the Theta-Gamma network performs as well as the best offline methods when tested over the tracking of syllabic boundaries, and that syllabic constituent can be recognized and classified using spiking activity of Gamma excitatory neurons. Crucially, the Theta-Gamma model performed better than control models where i) the connection between the two rhythms was cut and ii) when the Theta network receives no speech input. In Article 2 we focused on the synchronization properties of network analogous to the one we used in Article 1, employing the most advanced mathematical tools to analyze the different dynamical regimes arising in the parameter space.

## 6.2 Article 2: Analytical Insights on Theta-Gamma Coupled Neural Oscillators

Lorenzo Fontolan, Maciej Krupa, Alexandre Hyafil and Boris Gutkin

*Journal of Mathematical Neuroscience* (2013) 3:16

**Contribution to the article:** Conception and design of the model; Analysis and interpretation of data and results; Drafted and revised the manuscript.

### 6.2.1 Summary of results

We considered here a simple mathematical model of a fast spiking network (i.e. firing in the Gamma frequency band, 30 - 100 Hz) modulated by a slow periodic input (i.e. operating in the Delta-to-Theta frequency range, 1-8 Hz). In particular, we connected a single Excitatory Gamma (GE) neuron, modeled as a QIF, to an excitatory sinusoidal input, whose natural frequency lies in the Theta band. The GE neuron also participates in a Pyramidal Interneuron Network Gamma (PING) rhythm, although in our case the inhibitory neuron is instantaneously enslaved to the excitatory cell, meaning that every excitatory spike immediately prompts a virtually simultaneous inhibitory spike. We found that that this network can work in two distinct regimes (which we have named *excitable* and *oscillatory*), which respond in rather different ways to the Theta input. In the excitable case the GE neuron remains silent at troughs of Theta rhythm and then starts firing almost immediately when Theta input has reached a certain threshold. On the contrary, in the oscillator case the GE cell always fires, but its frequency of spikes is modulated by the phase of Theta. For both regimes, we provided analytical and numerical solutions to compute the time-to-first spike at the beginning of a Theta period and the average number of Gamma spikes inside one Theta cycle.

### **6.3 Article 3: The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex**

Lorenzo Fontolan\*, Benjamin Morillon\*, Catherine Liegeois-Chauvel, Anne-Lise Giraud

\*equal contributions

*Nature Communications* (2014) 5:4694

**Contribution to the article:** Analysis and interpretation of data and results; Drafted and revised the manuscript.

#### **6.3.1 Summary of results**

With the aim of exploring the frequency-specific content of information passing in auditory cortex, we investigated the directional influence in the frequency domain between A1 and associative auditory cortex (AAC) in epileptic patients that were recorded intracranially while they were listening to speech. We were able to identify the frequency channels of top-down and bottom-up communications during speech perception, being, respectively, Beta and Gamma bands, using *circular-to-linear correlation* to measure phase-amplitude coupling and *Granger causality* to reveal causal patterns between A1 and AAC at a specific frequency, the same for both areas. Lastly, we found that T-D and B-U causal patterns fluctuate over time, i.e. windows of prevalent T-D influence alternate with windows of B-U influence at a rate of about 2-3 Hz. This suggests that multiplexing does not occur simultaneously but in temporally distinct segments.

## **7 Discussion**

### **7.1 Foreword**

Brain oscillations have been recorded in thousands of experiments in humans and in more than a dozen animal models, throughout many different brain states and behavioral tasks. Recent efforts were aimed at revealing the potential role of oscillations in functions as diverse as the mapping of spatial locations (Buzsáki and Draguhn, 2004), facilitation of input selection (Hutcheon and Yarom, 2000; Akam and Kullmann, 2010), coordination of distant brain areas (Fries, 2005; Womelsdorf et al., 2007), binding of perceptual features (Engel et al., 2001), formation and persistence of memories (Jensen and Lisman, 2005; Axmacher et al., 2006), dynamics of synaptic plasticity (Bukalo et al., 2013), or the execution of sensory-motor functions (Schöffelen et al., 2005). Still, conclusive evidence on whether oscillations in cerebral cortex are a mere epiphenomenon or a useful computational strategy that allows the brain to better perform some specific classes of tasks is currently missing. In order to unravel the “cortical oscillations mystery” we must be able to answer a number of compelling questions that arose in past years. The work of this thesis is an attempt to find the answer to some of those questions. With our computational and analytical work we

tackled two questions: i) linking oscillations to a specific computational function (in our case in the auditory cortex) and ii) demonstrating that a hierarchy of oscillations is capable of enhancing the encoded stimulus information during speech processing, as compared to control conditions. We discuss the implications of this first part of our work in Section 7.2. On the other hand, the analysis of human intracranial recordings allowed us to establish the frequency content of the information flow within the auditory hierarchy and to uncover its temporal dynamics. The consequences of our experimental results are examined in Section 7.3.

## 7.2 Modeling speech processing in auditory cortex

Speech signal is an exceptionally attractive sensory stimulus to neuroscientists, given its rich spectrotemporal contents where features are encoded at several distinct timescales. Even more, speech is particularly suited for testing the contribution of brain oscillations in sensory perception, thanks to the quasi-periodic structure of its constituents. Nevertheless, speech perception is all but a simple matter: there seems to be no easy mapping between the continuous acoustic waveform and the perceived linguistic units such as abstract phonological representations, which have been found to be represented and categorized in the human brain (Mesgarani et al., 2014). Despite this complexity, the fact that there exist very few neural models of speech perception remains undoubtedly striking. On one side, a number of psycholinguistic and cognitive models of speech have been proposed. Among the first influential models it is worth mentioning the *motor theory*, advanced by Alvin Liberman and collaborators (Liberman et al., 1967; Liberman and Mattingly, 1985), that linked perceived phonemes to the articulatory movements of the vocal tract. The motor theory had the merit of uncovering the strong mapping between phonemes and the underlying speaker's gestures. However, its initial claim that speech production and perception share a unified and innate representation has been refuted by subsequent findings, beside the various perceptual effects that the model cannot account for. The attention of modelers then shifted to the apparent acoustic invariance of speech, in the attempt to find a set of invariant spectral properties that could lead to categorization of phonemes. Two interesting connectionist models attempted this approach: *TRACE* (McClelland and Elman, 1986) and *Shortlist* (Norris, 1994), which reached fairly good performances in recognizing words. Both these models are essentially based on multilayer neural networks, where the bottom layer encodes a number of features that are believed to be important for phonemic categorization, while nodes in the two upper layers correspond to phonemes and words, respectively. The main limitation of these models lies in the fact that the tuning of neural units to the phonemic set has to be imposed artificially, since brain recordings and behavioral results do not seem to suggest that phonemic spectral features can uniquely determine the classification of phonemes (Poeppel and Monahan, 2008). Although still highly debated, the key turning point came from results on how the brain encodes the temporal modulations of speech waveforms (Greenberg and Ainsworth, 2006). In fact, the fundamental sources of acoustic variability in the waveform all involve some kind of temporal transformation: i) variability in spectral modulations, ii) variability in the duration, iii) stretching or compression of acoustic subparts within speech constituents. Notably, brain oscillations have the ability to quickly adapt to temporal changes, a claim that is supported by the experimental evidence we reviewed in Sections 4.4 and 4.5.

While there is now a relatively large corpus of experimental evidences linking speech and oscillations, only the *TEMPO* model of Ghitza and colleagues (Ghitza and Greenberg, 2009; Shamir et al., 2009) has claimed a role for brain oscillations in the processing of speech. Our paper (Hyafil et al., 2015) is the first to propose a biophysically plausible, although much simplified, cortical circuit, making use of realistic neuronal networks to generate the rhythmic patterns observed in auditory cortex during speech perception. Concerning the fast rhythm, we implemented a set of biophysical parameters and connectivity structure, based on the PING model, that match the experimental values of Gamma rhythm in sensory cortices, (Cardin et al., 2009; Vierling-Claassen and Cardin, 2010), in particular in the auditory cortex (Nourski et al., 2009). It is less clear how slow frequency oscillations such as Theta are generated in cortical microcircuits, albeit having been extensively recorded in human auditory cortex both intracranially and from the scalp (Ahissar et al., 2001; Luo and Poeppel, 2007; Morillon et al., 2012). Hence, we used an implementation analogous to that we had used for Gamma oscillations, based on excitation followed by long decaying inhibition (lasting about 5 times longer than inhibition in the standard PING), which we called *PINTH*. We know in fact that there exist cells whose inhibitory decay constants are comparable to the value we used in the model, such as for example Martinotti cells, which are found in the rat's neocortex (Silberberg and Markram, 2007b). Obviously, the *PINTH* architecture has not yet been verified experimentally, but we believe that part of the value of a computational model lies in the possibility to offer a set of concrete predictions that can then be corroborated or disproved. In this particular case, our prediction would be that Theta rhythm arises from an excitatory-inhibitory loop similar to that of the PING. In addition, our low frequency circuit is not expected to increase Theta band power but to reset its phase when in presence of a steep rise in the speech waveform's envelope, as it has been shown in various human studies (Luo and Poeppel, 2007; Gross et al., 2013). The two rhythms received rather different inputs: Theta neurons were fed with the broad spectrum in order to follow the slow global fluctuations in power; each Gamma neuron, instead, was connected to a different cochlear channel to reflect the variety of spectrotemporal filters represented in the auditory cortex.

A second critical aspect of our model is Theta-Gamma PAC, which is realized through a bundle of synaptic connections between the excitatory population belonging to the *PINTH* (*TE*) and the one belonging to the PING circuit (*GE*). Since we assume, in our model, that *GE* neurons are the readout cells from which the relevant information regarding speech constituents can be decoded, their firing pattern becomes particularly important to examine. Intuitively, the spiking activity of these neurons should be more intense at the beginning of the syllable, so that it can provide maximal information and activate the appropriate predictions to anticipate subsequent phonemes, syllables and even words (Schroeder and Lakatos, 2009; Giraud and Poeppel, 2012). Indeed, our architecture nicely reproduces this effect, as a consequence of the fast Theta phase reset and the ensuing strong input reaching *GE* cells. It is worth noting a crucial feature of our model with respect to other previous computational works: with our implementation, the detection of the syllabic onset comes for free from the biophysical structure of our network, while other speech recognition models relied on either an ad

hoc neural code that has never been observed experimentally (Gütig and Sompolinsky, 2009) or an external, artificial, onset signal (Hopfield, 2004; Shamir et al., 2009), to which the brain cannot have access. Theta neurons also act as a sort of internal clock for other neuronal populations, signaling the onset of speech stimuli. Neurons with these properties have been found in A1 of monkeys (Brasselet et al., 2012; Panzeri et al., 2014).

Although the issue of whether Gamma rhythm can be used to process information or not is still debated, we think that this does not affect our model. Criticism over Gamma regards, in fact, the scarce contribution of Gamma to the power spectrum (about 10%), incompatible conduction delays that would make difficult to define a proper phase, and the strong dependence of Gamma activity on the presence of a stimulus (Ray and Maunsell, 2014). Altogether, these arguments cast into question whether Gamma can be used for coding or to process sensory information. However, for our model to work properly, Gamma does not necessarily have to be a rhythm *stricto sensu*; conversely, our Gamma rhythm is a rather sparse and weak rhythm, resulting from noisy excitatory-inhibitory feedback interactions, like those of the PING. As a consequence, individual GE cells may skip one or more Gamma cycle, producing a rather broad Gamma peak and making it hard to reliably measure a global phase. The sparseness of the rhythm is essential to maintain a balance between synchronization, which reduces the entropy of the spike trains and diminishes the information carried by GE spikes (Eyherabide and Samengo, 2013), and the preservation of a somewhat spread temporal distribution of GE spikes, which is essential to efficiently encode and classify different stimuli (Strong et al., 1998). To this end, we measured the mutual information (MI) between the stimuli and three neural codes: i) the spike count code, which simply computes the sum of all spikes within a specific time window; ii) the spike pattern code, which also considers the temporal order of spikes within that window; iii) the phase code, attributing an additional tag to each spike according the phase of Theta oscillation at which the spike occurs. Our MI plots are surprisingly similar to those obtained from *in vivo* recordings of monkey auditory cortex (Kayser et al., 2009), suggesting that our model correctly grasps, at least qualitatively, how information is encoded in that brain area.

We then established the model's performance to decode and classify two classes of stimuli: simple sawtooth waves, mimicking the syllabic shapes (Shamir et al., 2009), and syllables within English sentences, randomly selected from the TIMIT corpus (J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, 1993). The classifier was based on a clustering method (see Article 1), the readout neurons were, as mentioned above, the GE cells, and the neural codes we tested were the same we used to compute MI. The percentage of correctly decoded stimuli was very good for both sawtooth waves (60% using the best code, i.e. spike patterns; 10% chance level) and for real syllables (58%). As expected, it dropped dramatically in the two control conditions (i.e. in the absence of a Theta-Gamma connection or in presence of a Theta module disconnected from the input). We performed an additional classification analysis where, instead of aligning GE firing activity with the actual time-course of the sawtooth stimulus (*external* time reference), we aligned it with the onset of the theta oscillations produced by the network (*internal* time reference), which would correspond to the *perceived* time-course (see Section 4). This way we looked at the ability of the system to use an endogenous clue,

i.e. the onset of a Theta burst, as an internal clock. The performance degraded to about 42% for sawtooth waves, remaining however much above chance level.

Finally, we tested the model's ability to process time-compressed speech, as one of the significant advantages of using oscillations is the resilience of the phase-resetting mechanism to time-compression, at least to a certain degree. Behavioral data indicate that speech can be uniformly squeezed up to 3 times its standard duration before intelligibility (the degree to which speech can be understood) dramatically drops (Ghitza and Greenberg, 2009). Similarly, the preferential window provided by Theta oscillations can be aligned to the syllabic contours only within a certain compression range: the performance of our model was still above statistical significance up to a compression factor of 3, but it then plunged to chance level for higher compression factors. However, the difference in the classification performance of syllables was quite big between standard uncompressed speech and compression rates of 2 or 3. Our network is in fact purely bottom-up, therefore it lacks any predictive/anticipatory top-down effect that we know is a very important factor in speech processing, and which would probably explain the performance gap between standard and compressed intelligible speech.

We believe that more computational models are needed in neuroscience, and especially in audition where neuronal networks were only recently applied to model neural processes. Our model is an attempt to formulate a set of basic principles underpinning the chunking of speech and the extraction of its constituents in early auditory areas. We do not claim the model to be exceptionally close to the actual biophysical architecture of auditory cortex. Rather, although inspired by presently available biophysical evidence, we aimed at developing a reasonably simple network whose performance could be tested under the most important manipulations, and which could be used for making straightforward predictions. Nevertheless, one of the potential dangers of computational models comes from overfitting and fine tuning of simulations: any behavior can be reconstructed and reproduced with a sufficiently high number of parameter, whereas the underlying computational architecture most likely has nothing to do with the real architecture, resulting in very poor predictive performances. This particular concern was the main motivation for our purely theoretical investigation of nested oscillators, which lead to the publication of Article 2. The analytical and numerical results we reported in Article 2 (Fontolan et al., 2013) are important in order to understand the underpinning laws regulating the interaction between fast and slow neural oscillators. The theoretical discovery of two regimes with rather different synchronization properties allows theorists to make predictions regarding the dynamics of networks observed *in vivo* based on the dynamics of their elements. Evidence that these two coupling regimes might have been recorded in the hippocampus has appeared recently (Cabral et al., 2014).

### 7.3 Predictions and perception

Throughout the twentieth century, the majority of neuroscientists looked at the brain as a collection of passive filters, whose main job consisted in eliminating worthless information from external inputs. In the last twenty years, however, an alternative vision of perception has increasingly gained popularity, which implies that the brain actively maintains an internal model of reality, formed on the basis of previous

experience and used to predict future occurrences (see Section 5). One of the most famous and influential implementations of this view is predictive coding, as seen in Section 5.2. In this framework, perception is formed as the interaction between stimulus-driven, bottom-up, signals and top-down expectations, throughout the various stages of the sensory hierarchy (Rao and Ballard, 1999). Many recent experiments have indicated that oscillations are implicated in the formation and transmission of predictions, due to their dynamic properties that make them particularly attractive to explain T-D and B-U interactions.

Our work was motivated by three major inquiries: i) investigate the frequency channels employed in the abovementioned scheme by applying a directional measure, ii) unravel the presence of cross-frequency coupling, iii) search for any temporal pattern of modulation in the directional influence. With these questions in mind, we processed and analyzed a set of rare and thus precious intracortical human EEG data from epileptic patients, obtained while they were listening to uttered sentences. The stereotactic electrodes were placed along the auditory hierarchy, hence giving us the opportunity to investigate the relationship between an early sensory area such as A1 and a more associative, higher order area such as AAC.

To answer the first of our inquiries, and assess the causal influences between A1 and AAC as a function of frequency, we used a nonparametric version of Granger causality (Dhamala et al., 2008). Normally, GC is measured by first modeling the EEG time series using a multivariate autoregressive process, but this method does elicit several potential issues that would undermine the trustworthiness of the results (Dhamala et al., 2008; Barnett and Seth, 2011). Instead, nonparametric GC leans on a different estimation method that is more robust and reliable (Ding et al., 2008). Our method was relatively new in neuroscience, since very few articles had reported frequency-specific causal influence when we started analyzing this dataset, e.g. from Steven Bressler's lab to reveal Beta oscillations in sensorimotor cortex (Brovelli et al., 2004), and a few others appeared after we began our endeavor (Bosman et al., 2012; van Kerkoerle et al., 2014). Applied to our dataset, GC revealed a clear causal pattern between A1 and AAC: low frequencies, from Delta to high Beta, conveyed mainly Top-Down information from AAC to A1; conversely, high frequencies from low to high Gamma were associated with Bottom-Up flow going from A1 to AAC (Figure 3 in Fontolan et al., 2014). For the first time, we were able to link frequency specific channels to directional information flow in the human auditory hierarchy, partially confirming the proposed scheme of Gamma associated with Bottom-Up and Beta with Top-Down, presented in Sections 4 and 5. Our GC results indicate that a wider low frequency range, Delta-to-Beta, is involved in T-D flow, that would potentially complicate the hypotheses illustrated earlier in this thesis (Engel and Fries, 2010).

The frequency dissociation of GC causal peaks also translated into a number of cross-frequency coupling peaks between the phase of low frequency oscillation from one area and the amplitude of high frequency oscillations from the other area. In particular, we observed two clear PAC patterns in the left hemisphere of both patients (Figure 4 in Fontolan et al., 2014): the phase of low Delta frequencies (<3 Hz) of A1 coupled to the amplitude of Gamma frequencies of AAC, while the phase of Delta-to-Beta (~3-15 Hz) frequencies in AAC was coupled to the amplitude of Gamma

frequencies in A1.

Power spectra, GC and PAC showed that both intrinsic and extrinsic activity differed in the two hemispheres. Both A1 and AAC in the left auditory pathway (historically associated with language) displayed strong modulations in the time-frequency spectrum, at both high and low frequencies. Right A1 exhibited a pattern of activity similar to left A1, whereas right AAC appears less engaged in stimulus-driven activity. The GC plot were also distinct: simple in the right hemisphere, more complex in the left hemisphere, where T-D and B-U peaks alternated in frequency at the lower edge of the spectrum. This might reflect the greater involvement of the left hemisphere in handling speech, as it was confirmed by stimulus/brain correlations (Figure 2 in Fontolan et al., 2014) and previous analyses on the same dataset (Morillon et al., 2012). The partial overlap between GC peaks and PAC clusters ensured that the modulations of Gamma frequencies by the phase of low frequency bands were actually controlled by distant areas. The fact that a number of significant GC peaks did not find any correspondence in PAC clusters might have different explanations: i) it could reflect the absence of cross-frequency coupling at those peaks, ii) it could be due to the lower sensitivity of the circular-to-linear correlation method, or iii) the peaks could be connected through other kinds of coupling mechanisms (e.g. phase-phase).

Intriguingly, we also found that the pattern of causal influence was modulated in time: a fast Fourier transform (FFT), done after subtracting B-U Granger causality from T-D Granger causality, revealed that the causal influence swaps *in time* between T-D and B-U flows, at a rate of about 1-3 Hz and across several frequencies (Figure 5 in Fontolan et al., 2014). Interestingly, the temporal scale at which T-D and B-U causal dominance alternates in time, every ~300-1000 ms, roughly matches with the syllables/words rate in normal speech, potentially suggesting a timescale for the prediction of speech constituents. On the other hand, these slow modulations might be driven by intrinsic slow oscillations such as Delta or Alpha, which have been already linked to sensory perception of external stimuli (Ng et al., 2012; Arnal et al., 2014). A cautionary note should be taken when looking at the modulatory timescales we measured, since GC fluctuations might be rather irregular and thus only partially detectable using a method as simple as FFT.

Overall, the patterns we uncovered are compatible with the multiplexing scheme in the frequency domain: in this respect, ascending information would travel using a high-frequency channel (Gamma) while descending information would be handled by low frequency channels (Delta-to-Beta). According to our results, multiplexing does not appear to be continuous, instead it seems that the two channels dominate during alternating episodes, lasting about 300-1000 ms. The partial discretization of multiplexing in time could be of use in the context of predictive coding, as it has been showed that temporal predictive codes approximately matches the average syllabic duration (Gagnepain et al., 2012). Where and how these two channels are integrated remains to be investigated, although we know that predictive coding models implicate that, at each stage, T-D predictive signals must be formed upon the accumulation of a minimal amount of B-U information.

There are of course a number of limitations in our methodology: first of all, the low number of subjects restricts the statistical power of our study. However, besides the

fact that human intracortical recordings are rare and valued, we employed the most advanced statistical techniques and opted for rather conservative thresholds to ensure the statistical validity of our analyses. A second limitation comes from the linear framework of GC: although we have eliminated many of the flaws on MVAR methods, Granger causality can only grasp linear relationships while other methods, such as DCM, are sensitive to nonlinear features. Nonetheless, we were mainly interested in time-varying GC, and by estimating GC directly from the Wavelet spectrum we reduced the linearity requirement to within the Wavelet transform window. In addition, nonparametric GC is model-free, while DCM is strongly based on choosing a particular architecture *a priori*. GC can be linked to transfer entropy, another interesting measure that is based on information theory, when the variables are Gaussian (Barnett et al., 2009). Transfer entropy is nonlinear, but has some severe practical issues, for instance the estimation of the state-space (Bressler and Seth, 2010). We hence preferred GC in this particular case, our results being also corroborated by recent works in nonhuman primates (Bastos et al., 2014; van Kerkoerle et al., 2014). In these two papers, researchers investigated information passing in the visual cortex of rhesus monkeys using GC, finding a frequency separation between T-D and B-U pathways similar to what we found. The converging evidence clearly supports the hypothesis that frequency-domain multiplexing might be a universal strategy for information passing in sensory hierarchies, at least in primates.

Critics of predictive coding point out that, if the theory were to be true, we should observe much less brain activity than what we usually measure, since the B-U channel would only be activated when an unexpected input is received. Indeed, we do find this reduced activity when the subject is performing specific tasks, revealing effects such as mismatch negativity (MMN) or repetition suppression. Overall, however, the activity of the brain appears to be high enough to be a challenge for predictive coding theory, at least in its original form. As a partial answer, a theoretical implementation proposed in a recent model of visual cortex (Spratling, 2008) demonstrated that the predictive coding framework can also model cognitive functions, like attention, that preferentially enhance, rather than inhibit, cortical activity. Still, the theory of predictive coding requires a high degree of computational complexity to be implemented efficiently: further computational and experimental insights will be needed to definitively link predictive coding to cortical operations.

## 8 Conclusions and future directions

We have seen that an effective combination of experiments and models has so far uncovered some of the powerful advantages of brain rhythms as an information processing tool. Direct examples can be picked from this thesis: nested rhythms in early auditory processing are important for the discretization of speech signal (Article 1 and 2), while distinct frequency bands make up the channels in the multiplexed transmission of information across auditory cortex (Article 3).

This doctoral dissertation aimed at elucidating the role of oscillations in two different, although related, processes: i) we have developed a computational model

simulating the parsing of speech signal and the extraction of its constituents; ii) we have analyzed an exceptional set of human intracortical data to study the frequency channels responsible for information passing across the auditory hierarchy. In both cases we have demonstrated an active role for oscillations, by testing the performance of our nested-rhythms model in classifying speech chunks and by establishing the directional influences and the nonlinear couplings between A1 and AAC as a function of recorded frequency.

This work is part of a larger effort, involving several researchers, to get brain rhythms beyond simple correlations and prove their crucial role in brain functions. Many experimental works are providing convergent evidence that oscillations characterize many high- and low-level processes in the brain. Nevertheless, it remains very hard to univocally demonstrate that oscillations are *causing* a given perception or behavior, at least until it will become possible to falsify this theory in the sense of Popper. Recently, the optogenetics breakthrough has revolutionized the brain imaging techniques in mice, allowing for the online, selective control of neuronal firing. The ability to activate or inactivate a particular class of neurons is of capital importance, in order to test the correspondence between behavior and brain waves. Indeed, if we could, for example, reset the phase of a PING rhythm by simultaneously stimulating several fast-spiking interneurons: we could then examine the animal behavior and prove (or disprove) the causal role of oscillations. Optogenetics tools recently started to be employed to manipulate brain rhythms in rodents, in order to investigate Theta-Gamma coupling in the hippocampus (Vandecasteele et al., 2014). Researchers can now begin to answer the question of whether or not oscillations in the rodent neocortex have the same significance as in the primate neocortex.

Complementarily to experimental advances, I believe that neuroscience needs theoretical models, especially in order to be able to answer the question of the functional significance of oscillations. Even in the absence of compelling experimental proofs, it is very important, in my opinion, to discuss and model the computational contribution of oscillations to the coding scheme used in the brain. The main goal would be to explore the advantages of periodic over asynchronous neuronal firing, since it is well known that high correlations reduce the capacity of the network to process information but, at the same time, bringing the spikes together in time can increase the redundancy of the representation and, consequently, the noise robustness of the network. Future theoretical studies should address some of the core problems that, in my opinion, are far from being solved. First of all, whether is it computationally worth to spend several cortical neurons to produce a highly correlated, and thus less informative, spike pattern that gives rise to low frequency oscillations. To this end, it would be of great interest to examine, experimentally, if neurons that participate in generating cortical rhythms are selective to any of the stimulus features. A further interesting study should focus on the relationship between three classes of models that, until now, have not been connected: Bayesian models of probabilistic inference, mechanistic models of brain oscillations and models of predictive coding. Bringing together these three frameworks would constitute a critical advancement in understanding the basic principles of neural computation.

## 9 Appendices

### A. Circular-to-linear correlation

Circular-to-linear correlation is a simple measure to assess correlations between a phase (circular) variable  $\alpha$  and an amplitude (linear) variable  $x$  (Berens, 2009). The trick to compute circular-to-linear correlations consists in calculating the correlations between  $x$  and each of the two projections of the angular variable ( $\sin \alpha, \cos \alpha$ ) separately. From the Pearson correlation coefficient, computed for two linear variables  $x$  and  $y$ :

$$c(x, y) = \frac{\text{cov}(x, y)}{\text{var}(x) \text{var}(y)}, \quad (\text{A.1})$$

we define

$$r_{cx} = c(\cos \alpha, y), r_{sx} = c(\sin \alpha, y), r_{cs} = c(\sin \alpha, \cos \alpha). \quad (\text{A.2})$$

The formula for circular-to-linear coefficient simply reads:

$$\rho_{cl} = \sqrt{\frac{r_{cx}^2 + r_{sx}^2 - 2r_{cx}r_{sx}r_{cs}}{1 - r_{cs}^2}}. \quad (\text{A.3})$$

### B. Single neuron models

#### B.1. Integrate-and-fire neuron and the canonical model

Consider the simple circuit depicted in

Figure 4: the driving current  $I(t)$  passes through a resistor and a capacitance disposed in parallel, as in the following equation:

$$I(t) = \frac{V(t)}{R} + C \frac{dV(t)}{dt}. \quad (\text{B.4})$$

Introducing the membrane time constant  $\tau_m = RC$ , the previous equation might be rewritten as

$$\tau_m \frac{dV(t)}{dt} = -V(t) + R I(t). \quad (\text{B.5})$$

As soon as the voltage reaches its threshold value  $V_{th}$ , the neuron produces a spike and the potential is reset to zero. Although mathematically easy-to-handle, the linear version of the integrate-and-fire neuron (LIF) is overly unrealistic even to simulate spike generation, thus Ermentrout and Kopell introduced a variant (Ermentrout and Kopell, 1986) by adding a voltage-dependent nonlinear (quadratic or exponential) term that better approximates the fast activation of sodium channels:

$$\tau_m \frac{dV(t)}{dt} = f(V) + R I(t), \quad (\text{B.6})$$

Where  $f(V)$  is any nonlinear function of  $V$ . The quadratic ( $f(V) = V^2$ ) instance of the model (quadratic integrate-and-fire neuron, QIF) has proven very useful in network simulations and modeling studies, being both relatively realistic and mathematically tractable. Although it does not capture well the subthreshold dynamics, the QIF does a

much better job than the LIF in simulating the production of a spike. Also, it does not require any ad hoc threshold since for any positive value of  $I$ , the voltage goes to infinity in a finite time. Hence the reset condition becomes

$$V(t^* - 0) = \infty, V(t^* + 0) = -\infty. \quad (\text{B.7})$$

The QIF can be related to a canonical model type I model called *theta neuron* (Ermentrout and Kopell, 1986) via a simple change of variables  $V = \tan \frac{\theta}{2}$ , thereby mapping the membrane voltage space onto a one dimensional ring through an isomorphism. The relative location on the ring is described by the angle  $\theta \in (-\pi, \pi]$ , which represents the phase of the oscillator; by convention a spike is produced when  $\theta = \pi$ . Equation B.6 becomes then:

$$\frac{d\theta}{dt} = \frac{1}{\tau_m} (1 - \cos \theta) + I (1 + \cos \theta), \quad (\text{B.8})$$

with the reset condition

$$\theta(t^* - 0) = \pi, \theta(t^* + 0) = -\pi. \quad (\text{B.9})$$

The period can be easily computed by integrating Equation B.8 between  $-\pi$  and  $\pi$ , obtaining

$$T = \int_{\theta_1}^{\theta_2} \frac{d\theta}{\frac{1}{\tau_m} (1 - \cos \theta) + I (1 + \cos \theta)} = \sqrt{\frac{\pi}{I}} \left[ \tan^{-1} \frac{\tan(\theta/2)}{\sqrt{\tau I}} \right]_{-\pi}^{\pi} = \pi \sqrt{\frac{\pi}{I}}, \quad (\text{B.10})$$

from which we see that

$$\lim_{I \rightarrow 0^+} T(I) = \infty, \quad (\text{B.11})$$

i.e. the oscillations frequency will tend to zero.

## B.2. The effect of inhibition and the birth of a bifurcation

In the PING model a QIF neuron (or, equivalently, a theta model) receives input from an inhibitory cell that has been previously excited by QIF firing. The effect of inhibitory synapses can be incorporated into Equation B.8:

$$\frac{d\theta}{dt} = \frac{1}{\tau_m} (1 - \cos \theta) + (I - g_I s_I) (1 + \cos \theta), \quad (\text{B.12})$$

where  $g_I$  is the synaptic strength or conductance and  $s_I$  is any gating function, as for example instantaneous rise upon one excitatory spike, followed by exponential decay:

$$\frac{ds_I}{dt} = -\frac{s_I}{\tau_I} + \delta(t - t^*). \quad (\text{B.13})$$

The new negative term in Equation B.12 introduces a dramatic change in the phase space of the system, by creating a *bifurcation* point, i.e. a topological change in the dynamics (Shlizerman and Holmes, 2012). In fact, for  $g_I s_I > I$  this differential equation has two equilibrium points on the invariant circle for  $\theta_{\pm} = \cos^{-1} \left( \frac{1+I-g_I s_I}{1-I+g_I s_I} \right)$ , one stable and one unstable. As inhibition fades,  $g_I s_I$  approaches  $I$  and the two fixed points come close to each other, until they merge and disappear for  $g_I s_I = I$ , i.e. the bifurcation point. For  $g_I s_I < I$  the total input to the QIF is positive, there are no fixed points, and the neuron starts firing regularly although its dynamics is slowed down in the vicinity of the ghost of the merged equilibrium points ( $\theta = 0$ ). The theta model is an example of Type I, whose oscillatory dynamics, thanks to the particular bifurcation it bears, can be made arbitrarily slow (see Appendix B.1); this remains true in presence of inhibition as

long as  $g_I s_I < I$ .

### B.3. Phase response curves

Phase response curves have been introduced to quantify the advancements or regressions caused to a dynamical system by a perturbation of infinitesimal amplitude, arriving at time  $t$ . The PRC measures the difference between the standard period  $T$  and the time to next spike  $T'$ , as a function of the time or phase at which the perturbation has been applied:

$$PRC(t) = \frac{T - T'(t)}{T}. \quad (\text{B.14})$$

Similarly, PRCs can be computed as a function of phase  $\varphi = \frac{2\pi(t-t_0)}{T}$ , where  $t_0$  is the time at which the last spike has been emitted:

$$PRC(\varphi) = \varphi' - \varphi. \quad (\text{B.15})$$

an expression that is particularly useful for phase models like the theta neuron. Also, the PRC formalism remains effective even in presence of irregular oscillations, i.e. when the time difference between two spikes is subjected to a certain degree of randomness (Gutkin et al., 2005).

## 10 Bibliography

- Abbott L, van Vreeswijk C (1993) Asynchronous states in networks of pulse-coupled oscillators. *Phys Rev E* 48:1483–1490.
- Ahissar E, Nagarajan SS, Ahissar M, Protopapas A, Mahncke H, Merzenich MM (2001) Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc Natl Acad Sci U S A* 98:13367–13372.
- Akam T, Oren I, Mantoan L, Ferenczi E, Kullmann DM (2012) Oscillatory dynamics in the hippocampus support dentate gyrus–CA3 coupling. *Nat Neurosci* 15:763–768.
- Akam TE, Kullmann DM (2010) Oscillations and filtering networks support flexible routing of information. *Neuron* 67:308–320.
- Akam TE, Kullmann DM (2014) Oscillatory multiplexing of population codes for selective communication in the mammalian brain. *Nat Rev Neurosci* 15:111–122.
- Amari S-I (1972) Characteristics of Random Nets of Analog Neuron-Like Elements. *IEEE Trans Syst Man Cybern* 2:643–657.
- Arnal LH, Doelling KB, Poeppel D (2014) Delta–Beta Coupled Oscillations Underlie Temporal Prediction Accuracy. *Cereb Cortex* .
- Arnal LH, Wyart V, Giraud A-L (2011) Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nat Neurosci* 14:797–801.
- Arnal LHL, Giraud A-L AL (2012) Cortical oscillations and sensory predictions. *Trends Cogn Sci* 16:390–398.
- Atencio C a, Schreiner CE (2010) Columnar connectivity and laminar processing in cat primary auditory cortex. *PLoS One* 5:e9521.
- Axmacher N, Mormann F, Fernández G, Elger CE, Fell J (2006) Memory formation by neuronal synchronization. *Brain Res Rev* 52:170–182.
- Azevedo FAC, Carvalho LRB, Grinberg LT, Farfel JM, Ferretti REL, Leite REP, Jacob Filho W, Lent R, Herculano-Houzel S (2009) Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J Comp Neurol* 513:532–541.

- Baillie RT (1996) Long memory processes and fractional integration in econometrics. *J Econom* 73:5–59.
- Bak P, Paczuski M (1995) Complexity, contingency, and criticality. *Proc Natl Acad Sci U S A* 92:6689–6696.
- Bar M (2007) The proactive brain: using analogies and associations to generate predictions. *Trends Cogn Sci* 11:280–289.
- Barnett L, Barrett AB, Seth AK (2009) Granger causality and transfer entropy Are equivalent for gaussian variables. *Phys Rev Lett* 103:238701.
- Barnett L, Seth AK (2011) Behaviour of Granger causality under filtering: Theoretical invariance and practical application. *J Neurosci Methods* 201:404–419.
- Bastos AM, Vezoli J, Bosman CA, Schoffelen J-M, Oostenveld R, Dowdall JR, De Weerd P, Kennedy H, Fries P (2014) Visual Areas Exert Feedforward and Feedback Influences through Distinct Frequency Channels. *Neuron* 85:390–401.
- Bastos AMM, Usrey WMM, Adams RA a, Mangun GRR, Fries P, Friston KJJ (2012) Canonical microcircuits for predictive coding. *Neuron* 76:695–711.
- Bédard C, Destexhe A (2009) Macroscopic models of local field potentials and the apparent 1/f noise in brain activity. *Biophys J* 96:2589–2603.
- Bédard C, Kröger H, Destexhe a. (2006) Does the 1/f Frequency Scaling of Brain Signals Reflect Self-Organized Critical States? *Phys Rev Lett* 97:118102.
- Bédard C, Rodrigues S, Roy N, Contreras D, Destexhe A (2010) Evidence for frequency-dependent extracellular impedance from the transfer function between extracellular and intracellular potentials: intracellular-LFP transfer function. *J Comput Neurosci* 29:389–403.
- Belluscio M, Mizuseki K, Schmidt R, Kempter R, Buzsaki G (2012) Cross-Frequency Phase-Phase Coupling between Theta and Gamma Oscillations in the Hippocampus. *J Neurosci* 32:423–435.
- Berens P (2009) CircStat: A MATLAB Toolbox for Circular Statistics. *J Stat Softw* 31:1–21.
- Berger H (1929) Über das Elektrenkephalogramm des Menschen. *Arch Psychiatr Nervenkr* 87:527–570.
- Binzegger T, Douglas RJ, Martin K a C (2009) Topology and dynamics of the canonical circuit of cat V1. *Neural Netw* 22:1071–1078.
- Börgers C, Kopell NJ (2005) Effects of noisy drive on rhythms in networks of excitatory and inhibitory neurons. *Neural Comput* 17:557–608.
- Börgers C, Kopell NJ (2008) Gamma oscillations and stimulus selection. *Neural Comput* 20:383–414.
- Börgers C, Walker B (2013) Toggling between gamma-frequency activity and suppression of cell assemblies. *Front Comput Neurosci* 7:33.
- Bosman CA, Schoffelen J-M, Brunet N, Oostenveld R, Bastos AM, Womelsdorf T, Rubehn B, Stieglitz T, De Weerd P, Fries P (2012) Attentional Stimulus Selection through Selective Synchronization between Monkey Visual Areas. *Neuron* 75:875–888.
- Boyden ES, Zhang F, Bamberg E, Nagel G, Deisseroth K (2005) Millisecond-timescale, genetically targeted optical control of neural activity. *Nat Neurosci* 8:1263–1268.
- Bragin A, Jandó G, Nádasdy Z, Hetke J, Wise K, Buzsáki G (1995) Gamma (40-100 Hz) oscillation in the hippocampus of the behaving rat. *J Neurosci* 15:47–60.
- Brasselet R, Panzeri S, Logothetis NK, Kayser C (2012) Neurons with stereotyped and rapid responses provide a reference frame for relative temporal coding in primate auditory cortex. *J Neurosci* 32:2998–3008.

- Bressler SL, Seth AK (2010) Wiener-Granger Causality: A well established methodology. *Neuroimage* 58:323–329.
- Bressloff PC, Coombes S (2000) Dynamics of strongly-coupled spiking neurons. *Neural Comput* 12:91–129.
- Brovelli A (2012) Statistical analysis of single-trial Granger causality spectra. *Comput Math Methods Med* 2012:697610.
- Brovelli A, Ding M, Ledberg A, Chen Y, Nakamura R, Bressler SL (2004) Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality. *Proc Natl Acad Sci U S A* 101:9849–9854.
- Brown E, Moehlis J, Holmes P (2004) On the phase reduction and response dynamics of neural oscillator populations. *Neural Comput* 16:673–715.
- Brunel N (2000) Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J Comput Neurosci* 8:183–208.
- Brunel N, Hakim V (1999) Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comput* 11:1621–1671.
- Brunel N, Wang X-J (2003) What determines the frequency of fast network oscillations with irregular neural discharges? I. Synaptic dynamics and excitation-inhibition balance. *J Neurophysiol* 90:415–430.
- Buffalo E a, Fries P, Landman R, Buschman TJ, Desimone R (2011) Laminar differences in gamma and alpha coherence in the ventral stream. *Proc Natl Acad Sci U S A* 108:1–6.
- Bukalo O, Campanac E, Hoffman DA, Fields RD (2013) Synaptic plasticity by antidromic firing during hippocampal network oscillations. *Proc Natl Acad Sci U S A* 110:5175–5180.
- Burkitt AN (2006) A review on the integrate-and-fire neuron model: II. Inhomogeneous synaptic input and network properties. *Biol Cybern* 95:97–112.
- Buschman TJ, Miller EK (2007) Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* 315:1860–1862.
- Buzsáki G, Anastassiou C a, Koch C (2012) The origin of extracellular fields and currents--EEG, ECoG, LFP and spikes. *Nat Rev Neurosci* 13:407–420.
- Buzsáki G, Draguhn A (2004) Neuronal oscillations in cortical networks. *Science* 304:1926–1929.
- Buzsáki G, Wang X-J (2012) Mechanisms of gamma oscillations. *Annu Rev Neurosci* 35:203–225.
- Cabral HO, Vinck M, Fouquet C, Pennartz CMA, Rondi-Reig L, Battaglia FP (2014) Oscillatory dynamics and place field maps reflect hippocampal ensemble processing of sequence and place memory under NMDA receptor control. *Neuron* 81:402–415.
- Cannon J, McCarthy MM, Lee S, Lee J, Börgers C, Whittington MA, Kopell N (2014) Neurosystems: Brain rhythms and cognitive processing. *Eur J Neurosci* 39:705–719.
- Canolty R, Edwards E, Dalal S, Soltani M, SS (2006) High gamma power is phase-locked to theta oscillations in human neocortex. *Science* (80- ).
- Canolty RT, Knight RT (2010) The functional role of cross-frequency coupling. *Trends Cogn Sci*.
- Cardin J a, Carlén M, Meletis K, Knoblich U, Zhang F, Deisseroth K, Tsai L-H, Moore CI (2009) Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature* 459:663–667.
- Carlo CN, Stevens CF (2013) Structural uniformity of neocortex, revisited. *Proc Natl Acad Sci U S A* 110:1488–1493.

- Carracedo LM, Kjeldsen H, Cunningham L, Jenkins A, Schofield I, Cunningham MO, Davies CH, Traub RD, Whittington M a (2013) A neocortical delta rhythm facilitates reciprocal interlaminar interactions via nested theta rhythms. *J Neurosci* 33:10750–10761.
- Chen CC, Henson RN, Stephan KE, Kilner JM, Friston KJ (2009) Forward and backward connections in the brain: a DCM study of functional asymmetries. *Neuroimage* 45:453–462.
- Chen Y, Bressler SL, Ding M (2006) Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data. *J Neurosci Methods* 150:228–237.
- Chi T, Ru P, Shamma S a (2005) Multiresolution spectrotemporal analysis of complex sounds. *J Acoust Soc Am* 118:887–906.
- Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci* 36:181–204.
- Code RA, Winer JA (1985) Commissural neurons in layer III of cat primary auditory cortex (AI): pyramidal and non-pyramidal cell input. *J Comp Neurol* 242:485–510.
- Costa-Faidella J, Baldeweg T, Grimm S, Escera C (2011) Interactions between “What” and “When” in the Auditory System: Temporal Predictability Enhances Repetition Suppression. *J Neurosci* 31:18590–18597.
- Dale N, Kandel ER (1993) L-glutamate may be the fast excitatory transmitter of Aplysia sensory neurons. *Proc Natl Acad Sci U S A* 90:7163–7167.
- Dayan P, Hinton GE, Neal RM, Zemel RS (1995) The Helmholtz machine. *Neural Comput* 7:889–904.
- De Callataÿ AM (1992) Natural and artificial intelligence: misconceptions about brains and neural networks.
- Deco G, Jirsa VK (2012) Ongoing cortical activity at rest: criticality, multistability, and ghost attractors. *J Neurosci* 32:3366–3375.
- Detto M, Molini A, Katul G, Stoy P, Palmroth S, Baldocchi D (2012) Causality and persistence in ecological systems: a nonparametric spectral granger causality approach. *Am Nat* 179:524–535.
- Dhamala M, Rangarajan G, Ding M (2008) Estimating Granger Causality from Fourier and Wavelet Transforms of Time Series Data. *Phys Rev Lett* 100:018701.
- Diesmann M, Gewaltig MO, Aertsen A (1999) Stable propagation of synchronous spiking in cortical neural networks. *Nature* 402:529–533.
- Ding M, Chen Y, Bressler SL (2008) Granger Causality : Basic Theory and Application to Neuroscience arXiv : q-bio / 0608035v1 [ q-bio . QM ] 23 Aug 2006.
- Ding N, Simon JZ (2012) Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc Natl Acad Sci U S A* 109:11854–11859.
- Ding N, Simon JZ (2014) Cortical entrainment to continuous speech: functional roles and interpretations. *Front Hum Neurosci* 8:311.
- Doelling KB, Arnal LH, Ghitta O, Poeppel D (2014) Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* 85 Pt 2:761–768.
- Douglas RJ, Martin K a C (2004) Neuronal circuits of the neocortex. *Annu Rev Neurosci* 27:419–451.
- Dutta P, Horn PM (1981) Low-frequency fluctuations in solids: 1/f noise. *Rev Mod Phys* 53:497–516.
- Elliott TM, Theunissen FE (2009) The modulation transfer function for speech intelligibility. *PLoS Comput Biol* 5:e1000302.
- Engel a K, Fries P, Singer W (2001) Dynamic predictions: oscillations and synchrony in top-down processing. *Nat Rev Neurosci* 2:704–716.

- Engel AK, Fries P (2010) Beta-band oscillations--signalling the status quo? *Curr Opin Neurobiol* 20:156–165.
- Ermentrout B (1995) Type I Membranes , Phase Resetting Curves , and Synchrony 1 Introduction. :1–23.
- Ermentrout B (1996) Type I membranes, phase resetting curves, and synchrony. *Neural Comput* 8:979–1001.
- Ermentrout G, Kopell N (1986) Parabolic bursting in an excitable system coupled with a slow oscillation. *SIAM J Appl Math* 46:233–253.
- Eyherabide HG, Samengo I (2013) When and why noise correlations are important in neural decoding. *J Neurosci* 33:17921–17936.
- Feldman JA (1985) Connectionist models and parallelism in high level vision. *Comput Vision, Graph Image Process* 31:178–200.
- Fell J, Klaver P, Lehnertz K, Grunwald T, Schaller C, Elger CE, Fernández G (2001) Human memory formation is accompanied by rhinal-hippocampal coupling and decoupling. *Nat Neurosci* 4:1259–1264.
- Ferster D, Spruston N (1995) Cracking the neuronal code. *Science* 270:756–757.
- Fitzhugh R (1961) Impulses and Physiological States in Theoretical Models of Nerve Membrane. *Biophys J* 1:445–466.
- Fontolan L, Krupa M, Hyafil A, Gutkin B (2013) Analytical insights on theta-gamma coupled neural oscillators. *J Math Neurosci* 3:16.
- Fontolan L, Morillon B, Liegeois-Chauvel C, Giraud A-L (2014) The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. *Nat Commun* 5:4694.
- Freeman WJ, Holmes MD, Burke BC, Vanhatalo S (2003) Spatial spectra of scalp EEG and EMG from awake humans. *Clin Neurophysiol* 114:1053–1068.
- Frégnac Y, Bringuer V, Baranyi A (1994) Oscillatory Neuronal Activity in Visual Cortex: A Critical Re-Evaluation. In: *Temporal Coding in the Brain SE - 5* (Buzsáki G, Llinás R, Singer W, Berthoz A, Christen Y, eds), pp 81–102 *Research and Perspectives in Neurosciences*. Springer Berlin Heidelberg.
- Fries P (2005) A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn Sci* 9:474–480.
- Fries P, Nikolić D, Singer W (2007) The gamma cycle. *Trends Neurosci* 30:309–316.
- Fries P, Schröder J-H, Roelfsema PR, Singer W, Engel AK (2002) Oscillatory neuronal synchronization in primary visual cortex as a correlate of stimulus selection. *J Neurosci* 22:3739–3754.
- Friston K (2002) Functional integration and inference in the brain. *Prog Neurobiol* 68:113–143.
- Friston K (2003) Learning and inference in the brain. *Neural Netw* 16:1325–1352.
- Friston K (2005) A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* 360:815–836.
- Friston K (2008) Hierarchical models in the brain. *PLoS Comput Biol* 4.
- Friston K (2010) The free-energy principle: a unified brain theory? *Nat Rev Neurosci* 11:127–138.
- Friston K, Moran R, Seth AK (2013) Analysing connectivity with Granger causality and dynamic causal modelling. *Curr Opin Neurobiol* 23:172–178.
- Gagnepain P, Henson RN, Davis MH (2012) Temporal predictive codes for spoken words in auditory cortex. *Curr Biol* 22:615–621.
- Gerstein GL, Mandelbrot B (1964) RANDOM WALK MODELS FOR THE SPIKE ACTIVITY OF A SINGLE NEURON. *Biophys J* 4:41–68.

- Gevers MR, Anderson BDO (1981) Representations of jointly stationary stochastic feedback processes<sup>†</sup>. *Int J Control* 33:777–809.
- Geweke J (1982) Measurement of Linear Dependence and Feedback between Multiple Time Series. *J Am Stat Assoc* 77:304–313.
- Ghitza O (2011) Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front Psychol* 2:1–13.
- Ghitza O, Greenberg S (2009) On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica* 66:113–126.
- Giraud A-L, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci* 15:511–517.
- Goupillaud P, Grossmann A, Morlet J (1984) Cycle-octave and related transforms in seismic signal analysis. *Geoexploration* 23:85–102.
- Granger CWJ (1969) Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica* 37:424–438.
- Greenberg S, Ainsworth WA (2006) Listening to speech: an auditory perspective. Lawrence Erlbaum.
- 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:e1001752.
- Gütig R, Sompolinsky H (2009) Time-warp-invariant neuronal processing. *PLoS Biol* 7:e1000141.
- Gutkin BS, Ermentrout GB, Reyes AD (2005) Phase-response curves give the responses of neurons to transient inputs. *J Neurophysiol* 94:1623–1635.
- Hackett TA (2011) Information flow in the auditory cortical network. *Hear Res* 271:133–146.
- Haken H, Sauermann H (1963) Frequency shifts of laser modes in solid state and gaseous systems. *Zeitschrift fur Phys* 176:47–62.
- Hansel D, Mato G, Meunier C (1995) Synchrony in excitatory neural networks. *Neural Comput* 7:307–337.
- Hawkins RD, Kandel ER, Siegelbaum SA (1993) Learning to Modulate Transmitter Release: Themes and Variations in Synaptic Plasticity. *Annu Rev Neurosci* 16:625–665.
- Henry MJ, Herrmann B, Obleser J (2014) Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proc Natl Acad Sci U S A* 111:14935–14940.
- Herculano-Houzel S, Collins CE, Wong P, Kaas JH, Lent R (2008) The basic nonuniformity of the cerebral cortex. *Proc Natl Acad Sci U S A* 105:12593–12598.
- Hidalgo J, Grilli J, Suweis S (2013) Emergence of criticality in living systems through adaptation and evolution: Practice Makes Critical. *arXiv Prepr arXiv* ....
- Hidalgo J, Grilli J, Suweis S, Muñoz MA, Banavar JR, Maritan A (2014) Information-based fitness and the emergence of criticality in living systems. *Proc Natl Acad Sci U S A* 111:10095–10100.
- Hill A V. (1933) WAVE TRANSMISSION AS THE BASIS OF NERVE ACTIVITY. *Cold Spring Harb Symp Quant Biol* 1:146–151.
- Hodgkin AL, Huxley AF (1990) A quantitative description of membrane current and its application to conduction and excitation in nerve. *Bull Math Biol* 52:25–71.
- Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities *Biophysics : Hopfield I T ., V . 79:2554–2558.*

- Hopfield JJ (1995) Pattern recognition computation using action potential timing for stimulus representation. *Nature* 376:33–36.
- Hopfield JJ (2004) Encoding for computation: recognizing brief dynamical patterns by exploiting effects of weak rhythms on action-potential timing. *Proc Natl Acad Sci U S A* 101:6255–6260.
- Hosoya T, Baccus S a, Meister M (2005) Dynamic predictive coding by the retina. *Nature* 436:71–77.
- Howard MF, Poeppel D (2012) The neuromagnetic response to spoken sentences: co-modulation of theta band amplitude and phase. *Neuroimage* 60:2118–2127.
- Huerta PT, Lisman JE (1995) Bidirectional synaptic plasticity induced by a single burst during cholinergic theta oscillation in CA1 in vitro. *Neuron* 15:1053–1063.
- Hunter JD, Milton JG (2003) Amplitude and frequency dependence of spike timing: implications for dynamic regulation. *J Neurophysiol* 90:387–394.
- Hutcheon B, Yarom Y (2000) Resonance, oscillation and the intrinsic frequency preferences of neurons. *Trends Neurosci* 23:216–222.
- Hyafil A, Fontolan L, Kabdebon C, Gutkin B, Giraud A-L (2015) Speech encoding by coupled cortical theta and gamma oscillations. *eLife* (accepted Publ Minor Revis).
- Hyafil A, Giraud A-L, Fontolan L, Gutkin B (n.d.) Neural cross-frequency coupling: tying mechanisms to functions.
- Ijiri Y, Simon HA (1977) Skew distributions and the sizes of business firms (Studies in Mathematical and Managerial Economics). Elsevier Science.
- Isaacson JS, Scanziani M (2011) How inhibition shapes cortical activity. *Neuron* 72:231–243.
- Izhikevich EM (2004) Which model to use for cortical spiking neurons? *IEEE Trans Neural Networks* 15:1063–1070.
- J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus DSP and NLD (1993) TIMIT Acoustic Phonetic Continuous Speech Corpus.
- Jensen O, Colgin LL (2007) Cross-frequency coupling between neuronal oscillations. *Trends Cogn Sci* 11:267–269.
- Jensen O, Gips B, Bergmann TO, Bonnefond M (2014) Temporal coding organized by coupled alpha and gamma oscillations prioritize visual processing. *Trends Neurosci* 37:357–369.
- Jensen O, Lisman JEE (2005) Hippocampal sequence-encoding driven by a cortical multi-item working memory buffer. *Trends Neurosci* 28:67–72.
- Jirsa VK (2009) Neural field dynamics with local and global connectivity and time delay. *Philos Trans A Math Phys Eng Sci* 367:1131–1143.
- Jutras MJ, Fries P, Buffalo EA (2013) Oscillatory activity in the monkey hippocampus during visual exploration and memory formation. *Proc Natl Acad Sci U S A* 110:13144–13149.
- Kaas JH, Hackett TA, Tramo MJ (1999) Auditory processing in primate cerebral cortex. *Curr Opin Neurobiol* 9:164–170.
- Kahana MJ, Seelig D, Madsen JR (2001) Theta returns. *Curr Opin Neurobiol* 11:739–744.
- Kahana MJ, Sekuler R, Caplan JB, Kirschen M, Madsen JR (1999) Human theta oscillations exhibit task dependence during virtual maze navigation. *Nature* 399:781–784.
- Kanold PO, Nelken I, Polley DB (2014) Local versus global scales of organization in auditory cortex. *Trends Neurosci* 37:502–510.

- Kasdin NJ (1995) Discrete simulation of colored noise and stochastic processes and  $1/f[\text{up}]\alpha[\text{/up}]$  power law noise generation. Proc IEEE 83:802–827.
- Kayser C, Ince RAA, Panzeri S (2012) Analysis of Slow (Theta) Oscillations as a Potential Temporal Reference Frame for Information Coding in Sensory Cortices Behrens T, ed. PLoS Comput Biol 8:e1002717.
- Kayser C, Montemurro M a, Logothetis NK, Panzeri S (2009) Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. Neuron 61:597–608.
- Kello CT, Brown GD a, Ferrer-I-Cancho R, Holden JG, Linkenkaer-Hansen K, Rhodes T, Van Orden GC (2010) Scaling laws in cognitive sciences. Trends Cogn Sci 14:223–232.
- Kiebel SJ, Daunizeau J, Friston KJ (2008) A hierarchy of time-scales and the brain. PLoS Comput Biol 4:e1000209.
- Kilpatrick ZP, Ermentrout B (2011) Sparse gamma rhythms arising through clustering in adapting neuronal networks. PLoS Comput Biol 7:e1002281.
- King AJ, Walker KMM (2012) Integrating information from different senses in the auditory cortex. Biol Cybern 106:617–625.
- Koch C (1993) Good vibes. Proc Natl Acad Sci U S A 90:1637–1638.
- Koch C (1999) Complexity and the Nervous System. Science (80- ) 284:96–98.
- Kopell NJ, Börgers C, Pervouchine DD, Malerba P (2010) Gamma and theta rhythms in biophysical models of hippocampal circuits. In: Hippocampal ... (Cutsuridis V, ed), pp 423–457. Springer.
- Kruse K, Jülicher F (2005) Oscillations in cell biology. Curr Opin Cell Biol 17:20–26.
- Kubanek J, Brunner P, Gunduz A, Poeppel D, Schalk G (2013) The tracking of speech envelope in the human cortex. PLoS One 8:e53398.
- Kuramoto Y (1984) Chemical oscillations, waves, and turbulence. Berlin; New York: Springer.
- Lachaux J-P, Axmacher N, Mormann F, Halgren E, Crone NE (2012) High-frequency neural activity and human cognition: Past, present and possible future of intracranial EEG research. Prog Neurobiol 98:279–301.
- Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. Science (80- ) 320:110–113.
- Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE (2005) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. J Neurophysiol 94:1904–1911.
- Ledoux E, Brunel N (2011) Dynamics of networks of excitatory and inhibitory neurons in response to time-dependent inputs. Front Comput Neurosci 5:25.
- Lee H, Simpson G V, Logothetis NK, Rainer G (2005) Phase locking of single neuron activity to theta oscillations during working memory in monkey extrastriate visual cortex. Neuron 45:147–156.
- Levina a., Herrmann JM, Geisel T (2007) Dynamical synapses causing self-organized criticality in neural networks. Nat Phys 3:857–860.
- Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M (1967) Perception of the speech code. Psychol Rev 74:431–461.
- Liberman AM, Mattingly IG (1985) The motor theory of speech perception revised. Cognition 21:1–36.
- Lisman JE (2005) The theta/gamma discrete phase code occurring during the hippocampal phase precession may be a more general brain coding scheme. Hippocampus 15:913–922.

- Lisman JE, Buzsáki G (2008) A neural coding scheme formed by the combined function of gamma and theta oscillations. *Schizophr Bull* 34:974–980.
- Lisman JE, Idiart M (1995) Storage of  $7 \pm 2$  short-term memories in oscillatory subcycles. *Science* (80- ) 267:1512–1515.
- Lisman JE, Jensen O (2013) The Theta-Gamma Neural Code. *Neuron* 77:1002–1016.
- Llinás R, Yarom Y (1986) Oscillatory properties of guinea-pig inferior olivary neurones and their pharmacological modulation: an in vitro study. *J Physiol* 376:163–182.
- Llinás RR (1988) The intrinsic electrophysiological properties of mammalian neurons: insights into central nervous system function. *Science* 242:1654–1664.
- Llinás RR (2014) Intrinsic electrical properties of mammalian neurons and CNS function: a historical perspective. *Front Cell Neurosci* 8:320.
- Louie K, Wilson MA (2001) Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron* 29:145–156.
- 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* 3:170.
- Mancilla JG, Lewis TJ, Pinto DJ, Rinzel J, Connors BW (2007) Synchronization of electrically coupled pairs of inhibitory interneurons in neocortex. *J Neurosci* 27.
- Markov NT, Kennedy H (2013) The importance of being hierarchical. *Curr Opin Neurobiol* 23:187–194.
- Marner L, Nyengaard JR, Tang Y, Pakkenberg B (2003) Marked loss of myelinated nerve fibers in the human brain with age. *J Comp Neurol* 462:144–152.
- McClelland JL, Elman JL (1986) The TRACE model of speech perception. *Cogn Psychol* 18:1–86.
- Meddis R (2010) Computational Models of the Auditory System. Springer.
- Mehta M, Lee A (2002) Role of experience and oscillations in transforming a rate code into a temporal code. *Nature*:8–11.
- Mesgarani N, Cheung C, Johnson K, Chang EF (2014) Phonetic Feature Encoding in Human Superior Temporal Gyrus. *Science* 1006.
- Milotti E (2002) 1/f noise: a pedagogical review. *arXiv Prepr physics/0204033*.
- Mirollo RE, Strogatz SH (1990) Synchronization of Pulse-Coupled Biological Oscillators. *SIAM J Appl Math* 50:1645.
- Mongillo G, Amit D (2001) Oscillations and irregular emission in networks of linear spiking neurons. *J Comput Neurosci* 11:249–261.
- Morillon B, Liégeois-Chauvel C, Arnal LH, Bénar C-G, Giraud A-L (2012) Asymmetric Function of Theta and Gamma Activity in Syllable Processing: An Intra-Cortical Study. *Front Psychol* 3:1–9.
- Mountcastle VB (1997) The columnar organization of the neocortex. *Brain* 120 ( Pt 4):701–722.
- Mumford D (1992) On the computational architecture of neocortex. *Biol Cybern* 251:241–251.
- Murphy GG (1997) Mediation of Classical Conditioning in *Aplysia californica* by Long-Term Potentiation of Sensorimotor Synapses. *Science* (80- ) 278:467–471.

- Musha T, Yamamoto M (1997) 1/f fluctuations in biological systems. In: Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. "Magnificent Milestones and Emerging Opportunities in Medical Engineering" (Cat. No.97CH36136), pp 2692–2697. IEEE.
- Nagumo J, Arimoto S, Yoshizawa S (1962) An Active Pulse Transmission Line Simulating Nerve Axon. Proc IRE 50:2061–2070.
- Ng BSW, Schroeder T, Kayser C (2012) A precluding but not ensuring role of entrained low-frequency oscillations for auditory perception. J Neurosci 32:12268–12276.
- Norris D (1994) Shortlist: a connectionist model of continuous speech recognition. Cognition 52:189–234.
- Nourski K V, Reale R a, Oya H, Kawasaki H, Kovach CK, Chen H, Howard M a, Brugge JF (2009) Temporal envelope of time-compressed speech represented in the human auditory cortex. J Neurosci 29:15564–15574.
- Oprisan SA, Canavier CC (2006) Technique for eliminating nonessential components in the refinement of a model of dopamine neurons. Neurocomputing 69:1030–1034.
- Optican LM, Richmond BJ (1987) Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. J Neurophysiol 57:162–178.
- Panzeri S, Brunel N, Logothetis NK, Kayser C (2010) Sensory neural codes using multiplexed temporal scales. Trends Neurosci 33:111–120.
- Panzeri S, Ince R a a, Diamond ME, Kayser C (2014) Reading spike timing without a clock: intrinsic decoding of spike trains. Philos Trans R Soc Lond B Biol Sci 369:20120467.
- Pareti G, De Palma A (2004) Does the brain oscillate? The dispute on neuronal synchronization. Neurol Sci 25:41–47.
- Pasley BN, David S V., Mesgarani N, Flinker A, Shamma SA, Crone NE, Knight RT, Chang EF (2012) Reconstructing speech from human auditory cortex Zatorre R, ed. PLoS Biol 10:e1001251.
- Peelle JE, Gross J, Davis M (2013) Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb Cortex 23:1378–1387.
- Penny WD, Duzel E, Miller KJ, Ojemann JG (2008) Testing for nested oscillation. J Neurosci Methods 174:50–61.
- Peterson GE, Wang WS, Sivertsen E (1958) Segmentation Techniques in Speech Synthesis. J Acoust Soc Am 30:682.
- Pietersen ANJ, Ward PD, Hagger-Vaughan N, Wiggins J, Jefferys JGR, Vreugdenhil M (2014) Transition between fast and slow gamma modes in rat hippocampus area CA1 in vitro is modulated by slow CA3 gamma oscillations. J Physiol 592:605–620.
- Poeppel D (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as “asymmetric sampling in time.” Speech Commun 41:245–255.
- Poeppel D (2014) The neuroanatomic and neurophysiological infrastructure for speech and language. Curr Opin Neurobiol 28C:142–149.
- Poeppel D, Monahan P (2008) Speech Perception Cognitive Foundations and Cortical Implementation. Curr Dir Psychol ... 17:80–85.
- Quian Quiroga R, Sakowitz OW, Basar E, Schürmann M (2001) Wavelet Transform in the analysis of the frequency composition of evoked potentials. Brain Res Protoc 8:16–24.
- Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87.
- Ray S, Maunsell JHR (2014) Do gamma oscillations play a role in cerebral cortex? Trends Cogn Sci 19:78–85.

- Read HL, Winer J a, Schreiner CE (2002) Functional architecture of auditory cortex. *Curr Opin Neurobiol* 12:433–440.
- Recasens M, Grimm S, Wollbrink A, Pantev C, Escera C (2014) Encoding of nested levels of acoustic regularity in hierarchically organized areas of the human auditory cortex. *Hum Brain Mapp* 35:5701–5716.
- Reed W, Hughes B (2002) From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature. *Phys Rev E* 66:067103.
- Rockel AJ, Hiorns RW, Powell TP (1980) The basic uniformity in structure of the neocortex. *Brain* 103:221–244.
- Roopun AK, Kramer M a, Carracedo LM, Kaiser M, Davies CH, Traub RD, Kopell NJ, Whittington M a (2008a) Temporal Interactions between Cortical Rhythms. *Front Neurosci* 2:145–154.
- Roopun AK, Kramer M a, Carracedo LM, Kaiser M, Davies CH, Traub RD, Kopell NJ, Whittington M a (2008b) Period concatenation underlies interactions between gamma and beta rhythms in neocortex. *Front Cell Neurosci* 2:1.
- Rosen S (1992) Temporal information in speech : Acoustic, Auditory and Linguistic Aspects. *Philos Trans Biol Sci* 336:367–373.
- Rosen S, Howell P (2011) Signals and Systems for Speech and Hearing.
- Rotstein HG, Pervouchine DD, Acker CD, Gillies MJ, White JA, Buhl EH, Whittington MA, Kopell N, Horacio G, Miles A (2005) Slow and fast inhibition and an H-current interact to create a theta rhythm in a model of CA1 interneuron network. *J ...*:1509–1518.
- Roux F, Uhlhaas PJ (2013) Working memory and neural oscillations: alpha-gamma versus theta-gamma codes for distinct WM information? *Trends Cogn Sci* 18.
- Saalmann YB, Pinsk M a, Wang L, Li X, Kastner S (2012) The pulvinar regulates information transmission between cortical areas based on attention demands. *Science* 337:753–756.
- Salinas E, Sejnowski TJ (2001) Correlated neuronal activity and the flow of neural information. *Nat Rev Neurosci* 2:539–550.
- Sanchez-Vives M V, McCormick DA (2000) Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nat Neurosci* 3:1027–1034.
- Schabus M, Pelikan C, Chwala-Schlegel N, Weilhart K, Roehm D, Donis J, Michitsch G, Pichler G, Klimesch W (2011) Oscillatory brain activity in vegetative and minimally conscious state during a sentence comprehension task. *Funct Neurol* 26:31–36.
- Schirmer A, Fox PM, Grandjean D (2012) On the spatial organization of sound processing in the human temporal lobe: a meta-analysis. *Neuroimage* 63:137–147.
- Schoffelen J-M, Oostenveld R, Fries P (2005) Neuronal coherence as a mechanism of effective corticospinal interaction. *Science* 308:111–113.
- Schroeder CCE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci* 32:9–18.
- Schyndl PG, Thut G, Gross J (2011) Cracking the code of oscillatory activity. *PLoS Biol* 9:e1001064.
- Seth AK (2010) A MATLAB toolbox for Granger causal connectivity analysis. *J Neurosci Methods* 186:262–273.
- Shadlen MN, Movshon JA (1999) Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron* 24:67–77, 111–125.
- Shamir M, Ghitza O, Epstein S, Kopell NJ (2009) Representation of time-varying stimuli by a network exhibiting oscillations on a faster time scale. *PLoS Comput Biol* 5:e1000370.
- Shlizerman E, Holmes P (2012) Neural Dynamics, Bifurcations, and Firing Rates in a Quadratic Integrate-and-Fire Model with a Recovery Variable. I: Deterministic Behavior. *Neural Comput* 24:2078–2118.

- Siegel M, Donner TH, Engel AK (2012) Spectral fingerprints of large-scale neuronal interactions. *Nat Rev Neurosci* 13:121–134.
- Silberberg G, Markram H (2007a) Disynaptic inhibition between neocortical pyramidal cells mediated by Martinotti cells. *Neuron* 53:735–746.
- Silberberg G, Markram H (2007b) Disynaptic inhibition between neocortical pyramidal cells mediated by Martinotti cells. *Neuron* 53:735–746.
- Singer W (2013) Cortical dynamics revisited. *Trends Cogn Sci* 17:616–626.
- Sirota A, Montgomery S, Fujisawa S, Isomura Y, Zugaro M, Buzsáki G (2008) Entrainment of neocortical neurons and gamma oscillations by the hippocampal theta rhythm. *Neuron* 60:683–697.
- Smeal RM, Ermentrout GB, White J a (2010) Phase-response curves and synchronized neural networks. *Philos Trans R Soc Lond B Biol Sci* 365:2407–2422.
- Softky WR (1995) Simple codes versus efficient codes. *Curr Opin Neurobiol* 5:239–247.
- Spratling MW (2008) Predictive coding as a model of biased competition in visual attention. *Vision Res* 48:1391–1408.
- Staudigl T, Hanslmayr S (2013) Theta oscillations at encoding mediate the context-dependent nature of human episodic memory. *Curr Biol* 23:1101–1106.
- Steriade M (2006) Grouping of brain rhythms in corticothalamic systems. *Neuroscience* 137:1087–1106.
- Stevens KN (2005) Features in Speech Perception and Lexical Access. In: *The Handbook of Speech Perception*, pp 124–155. Blackwell Publishing Ltd.
- Strong S, Koberle R, Steveninck R van, Bialek W (1998) Entropy and information in neural spike trains. *Phys Rev Lett* 80:197–200.
- Tallon-Baudry C, Kreiter A, Bertrand O (1999) Sustained and transient oscillatory responses in the gamma and beta bands in a visual short-term memory task in humans. *Vis Neurosci* 16:449–459.
- Tateno T, Harsch A, Robinson HPC (2004) Threshold firing frequency-current relationships of neurons in rat somatosensory cortex: type 1 and type 2 dynamics. *J Neurophysiol* 92:2283–2294.
- Tateno T, Robinson HPC (2007) Phase resetting curves and oscillatory stability in interneurons of rat somatosensory cortex. *Biophys J* 92:683–695.
- Teramae JN, Tanaka D (2004) Robustness of the noise-induced phase synchronization in a general class of limit cycle oscillators. *Phys Rev Lett* 93:204103.
- Terman D, Kopell N, Bose A (1998) Dynamics of two mutually coupled slow inhibitory neurons. *Phys D Nonlinear Phenom* 117:241–275.
- Tiesinga PH, Sejnowski TJ (2009) Cortical Enlightenment: Are Attentional Gamma Oscillations Driven by ING or PING? *Neuron*.
- Tort AABL, Rotstein HHG, Dugladze T, Gloveli T, Kopell NJ (2007a) On the formation of gamma-coherent cell assemblies by oriens lacunosum-moleculare interneurons in the hippocampus. *PNAS* 104:13490–13495.
- Tort ABL, Rotstein H, Dugladze T, T (2007b) On the formation of gamma-coherent cell assemblies by oriens lacunosum-moleculare interneurons in the hippocampus. *PNAS*.
- Tsubo Y, Takada M, Reyes AD, Fukai T (2007) Layer and frequency dependencies of phase response properties of pyramidal neurons in rat motor cortex. *Eur J Neurosci* 25:3429–3441.

- Tsukada M, Ichinose N, Aihara K, Ito H, Fujii H (1996) Dynamical Cell Assembly Hypothesis - Theoretical Possibility of Spatio-temporal Coding in the Cortex. *Neural Networks* 9:1303–1350.
- Van der Meij R, Kahana MJ, Maris E (2012) Phase-amplitude coupling in human electrocorticography is spatially distributed and phase diverse. *J Neurosci* 32:111–123.
- Van Kerkoerle T, Self MW, Dagnino B, Gariel-Mathis M -a., Poort J, van der Togt C, Roelfsema PR (2014) Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc Natl Acad Sci* 111:14332–14341.
- Van Vreeswijk C, Abbott LF, Ermentrout G (1994) When inhibition not excitation synchronizes neural firing. *J Comput Neurosci* 1:313–321.
- Vandecasteele M, Varga V, Berényi A, Papp E, Barthó P, Venance L, Freund TF, Buzsáki G (2014) Optogenetic activation of septal cholinergic neurons suppresses sharp wave ripples and enhances theta oscillations in the hippocampus. *Proc Natl Acad Sci* 111:13535–13540.
- VanRullen R, Koch C (2003) Is perception discrete or continuous? *Trends Cogn Sci* 7:207–213.
- Varela F, Lachaux JP, Rodriguez E (2001) The brainweb: phase synchronization and large-scale integration. *Nat Rev* 2.
- Vierling-Claassen D, Cardin J (2010) Computational modeling of distinct neocortical oscillations driven by cell-type selective optogenetic drive: separable resonant circuits controlled by low-threshold spiking and fast-spiking interneurons. *Front Hum ...* 4:198.
- Vierling-Claassen D, Kopell NJ (2009) The Dynamics of a Periodically Forced Cortical Microcircuit, With an Application to Schizophrenia. *SIAM J Appl Dyn Syst* 8:710.
- Von Stein A, Sarnthein J (2000) Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization. *Int J Psychophysiol* 38:301–313.
- Wacongne C, Labyt E, van Wassenhove V, Bekinschtein T, Naccache L, Dehaene S (2011) Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc Natl Acad Sci*:1–6.
- Wang D (2001) Relaxation Oscillators and Networks. In: *Wiley Encyclopedia of Electrical and Electronics Engineering*. John Wiley & Sons, Inc.
- Wang X (2010) Neurophysiological and Computational Principles of Cortical Rhythms in Cognition.
- Ward L (2003) Synchronous neural oscillations and cognitive processes. *Trends Cogn Sci* 7:553–559.
- White JA, Banks M, Pearce R, Kopell NJ (2000) Networks of interneurons with fast and slow-aminobutyric acid type A (GABA $A$ ) kinetics provide substrate for mixed gamma-theta rhythm. *PNAS* 97:8128–8133.
- Whittington M a, Traub RD, Kopell N, Ermentrout B, Buhl EH (2000) Inhibition-based rhythms: experimental and mathematical observations on network dynamics. *Int J Psychophysiol* 38:315–336.
- Winkler I, Czigler I (2011) Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. *Int J Psychophysiol* 83:132–143.
- Womelsdorf T, Schoffelen J-M, Oostenveld R, Singer W, Desimone R, Engel AK, Fries P (2007) Modulation of neuronal interactions through neuronal synchronization. *Science* 316:1609–1612.
- Wulff P, Ponomarenko AA, Bartos M, Korotkova TM, Fuchs EC, Bähner F, Both M, Tort ABL, Kopell NJ, Wisden W, Monyer H (2009) Hippocampal theta rhythm and its coupling with gamma oscillations require fast inhibition onto parvalbumin-positive interneurons. *Proc Natl Acad Sci U S A* 106:3561–3566.
- Yamada WM, Koch C, Adams PR (1989) Multiple channels and calcium dynamics. :97–133.

Young CK, Eggermont JJ (2009) Coupling of mesoscopic brain oscillations: recent advances in analytical and theoretical perspectives. *Prog Neurobiol* 89:61–78.

Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon C a, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ, Poeppel D, Schroeder CE (2013) Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. *Neuron* 77:980–991.

## 11 Articles

### 11.1 Speech encoding by coupled cortical theta and gamma oscillations

#### Speech encoding by coupled cortical theta and gamma oscillations

##### Authors and Affiliations

**Hyafil A.<sup>1</sup>, Fontolan L.<sup>1,2</sup>, Kabdebon, C.<sup>1</sup>, Gutkin B.<sup>1,3</sup> and Giraud AL.<sup>2</sup>**

1. Inserm U960. Group for Neural Theory, Département d'Etudes Cognitives, Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris. France.

2. Department of Neuroscience, Biotech Campus, University of Geneva, 9 rue des Mines, 1211 Geneva. Switzerland.

3. Centre for Cognition and Decision Making, National Research University Higher School of Economics, Myasnitskaya St. 20, Moscow 101000, Russia

##### Abstract

Many environmental stimuli present a quasi-rhythmic structure at different timescales that the brain needs to decompose and integrate. Cortical oscillations have been proposed as instruments of sensory *de-multiplexing*, i.e., the parallel processing of different frequency streams in sensory signals. Yet their causal role in such a process has never been demonstrated. Here we used a neural microcircuit model to address whether coupled theta-gamma oscillations, as observed in human auditory cortex, could underpin the multiscale sensory analysis of speech. We show that, in continuous speech, theta oscillations can flexibly track the syllabic rhythm and temporally organize the phoneme-level response of gamma neurons into a code that enables syllable identification. The tracking of speech slow fluctuations by theta oscillations, and its coupling to gamma-spiking activity both appeared as critical features for accurate speech encoding. These results demonstrate that cortical oscillations can be a key instrument of speech de-multiplexing, parsing, and encoding.

## Introduction

The physical complexity of biological and environmental signals poses a fundamental problem to the sensory systems. Sensory signals are often made of different rhythmic streams organized at multiple time scales, which require to be processed in parallel and recombined to achieve unified perception. Speech constitutes an example of such a physical complexity, in which different rhythms index linguistic representations of different granularities, from phoneme to syllables and words (Rosen 1992; Zion Golumbic, Poeppel, and Schroeder 2012). Before meaning can be extracted from continuous speech, two critical pre-processing steps need to be carried out: a demultiplexing step, i.e. the parallel analysis of each constitutive rhythm, and a parsing step, i.e., the discretization of the acoustic signal into linguistically relevant chunks that can be individually processed (Poeppel 2003; Ghitza 2011; Stevens 2002). While parsing is presumably modulated in a top-down way, by knowing *a priori* through developmental learning (Ngon et al. 2013) where linguistic boundaries should lie, it is likely largely guided by speech acoustic dynamics. It has recently been proposed that speech de-multiplexing and parsing could both be handled in a bottom-up way by the combined action of auditory cortical oscillations in distinct frequency ranges, enabling parallel computations at syllabic and phonemic timescales (Ghitza 2011; Giraud and Poeppel 2012). Intrinsic coupling across cortical oscillations of distinct frequencies, as observed in electrophysiological recordings of auditory cortex (P. P. Lakatos et al. 2005; Fontolan et al. 2014), could enable the hierarchical combination of syllabic- and phonemic-scale computations, subsequently restoring the natural arrangement of phonemes within syllables (Giraud and Poeppel 2012).

The most pronounced energy fluctuations in speech occur at about 4 Hz (Zion Golumbic, Poeppel, and Schroeder 2012) and can serve as an acoustic guide for signalling the syllabic rhythm (Mermelstein 1975). Since the syllabic rate coincides with the auditory cortex theta rhythm (3-8 Hz), syllable boundaries could be viably signalled by a given *phase* in the theta cycle. The relevance of speech tracking by the theta neural rhythm (Henry, Herrmann, and Obleser 2014) is highlighted by experimental data showing that speech intelligibility depends on the degree of phase-locking of the theta-range neural activity in auditory cortex (Ahissar et al. 2001; Luo and Poeppel 2007; Peelle, Gross, and Davis 2013; Gross et al. 2013). By analogy with the spatial and mnemonic oscillatory processes that take place in the hippocampus (O Jensen and Lisman 1996; Lisman and Jensen 2013;

Lever, Kaplan, and Burgess 2014), the theta oscillation may orchestrate gamma neural activity to facilitate its subsequent decoding (Canolty et al. 2007): the phase of theta-paced neural activity could regulate faster neural activity in the low-gamma range ( $>30$  Hz) involved in linguistic coding of phonemic details (Ghitza 2011; Giraud and Poeppel 2012). The control of gamma by theta oscillations could hence both modulate the excitability of gamma neurons to devote more processing power to the informative parts of syllabic sound patterns, and constitute a reference time-frame aligned on syllabic contours for interpreting gamma-based phonemic processing (Shamir et al. 2009; Ghitza 2011; Kayser, Ince, and Panzeri 2012; Panzeri et al. 2014).

Compelling as this hypothesis may sound, direct evidence for neural mechanisms linking speech constituents and oscillatory components is still lacking. One way to address a causal role of oscillations in speech processing is computational modelling, as it permits to directly test the efficiency of cross-coupled theta and gamma oscillations as an instrument of speech de-multiplexing, parsing and encoding. Previous models of speech processing involved only gamma oscillations in the context of isolated speech segments (Shamir et al. 2009) or did not involve neural oscillations at all (Gütig and Sompolinsky 2009; Yildiz, von Kriegstein, and Kiebel 2013). On the other hand previous models of cross-frequency coupled oscillations did not address sensory functions as parsing and demultiplexing (O Jensen and Lisman 1996; Tort et al. 2007). Here, we examined how a biophysically-inspired model of coupled theta and gamma neural oscillations can process continuous speech (spoken sentences). Specifically we determined: (i) whether theta oscillations are able to accurately parse speech into syllables, (ii) whether syllable-related theta signal may serve as a reference time frame to improve gamma-based decoding of continuous speech; (iii) whether this decoding requires theta to modulate the activity of the gamma network. To address the last two points, we compared speech decoding performance of the model with two control versions of the network, in which we removed the neural connection entraining the theta neurons by speech fluctuations or the link that couples them to the gamma neurons.

## Results

### Model architecture and spontaneous behaviour

The model proposed here (Figure 1A) is inspired from cortical architecture (Douglas and Martin 2004; da Costa and Martin 2010) and function (P. Lakatos et al. 2007) as well as from previous biophysical models of cross-frequency coupled oscillation generation (Kopell et al. 2010; Tort et al. 2007; Vierling-Claassen et al. 2010). We used the well documented PING (Pyramidal Interneuron Gamma) model for implementing a gamma network: bursts of inhibitory neurons immediately follow bursts of excitatory neurons (Jadi and Sejnowski 2014), creating the overall all spiking rhythm. Given that gamma and theta oscillations are both locally present in superficial cortical layers (P. P. Lakatos et al. 2005), we assume a similar local generation mechanisms for theta and gamma with a direct connection between them. Direct evidence for a local generation of theta oscillations in auditory cortex is still scarce (Ainsworth et al. 2011) and we cannot completely rule out that they might spread from remote generators (e.g. in the hippocampus (Kopell et al. 2010; Tort et al. 2007)). Yet, we built the case for local generation from the following facts: 1) neocortical (somatosensory) theta oscillations are observed *in-vitro* (Fanselow, Richardson, and Connors 2008), 2) MEG, EEG and combined EEG/FMRI recordings in humans show that theta activity phase-locks to speech amplitude envelope in A1 and immediate association cortex - but not beyond - (Ahissar et al. 2001; Luo and Poeppel 2007; Cogan and Poeppel 2011; Morillon et al. 2012), and 3) theta phase-locking to speech is not accompanied by power increase, arguing for a phase restructuring of a local oscillation (Luo and Poeppel 2007). We assumed a similar generation mechanism for theta and gamma oscillations, with slower excitatory and inhibitory synaptic time constants for theta (Vierling-Claassen et al. 2010; Kopell et al. 2010). The distinct dynamics for the two modules reflect the diversity of inhibitory synaptic time scales observed experimentally, with Martinotti cells displaying slow synaptic inhibition ( $Ti$  neurons), and basket cells showing faster inhibition decay ( $Gi$  neurons) (Silberberg and Markram 2007). We refer to the theta network as PINTH (Pyramidal Interneuron Theta), by analogy with PING. The full model is hence composed of a theta-generating module with interconnected spiking excitatory ( $Te$ ) and inhibitory ( $Ti$ ) neurons that spontaneously synchronize at theta frequency (6-8 Hz) through slow decaying inhibition; and of a gamma-generating module with excitatory ( $Ge$ ) and inhibitory ( $Gi$ ) neurons that burst at a faster rate (25-45 Hz) synchronized by fast decaying inhibition (PING; Figure 1B) (Börgers and Kopell 2005). The firing pattern of our simulated neurons is sparse and weakly synchronous at rest, consistent with the low spiking rate of cortical neurons (Brunel and Wang 2003). Unlike

the classical 50-80 Hz PING seen in *in-vitro* preparations of rat auditory cortex (Ainsworth et al. 2011), our network produced a lower gamma frequency around 30 Hz, as observed in human auditory cortex in response to speech (Nourski et al. 2009; Pasley, David, and Mesgarani 2012),

At rest the PINTH population activity synchronizes at the theta time scale, and the PING population at the gamma time scale. Both the *Te* and *Ge* populations receive projections from a ‘subcortical’ module that mimics the non-linear filtering of acoustic input by subcortical structures, which primarily includes a signal decomposition into 32 auditory channels (Chi, Ru, and Shamma 2005). Individual excitatory neurons in the theta module received channel-averaged input while those in the gamma module received frequency selective input. Such a differential selectivity was motivated by experimental observations from intracranial recordings (Fontolan et al. 2014; Morillon et al. 2012) suggesting that unlike the gamma one, the theta response does not depend on the input spectrum. It also mirrors the dissociation in primate auditory cortex between a population of ‘stereotyped’ neurons responding very rapidly and non-selectively to any acoustic stimulus (putatively *Te* neurons) and a population of ‘modulated’ neurons responding selectively to specific spectro-temporal features (putatively *Ge* neurons) (Brasselet et al. 2012). Each *Ge* neuron receives input from one specific channel, preserving the auditory tonotopy, so that the whole *Ge* population represents the rich spectral structure of the stimulus. Each *Te* neuron receives input from all the channels, i.e. the *Te* population conveys a widely tuned temporal signal capturing slow stimulus fluctuations. Importantly, the two oscillating modules are connected through all-to-all connections from *Te* neurons to *Ge* neurons allowing the theta oscillations to control the activity of the faster gamma oscillations. This structure enables syllable boundary detection (through the theta module) to constrain the decoding of faster phonemic information. The output of the network is taken from the *Ge* neurons as we assume that the *Ge* neurons provide the input to higher-level cortical structures performing operations like phoneme categorization and providing access to lexicon. Accordingly, in the model the *Ge* neurons receive more spectral details about speech than the *Te* neurons (Figure 1B). *Ge* spiking is then referenced with respect to timing of theta spikes, and submitted to decoding algorithms.

#### **Model dynamics in response to natural sentences**

We first explored the dynamic behaviour of the model. As expected from its architecture and biophysical parameters (see *Methods*), the neural network produced activity in theta (6-8 Hz) and low gamma (25-45 Hz) ranges, both at rest and during speech presentation. Consistent with experimental observations (Luo and Poeppel 2007) there was no notable increase in theta spiking during speech presentation, but sentence onsets induced a phase-locking of theta oscillations as shown by the PSTH of theta neurons, which was further enhanced by all edges in speech envelope. Consequently, the resulting global evoked activity followed the acoustic envelope of the speech signal (Figure 1C) (Abrams et al. 2008). LFP indexes the global synaptic activity over the network (excitatory neurons of both networks) and its dynamics closely followed spiking dynamics. Unlike the LFP theta power pattern, the LFP theta phase pattern was robust across repetitions of the same sentence (Figure 1 – Figure Supplement 1A,1C), replicating LFP behaviour from the primate auditory cortex (Kayser et al. 2009), and human MEG data (Luo and Poeppel 2007; Luo, Liu, and Poeppel 2010). In line with other empirical data from human auditory cortex (Nourski et al. 2009) gamma oscillations followed the onset of sentences (Figure 1C). Owing to the feed-forward connection from the theta to the gamma sub-circuits, the gamma amplitude was coupled to the theta phase both at rest and during speech (Figure 1D). The coupling was visible both in the spiking (Figure 1 – Figure Supplement 1B) and LFP signal (Figure 1D). Critically, this coupling disappeared when the theta/gamma connection was removed, showing that a common input to *Te* and *Ge* cells is not sufficient to couple the two oscillations.

#### **Syllable boundary detection by theta oscillations**

Before testing the speech decoding properties of the model, we explored whether syllable boundaries could reliably be detected at the cortical level by a theta network (see *Methods*). This first study was based on a corpus consisting of 4620 phonetically-labelled English sentences (TIMIT Consortium, 1993). The acoustic analysis of these sentences confirmed a correspondence between the dominant peak of the speech modulation spectrum and the mean syllabic rate (3-6 Hz) (Figure 2 – Figure Supplement 1A), whereby syllabic boundaries correspond to trough in speech slow fluctuations (Peelle, Gross, and Davis 2013). The theta network in the model (Figure 2 – Figure Supplement 1B) was explicitly designed to exploit such regularities and infer syllable boundaries. When presenting sentences to the theta module, we observed a consistent theta burst within 50 ms

following syllable onset followed by a locking of theta oscillations to theta acoustic fluctuations in the speech signal (Figure 2 – Figure Supplement 1C-D). More importantly, neuronal theta bursts closely aligned to the timing of syllable boundaries in the presented sentences (Figure 2A). We compared the performance of the theta network to that of two alternative models also susceptible to predict syllable boundaries: a simple linear-nonlinear acoustic boundary detector (Figure 2 – Figure Supplement 1E), and Mermelstein algorithm, a state-of-the-art model which, unlike the model developed here, only permits “off-line” syllable boundary detection (Mermelstein 1975). The theta network performed better than both the linear model and the Mermelstein algorithm (Figure 2B, all p-values  $<10^{-12}$ ). Similar to results from behavioral studies of human perception (Miller, Grosjean, and Lomanto 1984; Nourski et al. 2009; Mukamel et al. 2011) the theta network could adapt to different speech rates. The model performed better than other algorithms, with a syllabic alignment accuracy remaining well above chance levels ( $p<10^{-12}$ ) in the twofold and threefold time compression conditions. (Figure 2B).

This first study demonstrates that theta activity provides a reliable, syllable-based, internal time reference that the neural system could use when reading out the activity of gamma neurons.

#### **Decoding of simple temporal stimuli from output spike patterns**

Our next step was to test whether the theta-based syllable chunks of output spike trains (*Ge* neurons) for the different input types could be properly classified. We first quantified the model’s ability to encode stimuli designed as simple temporal patterns. We used 50 ms sawtooth stimuli whose shape was parametrically varied by changing the peak position (Figure 3A), with inter-stimulus interval between 50 and 250 ms. This toy set of stimuli was previously used in a gamma-based speech encoding model and argued to represent idealized formant transitions (Shamir et al. 2009). We extracted spike patterns from all the *Ge* (output) neurons from -20 ms before each sawtooth onset to 20 ms after its offset. This procedure is referred to as “stimulus timing” since it uses the stimulus onset as time reference. Using a clustering method (see Methods), we observed that the identity of the presented sawtooth could be decoded from the output spike patterns (Figure 3A) with over 60% accuracy (Figure 3C, light gray bar). We also computed the decoding performance when we used an internal time reference provided by the theta timing rather than by the stimulus timing. When spike patterns were analysed within a window defined by two

successive theta bursts (Figure 3C, dark grey bar), sawtooth decoding was still possible and even relatively well preserved (mean decoding rate of 41.7%).

We then compared the decoding performance from the full model with that of two control models: one in which the theta module was not driven by the stimulus (*undriven theta* model) and one in which the theta module was not connected with the gamma module (*uncoupled theta/gamma* model) (Figure 3B, green and blue). Decoding performance of both control models, as revealed by the mean performance (Figure 3C) and confusion matrices (Figure 3E), was degraded for either neural code (theta onset and stimulus timing, all p-values  $<10^{-9}$ ). The details of the raw confusion matrices show that the temporal patterns are decoded correctly or as a neighbouring temporal shape only in the intact version of the model (Figure 3E). Furthermore, the intact model achieved better signal vs. rest discrimination than the two control models, notably avoiding false alarms (Figure 3D). In summary, these analyses show that gamma-spiking neurons within theta bursts provide a reliable internal code for characterizing simple temporal patterns, and that this ability is granted by the time-locking of theta neurons ( $Te$  units) to stimulus and the modulation they exert on the fast-scale output ( $Ge$ ) units.

#### **Continuous speech encoding by model output spike patterns**

The overarching goal of this theoretical work was to assess whether coupled cortical oscillations can achieve on-line speech decoding from *continuous* signal. We therefore set out to classify syllables from natural sentences. To decode  $Ge$  spiking, we used similar procedures as for the encoding/decoding of simple temporal patterns. Output  $Ge$  spikes were parsed into spike patterns based on the theta chunks, and the decoding analysis was used to recover syllable identity (Figure 4A). To evaluate the importance of the precise spike timing of gamma neurons, we compared decoding (see methods) using spike *patterns* (i.e. spikes labelled with their precise timing w.r.t. chunk onset) versus those obtained from plain spike *counts* (i.e. unlabelled spikes). When using spike patterns syllable decoding reached a high level of accuracy in the intact model: 58% of syllables were correctly classified within a set of 10 possible (randomly chosen) syllables (Figure 4B). Syllable decoding dropped when using spike counts instead of spike patterns ( $p < 10^{-12}$ ). Critically, decoding was poor in both control models (undriven theta and uncoupled theta/gamma) using either spike counts or spike patterns (significantly lower than decoding using spike

patterns in the full model, all p-values  $< 10^{-12}$ , and non-significantly higher than decoding using spike counts in the full model, all p-values  $> 0.08$  uncorrected).

We also explored the model performance for encoding syllables spoken by different speakers. We used a similar decoding procedure as above, but here the classifier was trained on different speakers pronouncing the same two sentences. Theta chunks were classified into syllables based on the network response to the two sentences uttered by 99 other speakers. The material included sentences spoken by 462 speakers of various ethnic and geographical origins, showing a marked heterogeneity in phonemic realization and syllable durations (as labelled by phoneticians). The syllable duration distribution was skewed with the median at 200 ms and tail values ranging from a few ms to over 800ms (Figure 4 Supp. 1A). Given that theta activity is meant to operate in a 3-9 Hz range, i.e. integrate speech chunks of about 100 to 300 ms (Ghitza 2011; Ghitza 2014), we did not expect the model to perform equally well along the whole syllable duration range. Accordingly, decoding accuracy was not uniform across the whole syllable duration range. When decoding from spike pattern, the intact model allowed 24% accuracy (chance level at 10%). It showed a peak in performance in the range in which it is expected to operate, i.e. for syllables durations between 100 and 300 ms. Given the cross-speaker phonemic variability such a performance is fairly good. Critically, the intact model outperformed control models both within the 100 to 300 ms range ( $p < 0.001$ ), and throughout the whole syllable duration span ( $p < 0.001$ ). These analyses overall show that the model can flexibly track syllables within a physiological operating window, and that syllable decoding relies on the integrity of the model architecture.

Lastly, we tested more directly the resilience of the spike pattern code to speech temporal compression and found that while degrading the decoding performance remained above chance for compression rates of 2 and 3 (Figure 4D). Altogether, the decoding of syllables from continuous speech showed that coupled theta and gamma oscillations provide a viable instrument for syllable parsing and decoding and that its performance relies on the coupling between the two oscillation networks.

#### Encoding properties of model neurons

We finally assessed the physiological plausibility of the model by comparing the encoding properties of the simulated neurons, without further parameter fitting, with those of

neurons recorded from primate auditory cortex (Kayser et al. 2009; Kayser, Ince, and Panzeri 2012). The first analysis of neural encoding properties consisted of comparing the ability to classify neural codes from the model into arbitrary speech segments of fixed duration (as opposed to classification into syllables as in previous section). We simulated data using natural speech and studied the spiking activity of *Ge* neurons by implementing the same methods of analysis as in the original experiment. We extracted fixed-size windows of spike patterns activity for individual *Ge* neurons, and assessed neural encoding characteristics using different neural codes. Speech encoding was first evaluated using a nearest-mean classifier and then using mutual information techniques (Kayser et al. 2009).

*Classifier analysis.* In this analysis neural patterns were classified not into syllables as above or into any linguistic constituent but into arbitrary segments of speech, allowing for a-theoretical insight into the encoding properties of neurons. We extracted a subset of 25 sentences from the TIMIT corpus and exposed the network to 50 presentations of each sentence from the subset. We defined 10 stimuli as 10 distinct windows of a given size (from 80 to 480 ms) randomly extracted from the 25 sentences, and then assessed the capacity to decode the identity of a stimulus from the activity of individual *Ge* neurons within that window (Kayser, Ince, and Panzeri 2012). Three different codes were used (Figure 5A): a simple *spike count* was used as reference code; a *time-partitioned code* where spikes were assigned to one of 8 bins of equal duration within the temporal window; a *phase-partitioned code* where spikes were labelled with the phase of LFP theta at the timing of spike (the spikes were then assigned into one of 8 bins according to their phase). We observed that for 80 to 240 ms windows (within one theta cycle), decoding was almost as good for the phase-partitioned code as for the time-partitioned code (Figure 5B, left). In other words, stimulus decoding using theta timing was nearly as good as when using stimulus timing. Performance using the spike count was considerably lower ( $p < 10^{-12}$  for all 6 window sizes). Overall, there was a qualitative and even quantitative match between the results from simulated data and the original experimental results (Figure 5B, right). When we removed either the input-to-theta (undriven theta model) or the theta-to-gamma connection (uncoupled theta/gamma model) in the network, the performance of the phase-partitioned code dropped to just above that of the spike count code (Figure 5 – Figure Supplement 1A; significantly lower increase in decoding performance using phase-partitioned instead of spike count code compared to full model,  $p < 10^{-12}$  for all 6 window sizes and both control models), and the simulations no longer predicted the experimental

results. Finally, experimental data and simulations from the intact model also matched when we investigated the dependence of decoding accuracy on the number of bins, which was not the case for any of the control models (Figure 5 – Figure Supplement 1B).

*Mutual information (MI) analysis.* MI between the input (acoustic stimulus) and the output (neural pattern) provides an alternative measure for how well stimuli are encoded in the output pattern (see Methods). We used the same simulation data as for the classification procedure, but the sentences were subdivided into shorter chunks using a non-overlapping time window (length T: 8–48 ms) (Kayser et al. 2009). We compared the MI between the stimulus and neural activity in individual *Ge* neurons as a function of the length of stimulus window, using four neural codes: spike count, time-partitioned code, phase-partitioned code combined with spike count and finally combined phase- and time-partitioned codes. These codes are qualitatively equivalent to the decoding strategies used in the previous classifier analysis. Figure 5C shows that taking into account the spike phase boosts the MI carried by the *Spike count code* or the *Time-partitioned code* alone ( $p < 10^{-12}$  for all 6 window sizes). In other words, spike phase provided additional rather than redundant information to more traditional codes. The gain provided by spike phase increased when enlarging the window and when combined with either spike count or spike pattern (*Spike Count vs. Time-partitioned*; *Spike count and Phase-partitioned code vs. Time- and Phase-partitioned code*). These results replicate the original experimental data from monkey auditory cortex (Kayser et al. 2009). Such a pattern was not reproduced using any of the control models (Figure 5 – Figure Supplement 1C). These results hence show that in addition to enhancing the reliability of the spike phase code, the theta-gamma connection enhanced the temporal precision of *Ge* neurons spiking in response to speech stimuli.

Critically, results from both classifier and mutual information analyses demonstrate that the full network architecture of the model provides an efficient way of boosting the encoding capacity of neurons in a way that bears remarkable similarities to actual neurons from primate auditory cortex.

## Discussion

Like most complex natural patterns, speech contains rhythmic activity at different scales that conveys different and sometimes non-independent categories of information. Using a

biophysically-inspired model of auditory cortex function, we show that cortical theta-gamma cross-frequency coupling provides a means of using the timing of syllables to orchestrate the readout of speech-induced gamma activity. The current modelling data demonstrate that theta bursts generated by a theta (PINTH) network can predict “on-line” syllable boundaries at least as accurately as state-of-the-art offline syllable detection algorithms. Syllable boundary detection by a theta network hence provides an endogenous time reference for speech decoding. Our simulated data further show that a gamma biophysical network, receiving a spectral decomposition of speech as input, can take advantage of the theta time reference to encode fast phonemic information. The central result of our work is that the gamma network could efficiently encode temporal patterns (from simple sawtooths to natural speech), as long as it was entrained by the theta rhythm driven by syllable boundaries. The proposed theta/gamma network displayed sophisticated spectral and encoding properties that compared both qualitatively and quantitatively to existing neurophysiological evidence including cross-frequency coupling properties (Schroeder and Lakatos 2009) and theta-referenced stimulus encoding (Kayser et al. 2009; Kayser, Ince, and Panzeri 2012). The projections from the *Te* to *Ge* neurons endowed the network with phase-amplitude and phase-frequency coupling between gamma and theta oscillations, at both the spike and the LFP levels (Ole Jensen and Colgin 2007). This closely reproduces the theta/gamma phase-amplitude coupling observed from intracortical recordings (Giraud and Poeppel 2012; P. P. Lakatos et al. 2005). Importantly, due to the dissociation of excitatory populations we obtained denser gamma spiking immediately after the theta burst evoked by the syllable onset. This validates a critical point of theta/gamma parsing system, namely that a more in-depth encoding is carried-out by the auditory cortex during the early phase of syllables, when more information needs to be extracted (Giraud and Poeppel 2012; Schroeder and Lakatos 2009).

The human auditory system, like other sensory systems, is able to produce invariant responses to different physical presentations of the same input. Importantly, it is relatively insensitive to the speed at which speech is being produced. Speech can double in speed from one speaker to another, and yet remain intelligible up to an artificial compression factor of 3. In the current model, theta bursts could still signal syllable boundaries when speech was compressed by a factor 2 and this alignment deteriorated for higher compression factors. Syllable decoding was significantly degraded for compressed speech, yet remained twice as accurate as chance. Our network is purely bottom-up and does not

include high level linguistic processes and representations, which in all likelihood plays an important role in speech perception (Davis et al. 2011; Peelle, Gross, and Davis 2013; Gagnepain, Henson, and Davis 2012): its relative resilience to speech compression is thus a fairly good performance. A previous model (Gütig and Sompolinsky 2009) proposed a neural code that was robust to speech warping, based on the notion that individual neurons correct for speech rate by their overall level of activity. While this model achieved very good speech categorization performance, it relied on extremely precise spiking behaviour (neurons spiked only once, when their associated channel reached a certain threshold), for which neurophysiological evidence is scarce. Another model developed by Hopfield proposes that a low gamma external current provides encoding neurons with reliable timing and dynamical memory spanning up to 200 ms, a long enough window to integrate information over a full syllable (Hopfield 2004). The utility of gamma oscillations for precise spiking is arguably similar in both Hopfield's model and ours, whereas the syllable integration process is irregularly ensured by intermittent traces of recent ( $\sim$ 200 ms) neural activity in Hopfield's, and in ours by regularly spaced theta bursts that are locked to the speech signal. The advantage of our model is that integration over long speech segments is *permanently* enabled by the phase of output spikes with respect to the ongoing theta oscillation. Our approach shows that accurate encoding can be achieved using a system that does not require explicit memory processes, and in which the temporal integration buffer is only emulated by a slow neural oscillator aligned to speech dynamics.

In the current combined theta/gamma model, theta oscillations do not only act as a syllable-scale integration buffer, but also as a precise neural timer. Because syllabic contours are reflected in the slow modulations of speech, the theta oscillator can flexibly entrain to them (3-7 Hz, Figure 2 – Figure Supplement 1A) and signal syllable boundaries. The spiking behaviour of theta neurons parallels experimental observations that a subset of neurons in A1 respond to the onset of naturalistic sounds (Fishbach, Nelken, and Yeshurun 2001; Phillips, Hall, and Boehnke 2002; Xiaoqin Wang et al. 2008), providing an endogenous time reference that serves as a landmark to decode from other neurons (Kayser, Ince, and Panzeri 2012; Brasselet et al. 2012; Panzeri and Diamond 2010; Panzeri et al. 2014). This parallels the dissociation between *Ge* and *Te* units in our model: while *Ge* units are channel specific, *Te* units cover the whole acoustic spectrum, which allow them to respond quickly and reliably to the onset of all auditory stimuli (Brasselet et al. 2012). In the model, however, theta neurons did not only discharge at stimulus onset but at

regular landmarks along the speech signal, the syllable boundaries (Zhou and Wang 2010). These neurons, hence, tie together the fast neural activity of gamma excitatory neurons into strings of linguistically relevant chunks (syllables), acting like punctuation in written language (Lisman and Buzsáki 2008). This mechanism for segmentation is conceptually similar to the segmentation of neural codes by theta oscillations in the hippocampus during spatial navigation (Gupta et al. 2012).

From an evolutionary viewpoint, because the theta rhythm is neither auditory- nor human-specific, it might have been incorporated as a speech-parsing tool in the course of language evolution. Likewise, human language presumably optimized the length of its main constituents, syllables, to the parsing capacity of the auditory cortex. As a result, syllables have the ideal temporal format to interface with, e.g., hippocampal memory processes, or with motor routines reflecting other types of rhythmic mechanical constraints, e.g. the natural motion rate of the jaw (4Hz) (Lieberman 1985).

Although conceptually promising, syllable tracking and speech encoding by a theta/gamma network, as proposed here, also shows some limitations. While our current model is purely bottom-up, top-down predictions play a significant role in guiding speech perception (Arnal and Giraud 2012; Gagnepain, Henson, and Davis 2012; Poeppel, Idsardi, and van Wassenhove 2008) presumably across different frequency channels and processing timescales (Xiao-jing Wang 2010; Bastos et al. 2012; Fontolan et al. 2014). How these predictions interplay with theta- and gamma-parsing activity remain unclear (Lee, Whittington, and Kopell 2013). Experimental findings suggest that theta activity might be at the interface of bottom-up and top-down processes (Peelle, Gross, and Davis 2013). Theta auditory activity is better synchronized to speech modulations when speech is intelligible, irrespective of its temporal or spectral structure (Luo and Poeppel 2007; Peelle, Gross, and Davis 2013). In the present model, theta activity bears an intrinsic temporal predictive function: it is driven by speech modulations, but is also resilient enough to syllable length variations to stay tuned to the global statistics of speech (average syllable duration). The model performed well above chance level when decoding syllables from a new speaker, showing flexibility in syllable tracking within a 3 to 9 Hz range. A natural follow-up of this work will hence be to explore how the intrinsic dynamics of theta and gamma activity interact not only with sensory input but also with linguistic top-down signals, e.g. word, sentence level predictions (Gagnepain, Henson, and Davis 2012), and even cross-modal predictions (Arnal et al. 2009). The trade-off between the autonomous

functioning of theta and gamma oscillatory activity on one hand, and their enslavement to sensory input on the other hand are at the core of future experimental and theoretical challenges.

In conclusion, our model provides a direct evidence that theta/gamma coupled oscillations can be a viable instrument to de-multiplex speech, and by extension to analyze complex sensory scenes at different timescales in parallel. By tying the gamma-organized spiking to the syllable boundaries, theta activity allows for decoding individual syllables in continuous speech streams. The model demonstrates the computational value of neural oscillations for parsing sensory stimuli based on their temporal properties, and offers new perspectives for syllable-based automatic speech recognition (Wu et al. 1997) and brain-machine interfaces using oscillation-based neuromorphic algorithms.

## Material and Methods

### Architecture of the full model

The model is composed of 4 types of cells: theta inhibitory neurons ( $T_i$ , 10 neurons), theta excitatory cells ( $T_e$ , 10 neurons), gamma inhibitory neurons ( $G_i$ , 32 neurons) and gamma excitatory neurons ( $G_e$ , 32 neurons) also called *output* neurons. All neurons were modeled as leaky integrate-and-fire neurons (LIF), where the dynamics of the membrane potential  $V_i$  of the neurons followed:

$$CdV_i/dt = g_L(V_L - V_i) + I_i^{SYN}(t) + I_i^{INP}(t) + I_i^{DC} + \eta(t)$$

where  $C$  is the capacitance of the membrane potential;  $g_L$  and  $V_L$  are the conductance and equilibrium potential of the leak current ;  $I^{SYN}$ ,  $I^{INP}$  and  $I^{DC}$  are the synaptic and constant currents, respectively ;  $\eta(t)$  is a Gaussian noise term of  $\sigma_t$  variance.

Whenever  $V_i$  reached the threshold potential  $V_{THR}$ , the neuron emitted a spike and  $V_i$  was turned back to  $V_{RESET}$ .

$I^{SYN}$  is the sum of all synaptic currents from all projecting neurons in the network:

$$I_i^{SYN}(t) = \sum_j g_{ij} s_{ij}(t)(V_j^{SYN} - V_i(t))$$

Where  $g_{ij}$  is the synaptic conductance of the  $j$ -to- $i$  synapse,  $s_{ij}(t)$  is the corresponding activation variable and  $V^{SYN}$  is the equilibrium potential of synaptic current (0 mV for excitatory neurons, -80 mV for inhibitory neurons). The activation variable  $s_{ij}(t)$  varies as

follow:

$$dx_j^R/dt = -1/\tau_j^R + \delta(t - t_j^{SPK})$$

$$ds_{ij}/dt = -1/\tau_j^D$$

where  $\tau_j^R$  and  $\tau_j^D$  are the time constants for synaptic rise and synaptic decay, respectively.

The connectivity among the cells is the following:

- $Te$  and  $Ti$  are reciprocally connected with all-to-all connections, generating the PINTH rhythm. There were also all-to-all connections within  $Ti$  cells.
- $Ge$  and  $Gi$  are also reciprocally connected with all-to-all connections, generating the PING rhythm.
- $Te$  projected with all-to-all connections to  $Ge$  cells, enabling cross-frequency coupling.

Input current  $I_i^{INP}(t)$  is non-null only for  $Te$  and  $Ge$  cells and follows the equation:

$$I_i^{INP}(t) = \sum_c \omega_{ci} x_c(t)$$

where  $x_c(t)$  is signal from channel  $c$  and  $\omega_{ci}$  is the weight of the projection from channel  $c$  to unit  $i$ .

Input to  $Te$  units is computed by filtering the auditory spectrogram by an optimized 2D spectro-temporal kernel (see section *LN model* below). LFP signal was simulated by summing the absolute values of all synaptic currents to all excitatory cells (both  $Ge$  and  $Te$ ), as in (Mazzoni et al. 2008). All simulations were run on Matlab. Differential equations were solved using Euler method with a time step of 0.005 ms. Values for all parameters are provided in Tables 1 and 2.

### Stimuli

We used oral recordings of English sentences produced by male and female speakers from the TIMIT database (Consortium 1993). The sentences were first processed through a model of subcortical auditory processing (Chi, Ru, and Shamma 2005) to the sentences. The model decomposes the auditory input into 128 channels of different frequency bands, reproducing the cochlear filterbank (<http://www.isr.umd.edu/Labs/NSL/Software.htm>). The frequency-decomposed signals undergo a series of nonlinear filters reflecting the computations taking place in the auditory nerve and other subcortical nuclei. We then reduced the number of channels from 128 to 32 by averaging the signal of each group of 4 consecutive channels, and used these 32 channels as input to the network. Each channel

projected onto a distinct  $Ge$  cell (i.e. specific connections,  $\omega_{ci} = 0.25\delta(c,i)$ ). As for  $Te$  input, each channel was convolved by the temporal filter and projected to all  $Te$  cells (all-to-all connections). Such a convolution can be implemented by a population of relay neurons that transmit their input with a certain delay, here between 0 and 50 ms.

Phoneme identity and boundaries have been labelled by phoneticians in every sentence of the corpus. We used the Tsylb2 program (Fisher 1996) that automatically syllabifies phonetic transcriptions (Kahn 1976) to merge these sequences of phonemes into sequences of syllables according to English grammar rules and thus get a timing for syllable boundaries.

To address the resilience of the model to speech compression we produced compressed sentences by applying a pitch-synchronous, overlap and add (PSOLA) procedure implemented by PRAAT, a speech analysis and modification software (<http://www.fon.hum.uva.nl/praat/>). The procedure retains all spectral properties from the original speech data in the compressed process. The same precortical filters were then applied as for uncompressed data before feeding into the network.

### Syllable boundary prediction algorithms

Syllable boundaries triggered average (STAs) were computed as follow: for each syllable boundary (syllable onsets excluding the first of each sentence), we extracted a 700 ms window of the corresponding locked to the syllable boundary and averaged over all syllable boundaries. STAs were computed for speech envelope and for each channel of the Chi et al. (Chi et al., 2005) model.

### Predictive models

We compared the performance of four distinct families of models to predict the timing of syllable boundaries based on speech envelope or speech audiogram: the Mermelstein algorithm (see below), a Linear Non-linear (LN) model (a simplified integration-to-threshold algorithm, see below), the entrained theta neural oscillator (see below) and a purely rhythmic control model (see below).

#### 1. Mermelstein algorithm

The Mermelstein algorithm is a standard algorithm that predicts syllable boundaries by identifying troughs in the power of the speech signal (Mermelstein, 1975; Villing et al.,

2004). The predicted boundaries are computed according to the following steps. First, extract the power of speech signal in the 500 – 4000 Hz range (grossly corresponding to formants) and low-pass filter at 40 Hz to remove fast fluctuations, defining a so-called *loudness function*. Second, for each sentence, compute the convex hull of the loudness signal and extract the maximum of the difference between the loudness signal and its convex hull. If that difference exceeds a certain threshold  $T_{min}$  and if the peak intensity of the interval of no more than  $P_{max}$  smaller than the peak intensity of the whole sentence, then that time of maximal difference is defined as a predicted boundary and the same procedure is applied recursively to the intervals to the left and right of that boundary. Parameters  $T_{min}$  and  $P_{max}$  were optimized to yield minimum prediction distance (see below), yielding  $T_{min} = 0.152 \text{ dB}$  and  $P_{max} = 15.85 \text{ dB}$ .

Note that this algorithm cannot be run *online* since the convex hull at a given time depends on the future value of speech power. Thus syllable boundaries can only be predicted after a certain delay, which makes it impractical for online speech comprehension as occurring in the human brain.

## 2. LN model and variations

To evaluate the capacity of a simplified neural system to predict syllable boundaries, we trained a generalized linear point process model on the syllable dataset. The model (figure 2 – Figure Supplement 1D) does not incorporate full neural dynamics but simply comprises a linear stimulus kernel followed by nonlinear function. The process issues a ‘spike’ or ‘syllable boundary signal’ whenever the output reaches a certain threshold (Pillow et al., 2008). This signal is fed back into the nonlinear function (another kernel  $Ih$  is used here): such negative feedback loop implements a relative refractory period. This model is a generalization of the Linear-Nonlinear Poisson model, hence we refer to it simply as *LN* model. We used the 32 auditory channels as input to the model and trained it to maximize its syllable boundary prediction performance.

We looked for a linear filter that is separable in its temporal and spectral component. We first computed the Spike Triggered Average (or rather ‘Syllable Boundary Triggered Average’) for all 32 channels from 600 ms to 0 ms prior to the actual boundary in 10 ms time steps. Yet STA provides the optimal estimate for the linear kernel in a LN model only when stimulus consists of uncorrelated white noise (Chichilnisky, 2001). To get the

optimal values out of the white noise condition, we looked at the separable filter  $H$  that yields best prediction of the output, i.e.  $\min H (< |Y(t) - \hat{Y}(t|H)|^2 >)$ , where:

- $Y(t)$  is a binary output equal to 1 if there is a syllabic boundary in the 10 ms interval, 0 otherwise
- $H$  is a separable spectrotemporal filter (i.e.  $H(\omega, u) = S(\omega)T(u)$  for all orders  $u$  and all frequencies  $\omega$ .  $S$  and  $T$  are, respectively, the spectral and temporal component of filter  $H$ .
- $\hat{Y}(t|H) = \sum_{u,w} H(w, u)X(\omega, t - u)$  where  $X(\omega, t)$  is the value of auditory channel  $\omega$  at time step  $t$ .

Optimal solutions of the system verify:

$$\begin{aligned}\sum_u T(u)R(\omega, u) &= \sum_{u,v,\xi} S(\xi)T(u)T(v)M(\omega, \xi, u, v) \quad \forall \omega \\ \sum_\omega S(\omega)R(\omega, u) &= \sum_{v,\omega,\xi} S(\omega)S(\xi)T(v)M(\omega, \xi, u, v) \quad \forall u\end{aligned}$$

where  $R(\omega, u) = \langle Y(t)X(\omega, t) \rangle_t$  (i.e.  $R$  is the Spike Triggered Average)

and  $M$  is the covariance tensor for  $X$ , ie.  $M(\omega, \xi, u, v) = \text{cov}(X(\omega, t - u), X(\xi, t - v))$ .

Solutions to  $T$  and  $S$  for that system of equations can be approximated numerically using the following iterative procedure:

$$\begin{aligned}S_0(\omega) &= 1 \quad \forall \omega, T_0(u) = 1 \quad \forall u \\ S_{n+1} &= \left( \frac{T_0 R}{\sum_{u,v} T_n(u)T_n(v)M(u, v, \dots)} \right)^T \\ T_{n+1} &= \left( \frac{RS_0}{\sum_{\omega,\xi} S_{n+1}(\omega)S_{n+1}(\xi)M(\omega, \xi, \dots)} \right)\end{aligned}$$

and then stopping when the resulting square error  $\|RS_0 - \sum_{\omega,\xi} S_{n+1}(\omega)S_{n+1}(\xi)T_n(v)M(\omega, \xi, v)\|_u^2$  goes below a minimum value (we used a threshold of  $10^{-1}$ ). The first 6 components (i.e. time bins) of the temporal kernel (i.e. 0 – 50 ms) were also used for input convolution in the theta model. We did not integrate further components (60 – 400 ms) since their weight was much lower and its implementation by relay neurons seemed less realistic.

To retrieve the optimal value for all parameters of the model, we used the GLM matlab toolbox developed in the Pillow lab ([http://pillowlab.cps.utexas.edu/code\\_GLM.html](http://pillowlab.cps.utexas.edu/code_GLM.html)), using as input the one-dimensional signal  $U(t) = \sum_\omega S(\omega)X(\omega, t)$ . Other parameters of the *LN* model including the self-inhibition temporal kernel  $Ih$  were optimized using the gradient descent implemented in the toolbox. This method provides estimation for a stochastic generalized *LN* model. We were interested in assessing the performance of a deterministic *LN* model. We then run a deterministic model with the same parameters as

the stochastic model plus one new free parameter describing the normalized time to next spike (in the stochastic model, that time is drawn from an exponential distribution). The value of  $t_{sp}^{next}$  was optimized using the same minimization procedure used for others models (see below). Two other parameters were also optimized again, since this procedure minimized a different score than the GLM toolbox score: time scale of self-inhibition  $\tau_{th}$  and constant input to the model  $DC$  (Table 2).

We made one last modification to this *LN* model. We optimized the model such that it would maximally fire not at the time of syllable boundaries but 10 ms posterior to that time (*de facto*, we simply slid the STA window by 10 ms). This provides a delayed signal but likely more reliable since it can use more information (notably the rebound in the auditory spectrogram that is present right after a syllable boundary).

### *3. Theta model*

The theta model is composed of the  $Te$  and  $Ti$  cells from the full network model described above, with the exact same parameter set. 11 parameters were optimized in the full model, 10 in the control model (see values in Table 3).

### *4. Control model*

The control model was used to provide a baseline for assessing the performance of other models. Under these control conditions, predicted syllable boundaries were generated rhythmically at a fixed time interval, irrespective of the stimulus. The rate of the rhythmic process was varied from 1 Hz to 15 Hz in 0.5 Hz intervals. Such control model yielded better performance than another control model consisting of a homogeneous Poisson process. It thus provides a more stringent control for estimating the efficiency of other algorithms.

### *Model performance evaluation*

We evaluated how well syllable boundaries predicted by any model matched with the boundaries derived from labelled speech data. As an evaluation metrics we used a point process distance that is used to compare distance between spike trains (Victor and Purpura, 1997). Shift cost was set to  $20 \text{ sec}^{-1}$  (in other words, a predicted and an actual boundary could be matched if they were no more than 50 msec apart).

To draw comparison between different models, for each level of compression, we computed the (non-normalized) distance measure for the theta model summed over all sentences in the test dataset, as well as the average number of predicted boundaries per sentence. We then matched the theta model to a control rhythmic model with the same predicted syllabic rate, and computed the difference between the non-normalized distance for the theta model and for that matched rhythmic model.

#### *Optimisation*

We optimized the parameters from all models to get the minimal normalized point process distance between predicted and actual boundaries in each sentence. Optimization was made using global gradient descent (function *fminsearch* in Matlab), and repeated with many initial points to avoid retaining a local minimum. Although both the theta model and the control model are intrinsically stochastic, the sample size was large enough for the objective function over the entire sample to be nearly deterministic, allowing for convergence of the gradient descent algorithm. The list of optimized parameters for each type of model is provided in the related model sections above. We split the entire TIMIT TRAIN dataset (4620 sentences) into two datasets: a first dataset of 1000 sentences was used to compute optimal parameters; final assessment of an algorithm performance with its optimal parameters was done on a separate set of 3620 sentences.

#### **Analysis of model behaviour**

##### *LFP Spectral analysis*

Simulated LFP was downsampled to 1000 Hz before applying a time-frequency decomposition using complex Morlet wavelet transform, with all frequencies between 2 and 100 Hz with a 0.5 Hz precision. Coherence between stimulus and LFP signal was then computed for each time point  $t$  and each frequency  $f$  over 100 simulations using 100 distinct sentences  $sen$ , using the formula from (Mitra and Pesaran 1999). Synchronized bursts of the PING or PINTH were detected using spike timings in  $Gi$  and  $Ti$  populations since spikes of inhibitory neurons were more synchronized than those of excitatory neurons. Synchronous bursts of spikes were detected within a given population whenever more than 10% of neurons in the population spikes within a 6 ms interval (15 ms for  $Ti$  cells).

##### *Cross-frequency coupling*

We computed cross-frequency coupling from 50 simulations of the model, each with a different TIMIT sentence preceded by 1000-1500 ms rest.

For the LFP phase-amplitude coupling, we extracted phase and amplitude from all frequencies from 2 Hz to 70 Hz in 1 Hz interval, and computed the Modulation Index (MI) for all pairs of frequencies (Tort et al. 2010). Data from all trials were concatenated (separately for spontaneous and speech-related activity) across all trials beforehand. To compute MI, in each condition, signal amplitude values  $x(f_{amp}, t, sen)$  were binned in  $N = 18$  different bins according to the simultaneous phase of  $x(f_{phase}, t, sen)$ . For spike phase-amplitude coupling, we defined spike gamma amplitude as the number of  $Gi$  neurons spiking at a given gamma burst, and the spike theta phase was defined by linear interpolation from  $-\pi$  for a theta spike burst to  $+\pi$  for the subsequent theta burst.

### Simple temporal patterns decoding

We first explored the model's performance using simple sawtooth signals (Shamir et al. 2009), representing prototypical realizations of formant transitions in a given frequency band. Each stimulus consisted of a rising component between 0 and 1, followed by a decay component from 1 back to 0. The overall length of the sawtooth was 50 ms, and the relative position of the maximal point  $t_{MAX}$  between the starting point  $t_{START}$  and end point  $t_{END}$  was defined by a variable  $\alpha = (t_{MAX}-t_{START})/(t_{END}-t_{START})$

The input connectivity had to be slightly modified since sawtooths are one-dimensional signals in contrast to the multi-dimensional channel signals that we have to use for speech stimuli: for  $Te$  units, we used  $I_{Te}^{EXT}=20$ ; and for the connections to  $Ge$  units in line with the original model (Shamir et al. 2009), we used different input levels across the population, ranging from 0.125 to 4 in 0.125 intervals. The rest of the model remained unchanged.

We simulated the response of the network to a series of 500 sawtooths with parameter  $\alpha$  taking one of 10 equally spaced values within the [0 1] interval. Interstimulus interval (ISI) varied randomly between 50 and 250 ms.

We compared the model's performance for different neural codes. For the “stimulus timing” code (see Results section), we extracted the spike pattern of output ( $Ge$ ) neurons between 20 ms before and 70 ms after of each sawtooth onset. We computed the distance between all output spike patterns using a spike train distance measure (Victor and Purpura 1997), implemented in the Spike Train Analysis Toolkit (<http://neuroanalysis.org/toolkit/>). We used a shift cost of 200 s<sup>-1</sup> corresponding to a timing resolution of 5 ms. We decoded the peak parameter using the simple leave-one-out clustering procedure of the STA toolkit,

using a clustering exponent of -10. By comparing the ‘decoded parameter’, i.e. the parameter corresponding to the closest cluster, to the input sawtooth parameter, we built confusion matrices and computed decoding performance.

In the “theta-timing” code, we extracted the spike pattern of output neuron in windows starting 20 before a theta burst and finishing 20 ms after the next theta burst (‘*theta chunks*’, Figure 4A). Spike times within each chunk were referenced with respect to the onset of the window. Each spike pattern was labelled with the corresponding value of the stimulus if the theta burst occurred during the presentation of the stimulus, or with the label ‘rest’ if the theta burst occurred during an interstimulus interval. The same decoding analysis was applied on such internally referenced neural patterns, yielding a 11 x 11 confusion matrix (10 stimulus shapes and rest). Detection theory measures (hits, misses, correct rejections and false alarms) were computed by summing values in blocks of the confusion matrix (of size 10 x 10, 10 x 1, 1 x 10 and 1 x 1 respectively). A classification confusion matrix was obtained by removing the last row and last column of that confusion matrix.

We run the same decoding analysis on variants of the network: the full network; a control model where  $T_\theta$  units do not receive the sawtooth input (*undriven theta network*) and another control where theta-gamma connections were removed (*uncoupled theta-gamma network*).

#### Syllable decoding from sentences

The classification procedure was similar for syllable decoding, where we tried to decode the identity of syllables within continuous stream of speech (full sentences) from the activity of output neurons. We stimulated the network by presenting 25 sentences from the TIMIT corpus repeated 100 times each. We extracted theta chunks of  $G_e$  spike patterns as explained previously. Each chunk was labelled with the identity of the syllable being presented at the time of the first theta burst of the chunk. We randomly selected 10 syllables from the whole set of syllables within the 25 sentences. As in some cases there were several consecutive theta chunks corresponding to the same syllable, we equated the total number of theta chunks per syllable by randomly selecting 100 theta chunks labelled with each of the 10 syllables. Syllable classification of theta-chunked  $G_e$  spike patterns was performed using two different neural codes. For the *spike pattern code*, we applied the same procedure as for sawtooth classification, using a smaller value of spike shift cost corresponding to a timing resolution of 60 ms. For the *spike count code*, we measured the

number of spikes emitted by each *Ge* neuron within a theta chunk. We then ran a simple nearest mean classification procedure to decode syllable identity corresponding to each theta chunk from the spike counts of all *Ge* neurons (see *Classification analysis* below). Both methods relied on the leave-one-out procedure that consists in identifying a chunk after the decoder was trained on all chunks but the to-be-decoded one. Decoding was repeated 200 times using each time a different set of 10 random syllables, and the analysis was performed over all three variants of the network.

For syllable classification across speakers, we used the two sentences from the TIMIT corpus that have been recorded for each of the 462 speakers ('*She had your dark suit in greasy wash water all year*' and '*Don't ask me to carry an oily rag like that*') and trained the network to classify syllables based on the neural output from other speakers, thus testing generalization across speakers. There is a wide variability of pronunciations over speakers as attested by the variability of chain of phonemes labelled of phoneticians, but the two sentences could nonetheless be parsed into 25 syllables overall for each speaker. We simulated the network presenting these 924 sentences and used the theta-chunked output to decode syllable identity. The method used was very similar to the syllable decoding analysis, where we classified theta-chunked neural patterns into one of 10 possible syllables (drawn randomly from the set of 25 syllables), with the only difference that here the classifier was based on theta chunks coming from different speakers. The classification was repeated 100 times for different subsets of syllables.

#### **Neural encoding properties: classification analysis**

The first analysis of neural encoding properties consisted in comparing the ability to classify neural codes from the model into arbitrary speech segments (as opposed to syllables as in previous section). The methods, as detailed below, were inspired by the decoding of neural auditory cortical activity recorded in monkeys in response to naturalistic sounds (Kayser, Ince, and Panzeri 2012). We simulated the network by presenting 25 different sentences from the TIMIT corpus repeated 50 times each. For a given window size (ranging from 80 to 480 ms in 80 ms intervals), we randomly extracted 10 windows (defined as *stimuli*) from the overall set of 25 sentences. We then retrieved stimulus identity based on the activity of a neuron that was randomly drawn from the *Ge* population using three different neural codes. In the *neural count code*, we counted the number of spikes emitted by that neuron within each window. In the *time-partitioned code*,

we divided each window into  $N$  equally size bins, and computed the number of spikes for each of the 8 bins separately. In the *phase-partitioned code*, we divided the window based on theta-phase- rather time- intervals: each spike was labelled with the phase of the theta oscillation at the corresponding spike time, and we computed the number of spikes falling into each of the  $N$  subdivisions of the  $[-\pi; \pi]$  interval.

We then used a nearest-mean template matching procedure to decode the stimuli. To classify each stimulus exemplar using each neural code, we averaged the vectors over all presentations of each stimulus using a leave-one-out procedure; we then computed the Euclidian distance from the current vector to each of the 10 stimulus-averaged template. Finally we ‘decoded’ the neural code by assigning it to the stimulus class with minimal distance to template. A more detailed explanation of the procedure is provided in the original experiment article (Kayser, Ince, and Panzeri 2012). The procedure was repeated 1000 times, each time with a different set of 10 random stimuli, and performed the 3 variants of network.

#### **Neural encoding properties: Mutual Information analysis**

We complemented the stimulus classification with a similar analysis using mutual information between the acoustic ‘stimulus’ and response of individual *Ge* neurons to further characterize the encoding properties of the network. Mutual Information ( $MI$ ) estimates the reduction of uncertainty about the acoustic ‘*stimulus*’ that is obtained from the knowledge of a single trial of neural response. The dataset was identical to the one previously used for stimulus classification analysis, where each stimulus was again segmented into non-overlapping windows of length  $T$  (here 8 to 48 ms) (Kayser et al. 2009; de Ruyter van Steveninck, Lewen, and Strong 1997).

Mutual Information was computed for the same neural codes as in Kayser et al. (2009). We used *Spike count code* and *Time-partitioned code* as described above (for the Time-partitioned code the size of the bins was kept constant to 8 bins; the number of bins in a window hence increased with window size. As slow LFP phase was more reliable over sentence repetitions than power, we combined spike count and LFP theta phase to get a *Spike count & Phase-partitioned code* (Montemurro et al. 2008). For this code, the phase of slow LFP was divided into  $N=4$  bins, and the firing rate in each window was labelled according to the phase at which the first spike occurred. Finally we explored the influence of slow LFP phase on MI when combined with temporal spiking patterns. Thus, in the *Time- & Phase-partitioned code* spikes carry two distinct tags, the first one referring to the

position of the spike inside one of the four subdivisions of the stimulus window, the second indicating the phase of the underlying LFP at the moment of the spike occurrence. We corrected for sampling bias (Kayser et al. 2009) first by using a *shuffling* method (Panzeri et al. 2007), then the quadratic extrapolation method (Strong et al. 1998). We further reduced the residual bias using a bootstrapping technique (200 resampled data) (Montemurro et al. 2008).

### Acknowledgements

This work was funded by the European Research Council (Compuslang project; Grant agreement 260347), the Swiss National Fund (grant 320030-149319), the Agence National de la Recherche, the CNRS. We warmly thank Oded Ghitza for stimulating discussions, Maoz Shamir and Andy Brughera for sharing elements of code with us, Adrien Wohrer for help with the mathematical analysis and Jean-Paul Haton for his input from the perspective of automatic speech recognition. BSG gratefully acknowledges partial support from the National Research University Higher School of Economics.

### References

- Abrams, DA, T Nicol, S Zecker, and N Kraus. 2008. "Right-Hemisphere Auditory Cortex Is Dominant for Coding Syllable Patterns in Speech." *The Journal of Neuroscience* 28 (15): 3958–3965.
- Ahissar, Ehud, Srikanth S Nagarajan, Merav Ahissar, Athanassios Protopapas, H Mahncke, and M M Merzenich. 2001. "Speech Comprehension Is Correlated with Temporal Response Patterns Recorded from Auditory Cortex." *Proceedings of the National Academy of Sciences of the United States of America* 98 (23) (November): 13367–72. doi:10.1073/pnas.201400998.
- Ainsworth, Matthew, Shane Lee, Mark O. Cunningham, Anita K Roopun, Roger D. Traub, Nancy J. Kopell, and Miles A. Whittington. 2011. "Dual Gamma Rhythm Generators Control Interlaminar Synchrony in Auditory Cortex." *Journal of Neuroscience* 31 (47): 17040–17051. doi:10.1523/JNEUROSCI.2209-11.2011.
- Arnal, Luc Henri, and Anne-Lise Giraud. 2012. "Cortical Oscillations and Sensory Predictions." *Trends in Cognitive Sciences* (June): 1–9. doi:10.1016/j.tics.2012.05.003.
- Arnal, Luc Henri, Benjamin Morillon, Christian Alexander Kell, and Anne-Lise Giraud. 2009. "Dual Neural Routing of Visual Facilitation in Speech Processing." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 29 (43) (October 28): 13445–53. doi:10.1523/JNEUROSCI.3194-09.2009.
- Bastos, Andre M., W. Martin Usrey, Rick A. Adams, George R. Mangun, Pascal Fries, and Karl J. Friston. 2012. "Canonical Microcircuits for Predictive Coding." *Neuron* 76 (4) (November): 695–711. doi:10.1016/j.neuron.2012.10.038.
- Börgers, Christoph, and Nancy J. Kopell. 2005. "Effects of Noisy Drive on Rhythms in Networks of Excitatory and Inhibitory Neurons." *Neural Computation* 17 (3): 557–608.

- Brasselet, R., Stefano Panzeri, Nikos K. Logothetis, and Christoph Kayser. 2012. "Neurons with Stereotyped and Rapid Responses Provide a Reference Frame for Relative Temporal Coding in Primate Auditory Cortex." *Journal of Neuroscience* 32 (9) (February 29): 2998–3008. doi:10.1523/JNEUROSCI.5435-11.2012.
- Brunel, Nicolas, and Xiao-jing Wang. 2003. "What Determines the Frequency of Fast Network Oscillations with Irregular Neural Discharges? I. Synaptic Dynamics and Excitation-Inhibition Balance." *Journal of Neurophysiology* 90 (1): 415.
- Canolty, Ryan T., Maryam Soltani, Sarang S Dalal, Erik Edwards, Nina F Dronkers, Srikanth S Nagarajan, Heidi E Kirsch, Nicholas M Barbaro, and Robert T. Knight. 2007. "Spatiotemporal Dynamics of Word Processing in the Human Brain." *Frontiers in Neuroscience* 1 (1) (November): 185–96. doi:10.3389/neuro.01.1.1.014.2007.
- Chi, Taiyih, Powen Ru, and Shihab A Shamma. 2005. "Multiresolution Spectrotemporal Analysis of Complex Sounds." *The Journal of Acoustical Society of America* 118 (2) (August): 887–906.
- Cogan, Gregory Brendan, and David Poeppel. 2011. "A Mutual Information Analysis of Neural Coding of Speech by Low Frequency MEG Phase Information." *Journal of Neurophysiology* (May 11). doi:10.1152/jn.00075.2011.
- Consortium, Linguistic Data. 1993. "TIMIT Acoustic-Phonetic Continuous Speech Corpus". Philadelphia.
- Da Costa, Nuno Maçarico, and Kevan a C Martin. 2010. "Whose Cortical Column Would That Be?" *Frontiers in Neuroanatomy* 4 (May) (January): 16. doi:10.3389/fnana.2010.00016.
- Davis, Matthew H., Michael A Ford, Ferath Kherif, and Ingrid S. Johnsrude. 2011. "Does Semantic Context Benefit Speech Understanding through 'Top-down' Processes? Evidence from Time-Resolved Sparse fMRI." *Journal of Cognitive Neuroscience* 23 (12) (December): 3914–32. doi:10.1162/jocn\_a\_00084.
- De Ruyter van Steveninck, Erik, GD Lewen, and SP Strong. 1997. "Reproducibility and Variability in Neural Spike Trains." *Science*.
- Douglas, Rodney J, and Kevan a C Martin. 2004. "Neuronal Circuits of the Neocortex." *Annual Review of Neuroscience* 27 (January): 419–51. doi:10.1146/annurev.neuro.27.070203.144152.
- Fanselow, EE, Kristen A. Richardson, and Barry W Connors. 2008. "Selective, State-Dependent Activation of Somatostatin-Expressing Inhibitory Interneurons in Mouse Neocortex." *Journal of Neurophysiology* 100: 2640–2652. doi:10.1152/jn.90691.2008.
- Fishbach, A, Israel Nelken, and Yeshurun. 2001. "Auditory Edge Detection: A Neural Model for Physiological and Psychoacoustical Responses to Amplitude Transients." *Journal of Neurophysiology* 85 (6) (June): 2303–23.
- Fisher, W. M. 1996. "tsylb2." *National Institute of Standards and Technology*. <http://www.nist.gov/speech/tools>.
- Fontolan, Lorenzo, Benjamin Morillon, Catherine Liégeois-Chauvel, Anne-Lise Giraud, and C. Liegeois-Chauvel. 2014. "The Contribution of Frequency-Specific Activity to Hierarchical Information Processing in Human Auditory Cortices." *Nature Communications* 5 (May) (September 2): 4694. doi:10.1038/ncomms5694.
- Gagnepain, Pierre, Richard N. Henson, and Matthew H. Davis. 2012. "Temporal Predictive Codes for Spoken Words in Auditory Cortex." *Current Biology : CB* 22 (7) (April 10): 615–21. doi:10.1016/j.cub.2012.02.015.
- Ghitza, Oded. 2011. "Linking Speech Perception and Neurophysiology: Speech Decoding Guided by Cascaded Oscillators Locked to the Input Rhythm." *Frontiers in Psychology* 2 (June): 1–13. doi:10.3389/fpsyg.2011.00130.
- . 2014. "Behavioral Evidence for the Role of Cortical  $\Theta$  Oscillations in Determining Auditory Channel Capacity for Speech." *Frontiers in Psychology* 5. doi:10.3389/fpsyg.2014.00652.
- Giraud, Anne-Lise, and David Poeppel. 2012. "Cortical Oscillations and Speech Processing: Emerging Computational Principles and Operations." *Nature Neuroscience* 15 (4) (March 18): 511–517. doi:10.1038/nn.3063.
- Gross, Joachim, Nienke Hoogenboom, Gregor Thut, Philippe Schyns, Stefano Panzeri, Pascal Belin, and Simon Garrod. 2013. "Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain." *PLoS Biology* 11 (12) (December): e1001752. doi:10.1371/journal.pbio.1001752.
- Gupta, Anoopum S, Matthijs a van der Meer, David S Touretzky, and a David Redish. 2012. "Segmentation of Spatial Experience by Hippocampal Theta Sequences." *Nature Neuroscience* 15 (7) (June 17): 1032–1039. doi:10.1038/nn.3138.
- Gütig, Robert, and Haim Sompolinsky. 2009. "Time-Warp-Invariant Neuronal Processing." *PLoS Biology* 7 (7) (July): e1000141. doi:10.1371/journal.pbio.1000141.
- Henry, M. J., B. Hermann, and J. Obleser. 2014. "Entrained Neural Oscillations in Multiple Frequency Bands Comodulate Behavior." *Proceedings of the National Academy of Sciences* 111 (41): 14935–14940. doi:10.1073/pnas.1408741111.

- Hopfield, J.J. 2004. "Encoding for Computation: Recognizing Brief Dynamical Patterns by Exploiting Effects of Weak Rhythms on Action-Potential Timing." *Proceedings of the National Academy of Sciences of the United States of America* 101 (16) (April): 6255–60. doi:10.1073/pnas.0401125101.
- Jadi, MP, and TJ Sejnowski. 2014. "Cortical Oscillations Arise from Contextual Interactions That Regulate Sparse Coding." *Proceedings of the National ...* doi:10.1073/pnas.1405300111/-DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1405300111.
- Jensen, O, and JE Lisman. 1996. "Theta/gamma Networks with Slow NMDA Channels Learn Sequences and Encode Episodic Memory: Role of NMDA Channels in Recall." *Learning & Memory* 3 (2-3) (January 1): 264–278. doi:10.1101/lm.3.2-3.264.
- Jensen, Ole, and Laura Lee Colgin. 2007. "Cross-Frequency Coupling between Neuronal Oscillations." *Trends in Cognitive Sciences* 11 (7) (July): 267–9. doi:10.1016/j.tics.2007.05.003.
- Kahn, D. 1976. *Syllable-Based Generalizations in English Phonology*.
- Kayser, Christoph, Robin A A Ince, and Stefano Panzeri. 2012. "Analysis of Slow (Theta) Oscillations as a Potential Temporal Reference Frame for Information Coding in Sensory Cortices." Edited by Tim Behrens. *PLoS Computational Biology* 8 (10) (October 11): e1002717. doi:10.1371/journal.pcbi.1002717.
- Kayser, Christoph, Marcelo a Montemurro, Nikos K. Logothetis, and Stefano Panzeri. 2009. "Spike-Phase Coding Boosts and Stabilizes Information Carried by Spatial and Temporal Spike Patterns." *Neuron* 61: 597–608.
- Kopell, Nancy J., Christoph Börgers, Dmitri D Pervouchine, and Paola Malerba. 2010. "Gamma and Theta Rhythms in Biophysical Models of Hippocampal Circuits." In *Hippocampal Microcircuits*, edited by V. Cutsuridis, 423–457. Springer.
- Lakatos, Péter, Chi-Ming Chen, Monica N O'Connell, and A Mills. 2007. "Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex." *Neuron* 53: 279–292.
- Lakatos, Péter Peter, AS Ankoor S Shah, KH Kevin H Knuth, István Ulbert, György George Karmos, and Charles E. E. Schroeder. 2005. "An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex." *Journal of ...* 94 (3) (September): 1904–11. doi:10.1152/jn.00263.2005.
- Lee, Jung H., Miles A. Whittington, and Nancy J. Kopell. 2013. "Top-Down Beta Rhythms Support Selective Attention via Interlaminar Interaction: A Model." Edited by Stephen Coombes. *PLoS Computational Biology* 9 (8) (August 8): e1003164. doi:10.1371/journal.pcbi.1003164.
- Lever, Colin, Raphael Kaplan, and Neil Burgess. 2014. *Space, Time and Memory in the Hippocampal Formation*. Edited by Dori Derdikman and James J. Knierim. Vienna: Springer Vienna. doi:10.1007/978-3-7091-1292-2.
- Lieberman, P. 1985. "On the Evolution of Human Syntactic Ability. Its Pre-Adaptive Bases—motor Control and Speech." *Journal of Human Evolution* 14 (7): 657–668.
- Lisman, John E., and György Buzsáki. 2008. "A Neural Coding Scheme Formed by the Combined Function of Gamma and Theta Oscillations." *Schizophrenia Bulletin* 34 (5) (September): 974–80. doi:10.1093/schbul/sbn060.
- Lisman, John E., and Ole Jensen. 2013. "The Theta-Gamma Neural Code." *Neuron* 77 (6) (March): 1002–1016. doi:10.1016/j.neuron.2013.03.007.
- Luo, Huan, Zuxiang Liu, and David Poeppel. 2010. "Auditory Cortex Tracks Both Auditory and Visual Stimulus Dynamics Using Low-Frequency Neuronal Phase Modulation." Edited by Robert Zatorre. *PLoS Biology* 8 (8) (August 10): e1000445. doi:10.1371/journal.pbio.1000445.
- Luo, Huan, and David Poeppel. 2007. "Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex." *Neuron* 54: 1001–1010.
- Mazzoni, Alberto, Stefano Panzeri, Nikos K. Logothetis, and Nicolas Brunel. 2008. "Encoding of Naturalistic Stimuli by Local Field Potential Spectra in Networks of Excitatory and Inhibitory Neurons." Edited by Karl J. Friston. *PLoS Computational Biology* 4 (12) (December): e1000239. doi:10.1371/journal.pcbi.1000239.
- Mermelstein, P. 1975. "Automatic Segmentation of Speech into Syllabic Units." *J. Acoust. Soc. Am* 58 (4): 880–883.
- Miller, J L, F Grosjean, and C Lomanto. 1984. "Articulation Rate and Its Variability in Spontaneous Speech: A Reanalysis and Some Implications." *Phonetica* 41: 215–225. doi:10.1159/000261728.
- Mitra, Partha P, and Bijan Pesaran. 1999. "Analysis of Dynamic Brain Imaging Data." *Biophysical Journal* 76 (February): 691–708.
- Montemurro, Marcelo a, MJ Rasch, Yusuke Murayama, Nikos K. Logothetis, and Stefano Panzeri. 2008. "Phase-of-Firing Coding of Natural Visual Stimuli in Primary Visual Cortex." *Current Biology* 18: 375–380.

- Morillon, Benjamin, Catherine Liégeois-Chauvel, Luc Henri Arnal, Christian-G. Bénar, and Anne-Lise Giraud. 2012. "Asymmetric Function of Theta and Gamma Activity in Syllable Processing: An Intra-Cortical Study." *Frontiers in Psychology* 3 (July): 1–9. doi:10.3389/fpsyg.2012.00248.
- Mukamel, EA Eran A, Kin Foon KF Wong, Michael J. Prerau, Emery N Brown, and Patrick L. Purdon. 2011. "Phase-Based Measures of Cross-Frequency Coupling in Brain Electrical Dynamics under General Anesthesia." ... in Medicine and ... 2011 (January): 1–10. doi:10.1109/IEMBS.2011.6090558.Phase-Based.
- Ngon, Céline, Andrew Martin, Emmanuel Dupoux, Dominique Cabrol, Michel Dutat, and Sharon Peperkamp. 2013. "(Non)words, (non)words, (non)words: Evidence for a Protolexicon during the First Year of Life." *Developmental Science* 16 (1) (January): 24–34. doi:10.1111/j.1467-7687.2012.01189.x.
- Nourski, Kirill V, Richard a Reale, Hiroyuki Oya, Hiroto Kawasaki, Christopher K Kovach, Haiming Chen, Matthew a Howard, and John F Brugge. 2009. "Temporal Envelope of Time-Compressed Speech Represented in the Human Auditory Cortex." *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 29 (49) (December 9): 15564–74. doi:10.1523/JNEUROSCI.3065-09.2009.
- Panzeri, Stefano, and Mathew E Diamond. 2010. "Information Carried by Population Spike Times in the Whisker Sensory Cortex Can Be Decoded Without Knowledge of Stimulus Time." *Frontiers in Synaptic Neuroscience* 2 (June) (January): 17. doi:10.3389/fnsyn.2010.00017.
- Panzeri, Stefano, Robin a Ince, Mathew E Diamond, and Christoph Kayser. 2014. "Reading Spike Timing without a Clock: Intrinsic Decoding of Spike Trains." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 369 (1637) (January): 20120467. doi:10.1098/rstb.2012.0467.
- Panzeri, Stefano, Riccardo Senatore, Marcelo a Montemurro, and Rasmus S Petersen. 2007. "Correcting for the Sampling Bias Problem in Spike Train Information Measures." *Journal of Neurophysiology* 98 (3) (September): 1064–72. doi:10.1152/jn.00559.2007.
- Pasley, Brian N., SV David, and Nima Mesgarani. 2012. "Reconstructing Speech from Human Auditory Cortex." Edited by Robert Zatorre. *PLoS Biology* 10 (1) (January 31): e1001251. doi:10.1371/journal.pbio.1001251.
- Peelle, Jonathan E, Joachim Gross, and Matthew H. Davis. 2013. "Phase-Locked Responses to Speech in Human Auditory Cortex Are Enhanced during Comprehension." *Cerebral Cortex* 23 (6): 1378–1387.
- Phillips, D P, S E Hall, and S E Boehnke. 2002. "Central Auditory Onset Responses, and Temporal Asymmetries in Auditory Perception." *Hearing Research* 167 (1-2) (May): 192–205.
- Poeppel, David. 2003. "The Analysis of Speech in Different Temporal Integration Windows: Cerebral Lateralization as 'Asymmetric Sampling in Time.'" *Speech Communication* 41 (1) (August): 245–255. doi:10.1016/S0167-6393(02)00107-3.
- Poeppel, David, William J Idsardi, and Virginie van Wassenhove. 2008. "Speech Perception at the Interface of Neurobiology and Linguistics." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363 (1493) (March 12): 1071–86. doi:10.1098/rstb.2007.2160.
- Rosen, Stuart. 1992. "Temporal Information in Speech : Acoustic, Auditory and Linguistic Aspects." *Philosophical Transactions: Biological Sciences* 336 (1278): 367–373.
- Schroeder, Charles E., and Péter Lakatos. 2009. "The Gamma Oscillation: Master or Slave?" *Brain Topography* 22 (1): 24–26. doi:10.1007/s10548-009-0080-y.The.
- Shamir, Maoz, Oded Ghitza, Steven Epstein, and Nancy J. Kopell. 2009. "Representation of Time-Varying Stimuli by a Network Exhibiting Oscillations on a Faster Time Scale." *PLoS Computational Biology* 5 (5): e1000370. doi:10.1371/journal.pcbi.1000370.
- Silberberg, Gilad, and Henry Markram. 2007. "Disynaptic Inhibition between Neocortical Pyramidal Cells Mediated by Martinotti Cells." *Neuron* 53 (5) (March 1): 735–46. doi:10.1016/j.neuron.2007.02.012.
- Stevens, Kenneth Noble. 2002. "Toward a Model for Lexical Access Based on Acoustic Landmarks and Distinctive Features." *The Journal of the Acoustical Society of America* 111 (4): 1872. doi:10.1121/1.1458026.
- Strong, SP, Roland Koberle, RRR van Steveninck, and William Bialek. 1998. "Entropy and Information in Neural Spike Trains." *Physical Review Letters* 80 (1): 197–200.
- Tort, Adriano B L, Robert W Komorowski, Howard Eichenbaum, and Nancy J. Kopell. 2010. "Measuring Phase-Amplitude Coupling between Neuronal Oscillations of Different Frequencies." *Journal of Neurophysiology* 104 (2) (May): 1195–210. doi:10.1152/jn.00106.2010.
- Tort, Adriano B L, Horacio G Rotstein, Tamar Dugladze, Tengis Gloveli, and Nancy J. Kopell. 2007. "On the Formation of Gamma-Coherent Cell Assemblies by Oriens Lacunosum-Moleculare Interneurons in the Hippocampus." *PNAS* 104 (33): 13490–13495.
- Victor, Jonathan, and Keith Purpura. 1997. "Metric-Space Analysis of Spike Trains: Theory, Algorithms and Application." *Network: Computation in Neural Systems* 8 (2) (May): 127–164. doi:10.1088/0954-898X/8/2/003.

- Vierling-Claassen, Dorec, Jessica a Cardin, Christopher I Moore, and Stephanie R Jones. 2010. "Computational Modeling of Distinct Neocortical Oscillations Driven by Cell-Type Selective Optogenetic Drive: Separable Resonant Circuits Controlled by Low-Threshold Spiking and Fast-Spiking Interneurons." *Frontiers in Human Neuroscience* 4 (November) (January): 198. doi:10.3389/fnhum.2010.00198.
- Wang, Xiao-jing. 2010. "Neurophysiological and Computational Principles of Cortical Rhythms in Cognition." *Physiological Reviews*. doi:10.1152/physrev.00035.2008.
- Wang, Xiaoqin, T Lu, Daniel Bendor, and E Bartlett. 2008. "Neural Coding of Temporal Information in Auditory Thalamus and Cortex." *Neuroscience* 157 (2) (November): 484–493. doi:10.1016/j.neuroscience.2008.07.050.
- Wu, S.L., M.L. Shire, Steven Greenberg, and Nelson Morgan. 1997. "Integrating Syllable Boundary Information into Speech Recognition." In *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, 2:987–990. IEEE.
- Yildiz, Izzet B., Katharina von Kriegstein, and Stefan J. Kiebel. 2013. "From Birdsong to Human Speech Recognition: Bayesian Inference on a Hierarchy of Nonlinear Dynamical Systems." Edited by Viktor K. Jirsa. *PLoS Computational Biology* 9 (9) (September 12): e1003219. doi:10.1371/journal.pcbi.1003219.
- Zhou, Y., and Xiaoqin Wang. 2010. "Cortical Processing of Dynamic Sound Envelope Transitions." *Journal of Neuroscience* 30 (49) (December): 16741–16754. doi:10.1523/JNEUROSCI.2016-10.2010.
- Zion Golumbic, Elana M., David Poeppel, and Charles E. Schroeder. 2012. "Temporal Context in Speech Processing and Attentional Stream Selection: A Behavioral and Neural Perspective." *Brain and Language* (January 26). doi:10.1016/j.bandl.2011.12.010.

## Figure legends

### Figure 1. Network architecture and dynamics

- A. Architecture of the full model.**  $T_e$  excitatory neurons ( $n=10$ ) and  $T_i$  inhibitory neurons ( $n=10$ ) form the PINTH loop generating theta oscillations.  $G_e$  excitatory neurons ( $n=32$ ) and  $G_i$  inhibitory neurons ( $n=32$ ) form the PING loop generating gamma oscillations.  $T_e$  neurons receive non-specific projections from all auditory channels, while  $G_e$  units receive specific projection from a single auditory channel, preserving tonotopy in the  $G_e$  population. PING and PINTH loops are coupled through all-to-all projections from  $T_e$  to  $G_e$  units.
- B. Network activity at rest and during speech perception.** Raster plot of spikes from representative  $T_i$  (dark green),  $T_e$  (light green),  $G_i$  (dark blue) and  $G_e$  (light blue). Simulated LFP is shown on top and the auditory spectrogram of the input sentence "*Ralph prepared red snapper with fresh lemon sauce for dinner*" is shown below.  $G_e$  spikes relative to theta burst (red boxes) form the output of the network. Gamma synchrony is visible in  $G_i$  spikes.
- C. Evoked potential (ERP) and Post-stimulus time histograms (PSTH)** of  $T_e$  and  $G_e$  population from 50 simulations of the same sentence: ERP (i.e., simulated LFP averaged over simulations, black line), acoustic envelope of the sentence (red line, filtered at 20 Hz), PSTH for theta (green line) and gamma (blue line) neurons. Vertical bars show scale of 10 spikes for both PSTH. The theta network phase-locks to speech slow fluctuations, and entrains the gamma network through the theta-gamma connection.
- D. Theta/gamma phase amplitude coupling in  $G_e$  spiking activity.** Top panel: LFP gamma envelope follows LFP theta phase in single trials. Bottom-Left panel: LFP phase-amplitude coupling (measured by Modulation Index, MI) for pairs of frequencies during rest, showing peak in theta-gamma pairs. Bottom-right panel: MI phase-amplitude coupling at the spiking level for the intact model and a control model with no theta-gamma connection (red arrow on A panel), during rest (blue bars) and speech presentation (brown bars).

Figure 1 – Figure Supplement 1: Spectral analysis

- A. **Theta phase pattern** (left panels) and **theta power pattern** (right panels) for 50 presentations of the same sentence in the uncoupled theta-gamma control model (top panels) and intact panels (bottom panels). Phase/power is binned into 4 different bins and colour coded. Theta phase is much more reliably imprinted by speech stimulus than power.
- B. (Left panel) **Spike phase-amplitude coupling:** mean value for PING amplitude (defined as the number of  $G_i$  neurons spiking within a gamma burst) as a function of PINTH phase (defined from interpolation between successive theta bursts). Intact model is shown in black while the uncoupled theta-gamma model is shown in blue. Data for rest (thick dashed lines) and during processing of speech (full thick lines) almost perfectly match. Thin dashed lines represent s.e.m. Spike PAC was very strong in the full model but quasi-absent when the theta-gamma connection was removed. (Right panel) **Spontaneous spike phase-frequency coupling:** mean value for PING frequency (defined from the duration between successive gamma bursts) as a function of PINTH phase. Same legend as left panel. Spike PFC is strong when and only when the theta-gamma connection is present (significant coupling  $p < 10^{-9}$  for both speech and rest).
- C. **Phase-locking of the theta and gamma oscillations to speech.** Phase concentration of the filtered LFP theta (top panel) and gamma (bottom panel) signals through time for 200 presentations of the same sentence (same as Figure 1B-C). The horizontal orange bar indicates the presentation of the sentence. There is a rapid transition from uniform theta distribution before sentence onset to perfectly phase-locked theta. Phase-locking vanishes at the end of sentence presentation.

**Figure 2. Theta entrainment by syllabic structure**

- A. **A. Theta spikes align to syllable boundaries.** Top graph shows the activity of the theta network at rest and in response to a sentence, including the LFP traces displaying strong theta oscillations, and raster plots for spikes in the  $T_i$  (light green) and  $T_e$  (dark green) populations. Theta bursts align well to the syllable boundaries obtained from labelled data (vertical black lines shown on top of auditory spectrogram in graph below).
- B. **Performance of different algorithms in predicting syllable onsets:** Syllable alignment score indexes how well theta bursts aligned onto syllable boundaries for each sentence in the corpus, and the score was averaged over the 3620 sentences in the test

dataset (error bars: standard error). Results compare Mermelstein algorithm (grey bar), linear-nonlinear predictor (LN, pink) and theta network (green), both for normal speed speech (compression factor 1) and compressed speech (compression factors 2 and 3). Performance was assessed on a different subsample of sentences than those used for parameter fitting.

**Figure 2 – Figure Supplement 1: TIMIT corpus and models used for syllable boundary detection**

- A. Acoustic analysis of TIMIT corpus.** *Left panel:* Speech modulation frequency increases with syllabic rate. All 4620 sentences of the TIMIT corpus (Test dataset) were sorted into quartiles according to syllabic rate (i.e. number of syllables per second). Speech envelope spectrum (with 1/f correction) was averaged over all sentences within each quartile, and the four averages are plotted. Color bars on top of the graphs represent the syllabic rate range for all four quartiles, showing a correspondence between the modal frequency and the syllabic rate over the corpus. *Middle panel:* Average channel spectrum. Spectrum were taken for each 128 auditory channels of the Chi and colleagues pre-cortical auditory model (Chi, Ru, and Shamma 2005), averaged over all sentences in the corpus. All channels show a clear peak in the same 4-8 Hz range, showing that the theta modulation is very present in the input to auditory cortex. *Right panel:* Syllable onset corresponds to a dip in spectrogram. Average of auditory spectrogram channels of sentences phase-locked to syllable onsets.  $t=0$  (green line) corresponds to syllable onset. Red colours correspond to high value, blue colours to low values. Dip at syllable onset is particularly pronounced over medium frequencies corresponding to formants. Auditory channels were averaged over all syllable onsets over the entire corpus (4620 sentences). This plot shows the connection between syllable boundaries and fluctuations of auditory channels that the auditory cortex may take advantage of in order to predict syllable boundaries.
- B. Theta network model.** Left panel: The architecture of the theta model is the same as the full model network without the PING component. Speech data is decomposed into auditory channels as in the LN model, and projected non-specifically onto 10  $T_e$  excitatory neurons. The  $T_e$  population interacts reciprocally with 10  $T_i$  inhibitory neurons, generating theta oscillations. Theta bursts provide the model prediction for syllable boundary timing.

- C. *Te* neurons burst at speech onset.** *Te* neurons provide onset-signalling neurons that respond non-specifically to the onset of all sentences. The spikes from one *Te* neuron were collected over presentation of 500 distinct sentences, and then referenced in time with respect to sentence onset. Here sentence onset was defined as the time when speech envelope first reached a given threshold (1000 a.u.). Spikes counts are then averaged in 20 ms bins, showing that this neuron displays a strong activity peak 0-60 ms after sentence onset. A secondary burst occurs around 200 ms after onset, as present in the example neuron shown in Brasselet et al. 2012.
- D. Model of linear-nonlinear (LN) predictor of syllable boundaries.** Auditory channels are filtered, summed and passed through a nonlinear function: the output determines the expected probability of syllable onset. A negative feedback loop prevents repeated onset at close timings. Values for filters, nonlinear function and feedback loops are optimized through fitting to a sub-sample of sentences.
- E. Stimulus-network coherence.** Theta phase (4-8 Hz) was extracted from both the simulated LFP and speech input. Coherence at each data point was computed as the Phase-Locking Value of the phase difference computed from 100 simulations with a distinct sentence. Coherence established in the 0-200 ms following sentence onset to a stable high coherence value of about 0.4.

**Figure 3. Sawtooth classification**

- A. Gamma spiking patterns in response to simple stimuli.** The model was presented with 50 ms sawtooth stimuli, where peak timing was parameterized between 0 (peak at onset) and 1 (peak at offset). Spiking is shown for different *Ge* neurons (y axis) in windows phase-locked to theta bursts (-20 to +70 ms around the burst, x-axis). Neural patterns are plotted below the corresponding sawtooths.
- B. Simulated networks.** The analysis was performed on simulated data from three distinct networks: ‘Undriven-theta model’ (no speech input to *Te* units, top), ‘Uncoupled theta/gamma model’ (no projection from *Te* to *Ge* units, middle), full intact model (bottom).
- C. Classification performance using stimulus vs. theta timing for the three simulated networks.** The stimulus timing (light bars) is obtained by extracting *Ge* spikes in a fixed-size window locked to the onset of the external stimulus; the theta timing (dark bars) is obtained by extracting *Ge* spikes in a window defined by consecutive theta

bursts (*theta chunk*, see Figure 3A). Classification was repeated 10 times for each network and neural code, and mean values and standard deviation were extracted. Average expected chance level is 10%.

- D. Stimulus detection performance**, for the intact and control models. Rest neural patterns were discriminated against any of the 10 neural patterns defined by the 10 distinct temporal shapes.
- E. Confusion matrices for stimulus- and theta-timing and the two control models (using theta-timing code)**. The colour of each cell represents the number of trials where a stimulus parameter was associated with a decoded parameter (blue: low numbers; red: high numbers). Values on the diagonal represent correct decoding.

**Figure 4. Continuous speech parsing and syllable classification**

- A. Decoding scheme.** Output spike patterns were built by extracting *Ge* spikes occurring within time windows defined by consecutive theta bursts (red boxes) during speech processing simulations. Each output pattern was then labelled with the corresponding syllable (grey bars).
- B. Syllable decoding average performance for uncompressed speech.** Performance for the three simulated models (Figure 3B) using two possible neural codes: *spike count* and *spike pattern*.
- C. Syllable decoding average performance across speakers, using the spike pattern code.** Syllable decoding was optimal when syllable duration was within the 100-300 ms range, i.e. corresponded to the duration of one theta cycle. The intact model performed better than the two controls irrespective of syllable duration range. Chance level is 10%. Colour code same as B.
- D. Syllable decoding performance for compressed speech for the intact model using the spike pattern code (same speaker, as in B).** Compression ranges from 1 (uncompressed) to 3. Average chance level is 10% (horizontal line in the right plot).

**Figure 4 – Figure Supplement 1. Syllable classification across speakers.**

- A. Distribution of syllable duration across 2 sentences and 462 speakers.** The shaded area (100-300 ms) indicates region of maximal density. Extreme values probably correspond to ill-defined syllables.

**Figure 5. Comparison with encoding properties of auditory cortical neurons.**

- A. Neural codes.** Stimulus decoding was performed on patterns of *Ge* spikes chunked in fixed-size windows (the figure illustrates the pattern for one neuron extracted from one window). *Spike count* consisted of counting all spikes for each neuron within the window. *Time-partitioned code* was obtained in dividing the window in  $N$  equal size bins (vertical grey bars) and counting spikes within each bin. *Phase-partitioned code* was obtained by binning LFP phase into  $N$  bins (depicted by the four colours in the top graph) and assigning each spike with the corresponding phase bin.
- B. Spike pattern decoding.** (Left) Decoding performance across *Ge* neurons for the intact model using  $N=8$  bins for each code: spike count (black curve), time-partitioned (blue curve) and phase-partitioned codes (green curve). (Right) Data from the original experiment. Adapted from Kayser et al., 2012.
- C. Mutual information (MI)** (Left) Mean MI between stimulus and individual output neuron activity during sentence processing in the intact model for spike count (black curve), time-partitioned (blue line), combined count and phase-partitioned (green line) and combined time- and phase-partitioned codes (red line). (Right) Comparison with experimental data from auditory cortex neurons (adapted from Kayser et al., 2009).

**Figure 5 – Figure Supplement 1: Speech decoding performance and MI (control models)**

- A. Stimulus decoding performance for each neural code across *Ge* neurons for the control models** (left:undriven theta; right: no theta-gamma connection): spike count (black line); time-partitioned neural code (blue line); phase-partitioned neural code (green line).
- B. Stimulus decoding performance as a function of bin number**, for all three variants of the model and experimental data. The number of bins used to partition the spikes was varied from 2 to 16, while the duration of the window was kept at 160 ms. Each dot corresponds to the average over 1000 different sets of stimuli and neuron (bars represent s.e.m.). Data from the original experiment, recording auditory cortex neurons from monkeys listening to naturalistic sounds. Experimental data are reproduced qualitatively by the intact model but not by the control model. Adapted from Figure 3E of Kayser et al., 2012.
- C. Mutual Information (MI) between acoustic stimulus and individual *Ge* neurons for the control models** (left: undriven theta; right: no theta-gamma connection): spike

count (black line); time-partitioned neural code (blue line); phase-partitioned & spike count neural code (green line); phase- and time- partitioned neural code (red lines). Both control models display low MI values and fail to display the pattern of experimental data shown in figure 5B.

## Tables

**Table 1. Full network parameter set**

Parameter	$C$	$V_{THR}$	$V_{RESET}$	$V_K$	$V_L$	$g_L$	$g_{Ge, Gi}$	$g_{Gi, Ge}$	$g_{Te, Ge}$
Value	1 F/cm <sup>2</sup>	-40 mV	-87 mV	-100 mV	-67 mV	0.1	5/N <sub>Ge</sub>	5/N <sub>Gi</sub>	0.3/N <sub>Te</sub>

Parameter	$\tau_{Ge}^R$	$\tau_{Te}^R$	$\tau_{Gi}^R$	$\tau_{Ti}^R$	$\tau_{Ge}^D$	$\tau_{Gi}^D$	$I_{Ge}^{DC}$	$I_{Gi}^{DC}$
Value	0.2 ms	4 ms	0.5 ms	5 ms	2 ms	20 ms	3	1

**Table 2. Optimal parameters for the LN model**

Parameter	$t_{sp}^{next}$	$\tau_{lh}$	$DC$
Value	0.0748	1.433	0.4672

**Table 3. Optimal parameters for the theta model**

Parameter	$\sigma_{Te}$	$\sigma_{Ti} = \sigma_{Ge} = \sigma_{Gi}$	$\tau_{Te}^D$	$\tau_{Ti}^D$	$I_{Te}^{ext}$	$I_{Te}^{DC}$	$I_{Ti}^{DC}$	$g_{Ti,Ti}$	$g_{Ti,Te}$	$g_{Te}^L$
Value	0.282	2.028	24.3	30.36	15	1.25	0.0851	0.432	0.207	0.264

Figure 1 - Model Architecture

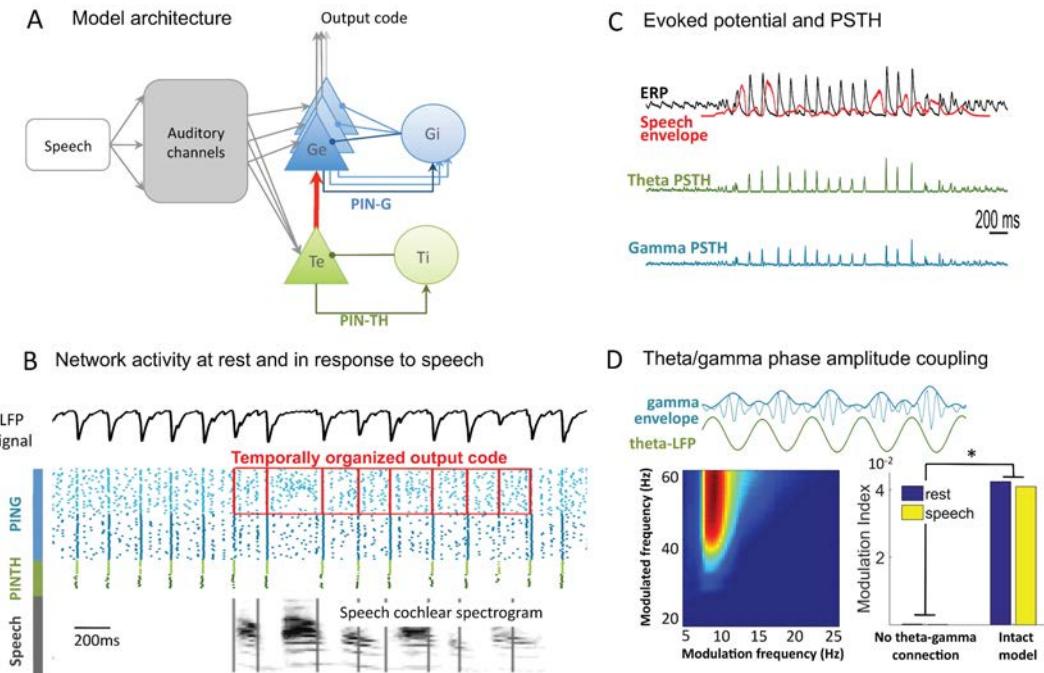


Figure 2 – Syllable boundary detection

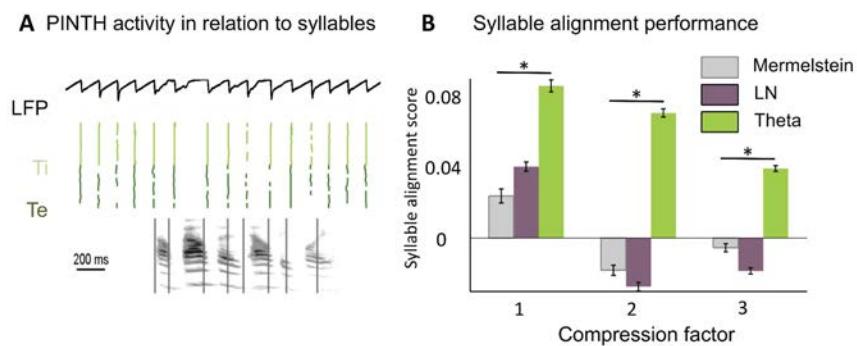


Figure 3 – Classification of model responses to simple stimuli

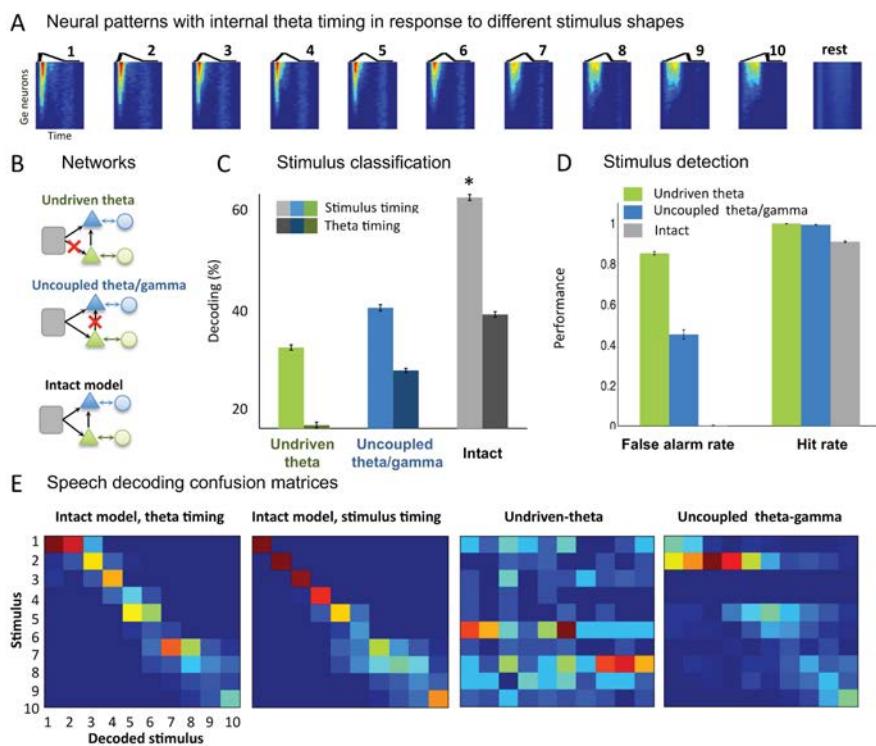


Figure 4 – Syllable parsing and classification from natural speech

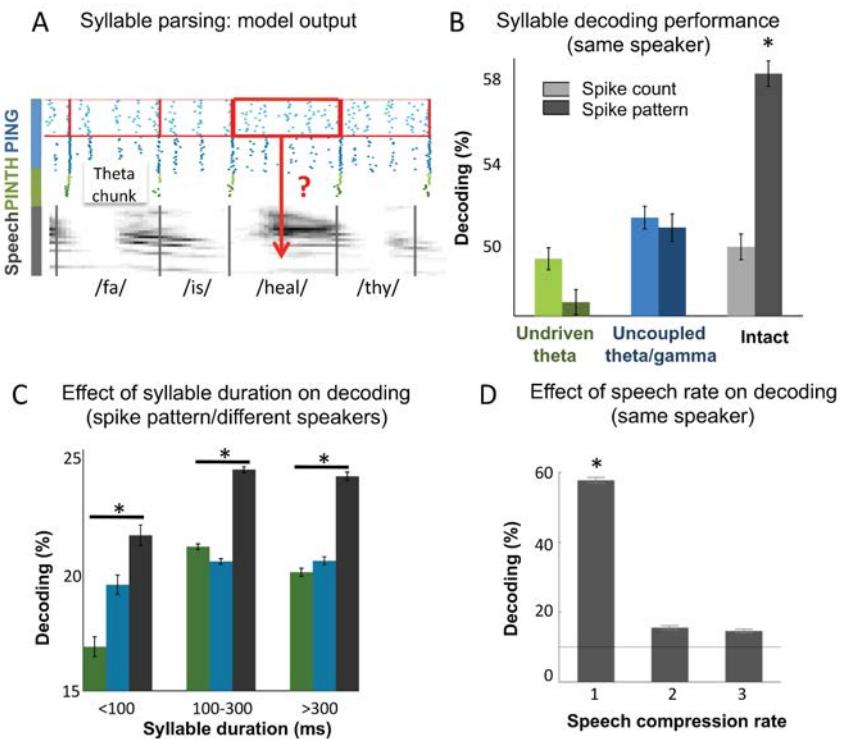
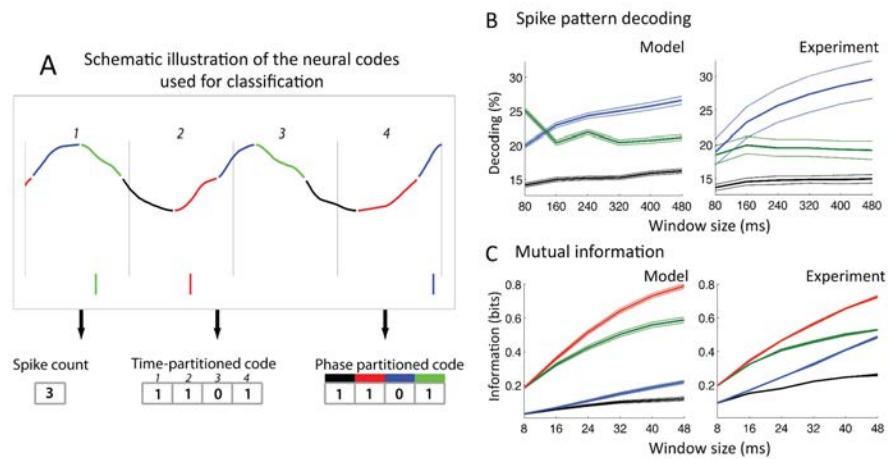


Figure 5 – Comparison of different neural codes



## SUPPLEMENTARY METHODS

### Syllable boundary prediction algorithms

Syllable boundaries triggered average (STAs) were computed as follow: for each syllable boundary (syllable onsets excluding the first of each sentence), we extracted a 700 ms window of the corresponding locked to the syllable boundary and averaged over all syllable boundaries. STAs were computed for speech envelope and for each channel of the Chi et al. (Chi et al., 2005) model.

### *Predictive models*

We compared the performance of four distinct families of models to predict the timing of syllable boundaries based on speech envelope or speech audiogram: the Mermelstein algorithm (see below), a Linear Non-linear (LN) model (a simplified integration-to-threshold algorithm, see below), the entrained theta neural oscillator (see below) and a purely rhythmic control model (see below).

#### 1. *Mermelstein algorithm*

The Mermelstein algorithm is a standard algorithm that predicts syllable boundaries by identifying troughs in the power of the speech signal (Mermelstein, 1975; Villing et al., 2004). The predicted boundaries are computed according to the following steps. First, extract the power of speech signal in the 500 – 4000 Hz range (grossly corresponding to formants) and low-pass filter at 40 Hz to remove fast fluctuations, defining a so-called *loudness function*. Second, for each sentence, compute the convex hull of the loudness signal and extract the maximum of the difference between the loudness signal and its convex hull. If that difference exceeds a certain threshold  $T_{min}$  and if the peak intensity of the interval of no more than  $P_{max}$  smaller than the peak intensity of the whole sentence, then that time of maximal difference is defined as a predicted boundary and the same procedure is applied recursively to the intervals to the left and right of that boundary. Parameters  $T_{min}$  and  $P_{max}$  were optimized to yield minimum prediction distance (see below), yielding  $T_{min} = 0.152 \text{ dB}$  and  $P_{max} = 15.85 \text{ dB}$ .

Note that this algorithm cannot be run *online* since the convex hull at a given time depends on the future value of speech power. Thus syllable boundaries can only be predicted after a

certain delay, which makes it impractical for online speech comprehension as occurring in the human brain.

## 2. LN model and variations

To evaluate the capacity of a simplified neural system to predict syllable boundaries, we trained a generalized linear point process model on the syllable dataset. The model (figure 2 – Figure Supplement 1D) does not incorporate full neural dynamics but simply comprises a linear stimulus kernel followed by nonlinear function. The process issues a ‘spike’ or ‘syllable boundary signal’ whenever the output reaches a certain threshold (Pillow et al., 2008). This signal is fed back into the nonlinear function (another kernel  $Ih$  is used here): such negative feedback loop implements a relative refractory period. This model is a generalization of the Linear-Nonlinear Poisson model, hence we refer to it simply as *LN* model. We used the 32 auditory channels as input to the model and trained it to maximize its syllable boundary prediction performance.

We looked for a linear filter that is separable in its temporal and spectral component. We first computed the Spike Triggered Average (or rather ‘Syllable Boundary Triggered Average’) for all 32 channels from 600 ms to 0 ms prior to the actual boundary in 10 ms time steps. Yet *STA* provides the optimal estimate for the linear kernel in a LN model only when stimulus consists of uncorrelated white noise (Chichilnisky, 2001). To get the optimal values out of the white noise condition, we looked at the separable filter  $H$  that yields best prediction of the output:

$$\min H \left( < |Y(t) - \hat{Y}(t|H)|^2 > \right), \text{ where :}$$

- $Y(t)$  is a binary output equal to 1 if there is a syllabic boundary in the 10 ms interval, 0 otherwise
- $H$  is separable spectrotemporal filter (i.e.  $H(\omega, u) = S(\omega)T(u)$  for all orders  $u$  and all frequencies  $\omega$ .  $S$  and  $T$  are, respectively, the spectral and temporal component of filter  $H$ .
- $\hat{Y}(t|H) = \sum_{u,w} H(w, u)X(\omega, t - u)$  where  $X(\omega, t)$  is the value of auditory channel  $\omega$  at time step  $t$ .

Optimal solutions of the system verify:

$$\begin{aligned}\sum_u T(u)R(\omega, u) &= \sum_{u,v,\xi} S(\xi)T(u)T(v)M(\omega, \xi, u, v) \quad \forall \omega \\ \sum_{\omega} S(\omega)R(\omega, u) &= \sum_{v,\omega,\xi} S(\omega)S(\xi)T(v)M(\omega, \xi, u, v) \quad \forall u\end{aligned}$$

where  $R(\omega, u) = \langle Y(t)X(\omega, t) \rangle_t$  (i.e.  $R$  is the Spike Triggered Average) and  $M$  is the covariance tensor for  $X$ , ie.

$$M(\omega, \xi, u, v) = \text{cov}(X(\omega, t - u), X(\xi, t - v)).$$

Solutions to  $T$  and  $S$  for that system of equations can be approximated numerically using the following iterative procedure:

$$\begin{aligned}S_0(\omega) &= 1 \quad \forall \omega, T_0(u) = 1 \quad \forall u \\ T_0 R &\\ S_{n+1} &= \left( \frac{\sum_{u,v} T_n(u)T_n(v)M(u, v, \dots)}{\sum_{\omega,\xi} S_{n+1}(\omega)S_{n+1}(\xi)M(\omega, \xi, \dots)} \right)^T \\ T_{n+1} &= \left( \frac{RS_0}{\sum_{\omega,\xi} S_{n+1}(\omega)S_{n+1}(\xi)M(\omega, \xi, \dots)} \right)\end{aligned}$$

and then stopping when the resulting square error  $\|RS_0 - \sum_{\omega,\xi} S_{n+1}(\omega)S_{n+1}(\xi)T_n(v)M(\omega, \xi, v)\|_u^2$  goes below a minimum value (we used a threshold of  $10^{-1}$ ). The first 6 components (i.e. time bins) of the temporal kernel (i.e. 0 – 50 ms) were also used for input convolution in the theta model. We did not integrate further components (60 – 400 ms) since their weight was much lower and its implementation by relay neurons seemed less realistic.

To retrieve the optimal value for all parameters of the model, we used the GLM matlab toolbox developed in the Pillow lab ([http://pillowlab.cps.utexas.edu/code\\_GLM.html](http://pillowlab.cps.utexas.edu/code_GLM.html)), using as input the one-dimensional signal  $U(t) = \sum_{\omega} S(\omega)X(\omega, t)$ . Other parameters of the *LN* model including the self-inhibition temporal kernel  $Ih$  were optimized using the gradient descent implemented in the toolbox. This method provides estimation for a stochastic generalized *LN* model. We were interested in assessing the performance of a deterministic *LN* model. We then run a deterministic model with the same parameters as the stochastic model plus one new free parameter describing the normalized time to next spike (in the stochastic model, that time is drawn from an exponential distribution). The value of  $t_{sp}^{next}$  was optimized using the same minimization procedure used for others models (see below). Two other parameters were also optimized again, since this procedure

minimized a different score than the GLM toolbox score: time scale of self-inhibition  $\tau_{Th}$  and constant input to the model  $DC$  (Table 2).

We made one last modification to this  $LN$  model. We optimized the model such that it would maximally fire not at the time of syllable boundaries but 10 ms posterior to that time (*de facto*, we simply slid the STA window by 10 ms). This provides a delayed signal but likely more reliable since it can use more information (notably the rebound in the auditory spectrogram that is present right after a syllable boundary).

### 3. Theta model

The theta model is composed of the  $Te$  and  $Ti$  cells from the full network model described above, with the exact same parameter set. 11 parameters were optimized in the full model, 10 in the control model (see values in the Table 3).

### 4. Control model

The control model was used to provide a baseline for assessing the performance of other models. Under these control conditions, predicted syllable boundaries were generated rhythmically at a fixed time interval, irrespective of the stimulus. The rate of the rhythmic process was varied from 1 Hz to 15 Hz in 0.5 Hz intervals. Such control model yielded better performance than another control model consisting of a homogeneous Poisson process. It thus provides a more stringent control for estimating the efficiency of other algorithms.

#### Model performance evaluation

We evaluated how well syllable boundaries predicted by any model matched with the boundaries derived from labelled speech data. As an evaluation metrics we used a point process distance that is used to compare distance between spike trains (Victor and Purpura, 1997). Shift cost was set to  $20 \text{ sec}^{-1}$  (in other words, a predicted and an actual boundary could be matched if they were no more than 50 msec apart). The distance was summed over all sentences used in the dataset, and normalized by the sum of all actual and predicted boundaries over the dataset.

We also assessed performance of the models by reporting standard measures of the detection theory. Each actual boundary was counted as a *hit* if it was distant by no more than a time jitter of 50 ms to any predicted boundary. Actual boundaries with no associated

predicted boundary was termed *misses*. Predicted boundaries with no associated actual boundary were termed *false alarms*.

We finally computed *true positive rate (TPR)* and *false positive rate (FPR)* over a set of sentences as follows:

$$TPR = \Sigma \text{hits} / (\Sigma \text{hits} + \Sigma \text{misses})$$

$$FPR = \Sigma \text{false alarms} / (\Sigma \text{false alarms} + \Sigma \text{hits})$$

where  $\Sigma$  represents the sum over all simulations. Note that FPR here differs from the standard definition of False Positive Rate since correct rejection events cannot be defined in a point process detection task. TPR and FPR for each model are then plotted in a ROC curve for model comparison.

Neither of those measures was adequate to compare the performance of the same model (here, the theta model) across different levels of speech compression. To draw such comparison, for each level of compression, we computed the (non-normalized) distance measure for the theta model summed over all sentences in the test dataset, as well as the average number of predicted boundaries per sentence. We then matched the theta model to a control rhythmic model with the same predicted syllabic rate, and computed the difference between the non-normalized distance for the theta model and for that matched rhythmic model.

#### *Optimisation*

We optimized the parameters from all models (except the control model for which we used the whole range of values of the rate parameter) to get the minimal normalized point process distance between predicted and actual boundaries in each sentence. Optimization was made using global gradient descent (function *fminsearch* in Matlab), and repeated with many initial points to avoid retaining a local minimum. Although both the theta model and the control model are intrinsically stochastic, the sample size was large enough for the objective function over the entire sample to be nearly deterministic, allowing for convergence of the gradient descent algorithm. The list of optimized parameters for each type of model is provided in the related model sections above. We split the entire TIMIT TRAIN dataset (4620 sentences) into two datasets: a first dataset of 1000 sentences was used to compute optimal parameters; final assessment of an algorithm performance with its optimal parameters was done on a separate set of 3620 sentences.

Table 1. Parameter set

Parameter	$C$	$V_{THR}$	$V_{RESET}$	$V_K$	$V_L$	$g_L$	$g_{Ge,Gi}$	$g_{Gi,Ge}$	$g_{Te,Ge}$
Value	$1\text{ F/cm}^2$	-40 mV	-87 mV	-100 mV	-67 mV	0.1	$5/N_{Ge}$	$10/N_{Gi}$	$0.3/N_{Te}$

Parameter	$\tau_{Ge}^R$	$\tau_{Te}^R$	$\tau_{Gi}^R$	$\tau_{Ti}^R$	$\tau_{Ge}^D$	$\tau_{Gi}^D$	$I_{Ge}^{DC}$	$I_{Gi}^{DC}$
Value	0.2 ms	4 ms	0.5 ms	5 ms	2 ms	20 ms	3	

Table 2. Optimal parameters for the LN model

Parameter	$t_{sp}^{next}$	$\tau_{Ih}$	$DC$
Value	0.0748	1.433	0.4672

Table 3. Optimal parameters for the theta model

Parameter	$\eta_{Te}$	$\eta_{Ti} = \eta_{Ge} = \eta_{Gi}$	$\tau_{Te}^D$	$\tau_{Ti}^D$	$I_{Te}^{ext}$	$I_{Te}^{DC}$	$I_{Ti}^{DC}$	$g_{Ti,Ti}$	$g_{Ti,Te}$	$g_{Te}$
Value	0.282	2.028	24.3	30.36	15	1.25	0.0851	0.432	0.207	0.264

## 11.2 Analytical insights on Theta-Gamma coupled neural oscillators

Journal of Mathematical Neuroscience (2013) 3:16  
DOI 10.1186/2190-8567-3-16

 The Journal of Mathematical Neuroscience  
a SpringerOpen Journal

RESEARCH

Open Access

### Analytical Insights on Theta-Gamma Coupled Neural Oscillators

Lorenzo Fontolan · Maciej Krupa ·  
Alexandre Hyafil · Boris Gutkin

Received: 11 March 2013 / Accepted: 13 June 2013 / Published online: 14 August 2013  
© 2013 L. Fontolan et al.; licensee Springer. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Abstract** In this paper, we study the dynamics of a quadratic integrate-and-fire neuron, spiking in the gamma (30–100 Hz) range, coupled to a delta/theta frequency (1–8 Hz) neural oscillator. Using analytical and semianalytical methods, we were able to derive characteristic spiking times for the system in two distinct regimes (depending on parameter values): one regime where the gamma neuron is intrinsically oscillating in the absence of theta input, and a second one in which gamma spiking is directly gated by theta input, i.e., windows of gamma activity alternate with silence periods depending on the underlying theta phase. In the former case, we transform the equations such that the system becomes analogous to the Mathieu differential equation. By solving this equation, we can compute numerically the time to the first gamma spike, and then use singular perturbation theory to find successive spike times. On the other hand, in the excitable condition, we make direct use of singular perturbation theory to obtain an approximation of the time to first gamma spike, and then extend the result to calculate ensuing gamma spikes in a recursive fashion. We thereby give explicit

---

L. Fontolan (✉)  
Department of Fundamental Neurosciences, CMU, University of Geneva, 1 rue Michel Servet, 1211 Geneva, Switzerland  
e-mail: [lorenzo.fontolan@unige.ch](mailto:lorenzo.fontolan@unige.ch)

M. Krupa  
INRIA Paris-Rocquencourt Research Centre, Domaine de Voluceau BP 105, 78153 Le Chesnay, France  
e-mail: [maciej.p.krupa@gmail.com](mailto:maciej.p.krupa@gmail.com)

A. Hyafil · B. Gutkin  
Group for Neural Theory, Département des Etudes Cognitives, Ecole Normale Supérieure, 5 rue d'Ulm, 75005 Paris, France

A. Hyafil  
e-mail: [alexandre.hyafil@gmail.com](mailto:alexandre.hyafil@gmail.com)

B. Gutkin  
e-mail: [boris.gutkin@ens.fr](mailto:boris.gutkin@ens.fr)

 Springer

formulas for the onset and offset of gamma spike burst during a theta cycle, and provide an estimation of the total number of spikes per theta cycle both for *excitable* and *oscillator* regimes.

**Keywords** Oscillations · PING · Dynamical systems · Geometric singular perturbation theory · Blow-up method · Spike times · Theta-gamma rhythms · Type I neuron · SNIC bifurcation

## 1 Introduction

Oscillations of neural activity are ubiquitous in the brain in many frequency bands [1], and it has been often argued that they play a functional role in cortical processing [2–4]. Physiological experiments and computational models have shown that ongoing brain oscillations are involved in sensory-motor functions [5], synaptic plasticity [6], memory formation and maintenance [7], among many other cognitive tasks. Indeed, it has been reported [2] that intrinsic brain rhythms can bias input selection, temporally link neurons into assemblies, and facilitate mechanisms that cooperatively support temporal representation and long-term consolidation of information. Notably gamma oscillations ( $>30$  Hz) are prominent in neocortex during attention [8], sensory processing [9, 10], or motor control tasks [11], together with slower rhythms in the theta (3–8 Hz) or delta (1–3 Hz) range that have also been linked to various aspects of cognitive processes like working memory or the transmission of sensory and motor signals.

Many recent contributions point to nontrivial interactions among different frequency bands [12–14], such as phase-amplitude [15, 16] or phase-phase coupling [17, 18] that can facilitate the simultaneous integration of multiple layers of information [19]. The hippocampus is a privileged site for observing such interactions [11, 20], since theta and gamma waves are particularly strong and reliable in that region [21]. Another particular case is represented by perception of speech signal performed by auditory cortex. In fact, to capture the many different relevant features of speech (i.e., syllables, vowels, consonants, etc.), the brain must be able to parse the speech signal over these many time-scales at the same time. A number of recent works introduced the hypothesis that a network of nested theta (3–8 Hz) and gamma (30–100 Hz) rhythms could accomplish this task [22–24], given their matching in frequency with syllabic and phonemic time-scale, respectively. Since there is no external onset signaling the presence of an incoming syllabic content, the phase of the gamma rhythm needs to be reset by some intrinsic mechanism, e.g., by theta input [23]. It becomes therefore important to know the time to first spike, which would be a measure of the speed of gamma phase resetting, as well as the time to last spike and the spiking frequency during excitable period.

There is a large literature on mathematical analysis of single frequency oscillators in networks of cortical circuits [25–31], and much work has been done in computational modeling of neural oscillations [2, 32, 33]. There is also a significant number of mathematical studies on cross-frequency interactions, however, most of that analysis is limited to the cases of weak coupling [34–37]. Strong coupling case has been

analyzed either with pulsatile coupling [25, 38, 39] or with semianalytical and computational techniques [40–42]. Importantly, the question of how strong continuous coupling between slow and fast oscillations influences frequency and time of fast spikes has not been treated analytically, at least to the best of our knowledge. Yet experimental data suggest that phase-amplitude coupling in the brain is continuous (i.e., low-frequency phase is conveyed through local field potential, a continuous signal) and strong [15, 39, 41], so this will be the regime we aim to study in the present work.

In this article, we provide analytical insights on the precise spiking times of a simplified Pyramidal Interneuron Network Gamma (PING) [41] during theta modulation. Two separate cases are studied: In the first setting, which we will refer to as *oscillatory regime*, the gamma network behaves as an intrinsic oscillator whose spike frequency is modulated by the theta phase; in the second, named *excitable regime*, gamma spikes are only evoked when input coming from the theta oscillator is strong enough. In the latter case, the system is in an “excitable” regime, where theta pushes gamma back and forth across a Saddle-Node on Invariant Circle (SNIC) bifurcation. The analysis can be generalized beyond theta-gamma nested oscillations; indeed it describes any coupling between low and a high frequency rhythms [43], provided that the latter is produced through feedback inhibition to the excitatory cell. To compute the time to the first gamma spike, we used different approaches for the two regimes: In the oscillatory case, we reduce the system in order to describe its dynamics with the Mathieu equation [44], and in the excitable case we apply an extension of geometric singular perturbation theory [45–47]. We then use a combination of the two to get successive spike times and an estimation of the total number of spikes per theta cycle.

The paper is organized as follows:

1. In Sect. 1, we introduce the system to be studied.
2. In Sect. 2, we consider the system in the oscillatory regime and compute time to first gamma spike using Mathieu functions. We found that spike time is mainly determined by the magnitude of theta-gamma coupling ( $\lambda$ ) and of theta frequency.
3. In Sect. 3, we turn our attention to the excitable regime where theta phase determines the magnitude of input, thereby causing the gamma circuit to spike.
4. Finally, we show that our approach gives results in agreement with direct numerical simulations of the system of interest.

In our analysis, we use tools from geometric singular perturbation theory. This approach normally fails in proximity of nonhyperbolic points, as it would be the case for the system considered in the present paper, but the *blow-up method* extension provided in [48] allows us to compute approximations of the passage time to the first spike in the excitable case, and it is used both in the oscillator and excitable cases to estimate the duration of inhibition and the passage time of subsequent spikes. The latter estimates are based on the idea that inhibition puts the system in a state of quasi equilibrium; consequently, they work well if inhibition is strong and excitation not too high.

decay, so that almost no residual inhibition is present at the beginning of a new theta cycle. In order to obtain an equation in the form of Mathieu equation, we first perform a change of variables in system (1)  $V_E = \tan \frac{\theta_E}{2}$ , going from  $\theta$ -variable to membrane voltage  $V_E$ . As it is known [40],  $\theta$ -neuron is formally equivalent to the Quadratic Integrate and Fire (QIF) neuron:

$$\begin{aligned} \frac{dV_E(t)}{dt} &= V_E^2(t) + I_E + \lambda(1 + \cos(\Theta)), \\ \frac{d\Theta}{dt} &= \varepsilon_\Theta \omega. \end{aligned} \quad (2)$$

The two neural models are formally equivalent if we define the reset conditions as

$$V_E(t^* - 0) = +\infty, \quad V_E(t^* + 0) = -\infty,$$

where  $t^*$  the time of spike. We have omitted the synaptic input dynamics since we assume that inhibition is directly enslaved to spikes coming from EG neuron, hence the inhibitory synapse  $s_I$  stays inactive up to the first EG spike. We restate the system in (2) as a single equation, assuming by convention that  $\Theta = -\pi$  at  $t = 0$  (or at the beginning of a new theta cycle):

$$\frac{dV_E(t)}{dt} = V_E^2(t) + I_E + \lambda(1 + \cos(\varepsilon_\Theta \omega t - \pi)). \quad (3)$$

For  $I_E > 0$ , this equation has an exact solution in terms of Mathieu functions. This can be found by imposing one more change of variable:

$$V_E = -\frac{u'}{u}, \quad u(t) = e^{-\int V_E(\tau)d\tau}, \quad (4)$$

where the prime mark denotes the time derivative (a similar transformation is used in [50] where the cosinusoidal forcing term was replaced by exponential decay, leading to a different solution of the corresponding differential equation). Hence, we write (3) as a second-order differential equation:

$$u'' = -(I_E + \lambda(1 + \cos(\varepsilon_\Theta \omega t - \pi)))u. \quad (5)$$

If parameters  $a, q, z$  are rescaled as following:

$$z = \frac{\varepsilon_\Theta \omega t}{2}, \quad a = \frac{4(I_E + \lambda)}{\varepsilon_\Theta^2 \omega^2}, \quad q = \frac{2\lambda}{\varepsilon_\Theta^2 \omega^2}, \quad (6)$$

then Eq. (5) has the form of a Mathieu equation:

$$\frac{d^2u}{dz^2} = -(a - 2q \cos(2z))u. \quad (7)$$

To interpret Eq. (7), we need temporal rescaling from  $t$  to  $z$ , and as a consequence the period of cosinusoidal term, which in Eqs. (1) and (2) was  $T = \frac{2\pi}{\varepsilon_\Theta \omega}$ , becomes

## 2 Theta-Gamma Coupled Oscillator

We consider a minimal formulation of a theta modulated gamma spiking network. One single Excitatory Gamma (EG) neuron ( $\theta_E$ ), modeled as a  $\theta$ -neuron [49] receives an excitatory input coming from an oscillator ( $\Theta$ ) whose natural frequency lies in the theta band. The canonical  $\theta$ -neuron model is described by a phase variable lying on a one-dimensional circle in the range  $\theta_E \in [-\pi, \pi]$ , a spike is produced when  $\theta_E = \pi$ . The EG neuron participates in a PING rhythm, although in our case the inhibitory gamma neuron is instantaneously enslaved to the excitatory cell, meaning that every excitatory spike would immediately prompt a simultaneous inhibitory spike [32]. This allows us to suppress the explicit dynamics of the inhibitory gamma neuron and focus on inhibitory synaptic dynamics only. Our system can be described by the following equations:

$$\begin{aligned}\frac{d\theta_E}{dt} &= (1 - \cos(\theta_E)) + (I_E + \lambda(1 + \cos(\Theta)) - g_{EI}s_I)(1 + \cos(\theta_E)), \\ \frac{ds_I}{dt} &= -\varepsilon_I s_I + \delta(\theta_E - \pi), \\ \frac{d\Theta}{dt} &= \varepsilon_\Theta \omega,\end{aligned}\tag{1}$$

where  $\Theta \in [-\pi, \pi]$  is the instantaneous phase of the slow rhythm variable (delta/theta frequency band, i.e., 1–8 Hz), which provides the sinusoidal modulatory input to the EG cell;  $s_I$  is the variable representing the activation of the inhibitory synapse;  $I_E$  represents constant driving input to excitatory gamma neuron;  $\lambda$  is the strength of theta-gamma coupling;  $g_{EI}$  is the inhibitory synaptic strength;  $\omega$  has been chosen so that frequency  $\varepsilon_\Theta \omega$  falls into the theta range;  $\varepsilon_I$  is a scaling parameter that scales inversely with the time constant of synaptic inhibition;  $\varepsilon_\Theta$  is a second, slower, scaling parameter that has been chosen such that  $\varepsilon_\Theta \sim \varepsilon_I^2$ , an assumption that is motivated by biophysical considerations and, in addition, keeps the three time scales (theta rhythm, synaptic inhibition, and excitatory membrane potential) well separate.

We will consider two cases: the *oscillator* case, defined by  $I_E > 0$ , and the *excitable* case, defined by  $I_E < 0$ . The characterizing feature of the oscillator setting is that  $\theta_E - s_I$  subsystem in (1) is an intrinsic oscillator at every stage of a  $\Theta$ -cycle, i.e., the total current input to EG neuron is always positive. In the *excitable* case, on the other hand, part of theta oscillation period is such that  $\theta_E$  subsystem of (1) has an attracting quasisteady state, i.e., the total input to the EG neuron is negative or positive depending on  $\Theta$ -oscillator phase. If  $I_E < -2\lambda$ , the net input to EG neuron is always negative and the gamma circuit is always silent.

## 3 Time to First Spike, Oscillator Case

Let us consider the case in which constant driving term  $I_E$  in system (1) is positive and such that, in absence of theta modulation, the EG neuron would fire periodically with a spiking frequency in the gamma range (30–150 Hz). We assume that the dynamics of the theta oscillator is at least one order of magnitude slower than synaptic

decay, so that almost no residual inhibition is present at the beginning of a new theta cycle. In order to obtain an equation in the form of Mathieu equation, we first perform a change of variables in system (1)  $V_E = \tan \frac{\theta_E}{2}$ , going from  $\theta$ -variable to membrane voltage  $V_E$ . As it is known [40],  $\theta$ -neuron is formally equivalent to the Quadratic Integrate and Fire (QIF) neuron:

$$\begin{aligned} \frac{dV_E(t)}{dt} &= V_E^2(t) + I_E + \lambda(1 + \cos(\Theta)), \\ \frac{d\Theta}{dt} &= \varepsilon_\Theta \omega. \end{aligned} \quad (2)$$

The two neural models are formally equivalent if we define the reset conditions as

$$V_E(t^* - 0) = +\infty, \quad V_E(t^* + 0) = -\infty,$$

where  $t^*$  the time of spike. We have omitted the synaptic input dynamics since we assume that inhibition is directly enslaved to spikes coming from EG neuron, hence the inhibitory synapse  $s_I$  stays inactive up to the first EG spike. We restate the system in (2) as a single equation, assuming by convention that  $\Theta = -\pi$  at  $t = 0$  (or at the beginning of a new theta cycle):

$$\frac{dV_E(t)}{dt} = V_E^2(t) + I_E + \lambda(1 + \cos(\varepsilon_\Theta \omega t - \pi)). \quad (3)$$

For  $I_E > 0$ , this equation has an exact solution in terms of Mathieu functions. This can be found by imposing one more change of variable:

$$V_E = -\frac{u'}{u}, \quad u(t) = e^{-\int V_E(\tau)d\tau}, \quad (4)$$

where the prime mark denotes the time derivative (a similar transformation is used in [50] where the cosinusoidal forcing term was replaced by exponential decay, leading to a different solution of the corresponding differential equation). Hence, we write (3) as a second-order differential equation:

$$u'' = -(I_E + \lambda(1 + \cos(\varepsilon_\Theta \omega t - \pi)))u. \quad (5)$$

If parameters  $a, q, z$  are rescaled as following:

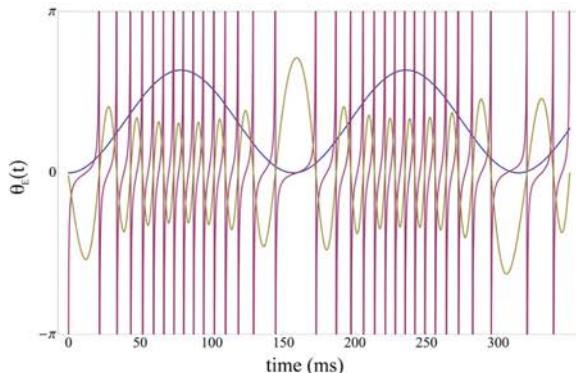
$$z = \frac{\varepsilon_\Theta \omega t}{2}, \quad a = \frac{4(I_E + \lambda)}{\varepsilon_\Theta^2 \omega^2}, \quad q = \frac{2\lambda}{\varepsilon_\Theta^2 \omega^2}, \quad (6)$$

then Eq. (5) has the form of a Mathieu equation:

$$\frac{d^2u}{dz^2} = -(a - 2q \cos(2z))u. \quad (7)$$

To interpret Eq. (7), we need temporal rescaling from  $t$  to  $z$ , and as a consequence the period of cosinusoidal term, which in Eqs. (1) and (2) was  $T = \frac{2\pi}{\varepsilon_\Theta \omega}$ , becomes

**Fig. 1** Oscillatory regime—dynamics of theta-modulated EG neuron in absence of inhibition. Red line: membrane voltage of EG neuron in presence of theta modulation (blue line) without inhibitory synaptic input ( $I_E = 0$ ). Solution  $u(t)$  to Mathieu equation is plotted in yellow



$T^* = \pi$ . The solutions to Eq. (7) are linear combinations of the even and odd Mathieu functions [44],  $Ce(a, q, z)$  and  $Se(a, q, z)$ , respectively. The solution

$$\begin{aligned} u(z) = & 2(-2\pi Ce(a, q, 0) + \varepsilon_\Theta \omega \dot{Ce}(a, q, 0) Se(a, q, z) \\ & + Ce(a, q, z)(4\pi Se(a, q, 0) - 2\varepsilon_\Theta \omega \dot{Se}(a, q, 0))), \end{aligned}$$

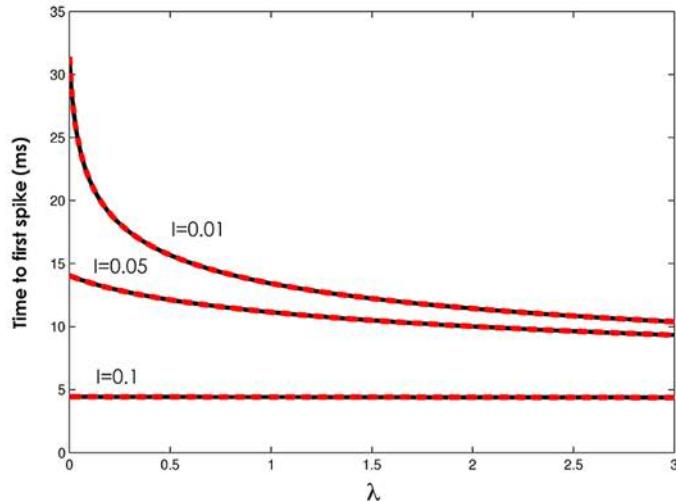
obeys the desired initial conditions, where the dot indicates the derivative with respect to  $z$ . Because of the change of variable in (4), the spiking times in the absence of inhibition correspond to the zeros of the solution of (7)  $u(z)$  (Fig. 1). Hence, by scaling back to the original variables and looking at the first zero of  $u(z)$  we obtain the time to the very first spike  $T_1$ . We numerically compute the time to the first spike as a function of parameters  $a$  and  $q$ , i.e.,  $I_E$  and  $\lambda$ . The subsequent spikes, on the other hand, depend on inhibition, and thus cannot be described by (3) alone. We looked for solutions of (2) with initial condition  $V_E(0) = -\infty$  and  $\Theta(0) = -\pi$ . Figure 2 shows the time to first spike  $T_1$  as a function of  $\lambda$  with  $I_E$  fixed at three values,  $I_E = 0.01$ ,  $I_E = 0.05$ , and  $I_E = 0.1$ . Note that for  $I_E = 0.01$  the dependence on  $\lambda$  is strong, but for larger values of  $I_E$  the sensitivity of  $T_1$  with respect to  $\lambda$  is smaller, since  $I_E$  becomes the dominant input term. In the next section, we will consider the case when inhibition is strong and fully controls the gamma spikes.

#### 4 Time to First Spike, Excitable Case

The excitable case implies that  $I_E < 0$ , and  $2\lambda + I_E > 0$ . Under these assumptions, the gamma spikes are only possible when  $\Theta$  lies in a proper subinterval of  $[-\pi, \pi]$ , which corresponds to the values of  $\Theta$  for which  $\lambda(1 + \cos(\Theta)) + I_E > 0$ . This ensures that the dynamics of (1) cross the SNIC bifurcation for a certain value of  $\cos(\Theta)$ . We carry out the computation with the initial conditions

$$\theta_E = \theta_0, \quad \Theta = -\pi, \quad s_I = 0, \quad (8)$$

where  $-\pi < \theta_0 < 0$  is defined by the condition  $(1 - I_E)\cos\theta_0 = 1 + I_E$ , i.e.  $\theta_0$  is a stationary fixed point in the absence of  $\Theta$  positive input. The gamma neuron relaxes



**Fig. 2** Oscillatory regime—time to first spike. Time to first spike as a function of coupling constant  $\lambda$ , for  $I_E = 0.01$ ,  $I_E = 0.05$  and  $I_E = 0.1$ . Black line: simulation; red dotted line: solution to Mathieu equation.  $\varepsilon_\Theta = 0.01$ ,  $\omega = 4$

to  $\theta_0$  when theta modulation is turned off. Note that  $\theta_E = \theta_0$  is not required for our solution to be applicable, since, for any initial value  $\theta_{\text{in}} < \theta_0$ ,  $\theta_E$  quickly converges to  $\theta_0$ . At the end of every theta cycle  $s_I$  goes back to zero, since its decay constant  $\varepsilon_I$  is one order of magnitude bigger than  $\varepsilon_\Theta$ , and the EG cell has stopped firing once inhibition pushed it below the SNIC bifurcation. We start by computing an estimate of the time to the first gamma spike. System (1) involves two time scales, one that controls the intrinsic dynamics of the EG neuron and the other comes from  $\Theta$  modulation. In the excitable case, rather than using the approach based on Mathieu functions, we use geometric singular perturbation theory. This approach leads to explicit estimates of the onset and duration of the gamma burst, it gives some geometric insights and can be applied in a more general setting. In order to compute the time at which the fast and the slow dynamics intersect, we need the value of  $\Theta$  corresponding to  $I_E + \lambda(1 + \cos(\Theta)) = 0$ , i.e., where the SNIC bifurcation takes place. Simple algebra shows that this occurs when

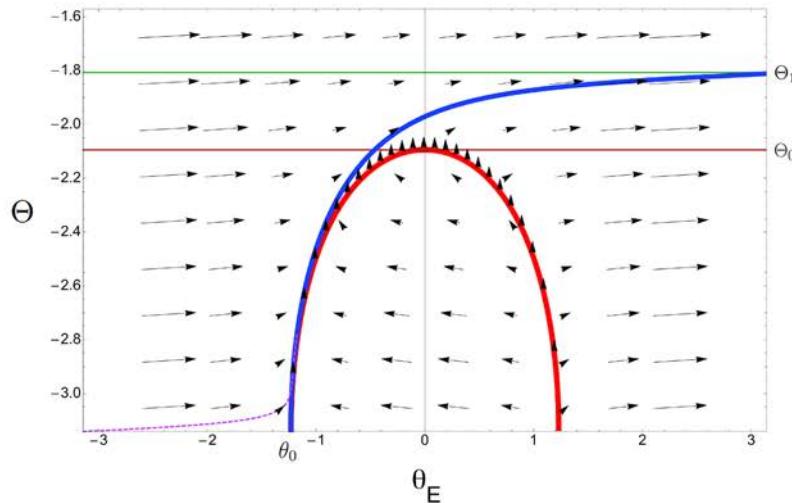
$$\cos(\Theta) = -\frac{\lambda + I_E}{\lambda}. \quad (9)$$

To ensure that (9) has solutions, we verify that the RHS of (9) is in the interval  $(-1, 1)$ . For the upper bound, we have

$$2\lambda + I_E > 0 \quad \Rightarrow \quad \lambda > -(I_E + \lambda) \quad \Rightarrow \quad 1 > -\frac{\lambda + I_E}{\lambda}.$$

The lower bound is obtained as follows:

$$-\frac{\lambda + I_E}{\lambda} = -1 + \frac{-I_E}{\lambda} > -1$$



**Fig. 3** Excitatory regime—phase plane. Phase portrait of system (1) for  $I_E = -0.5$ ,  $\lambda = 1$ ,  $\varepsilon_\Theta = 0.01$ ,  $\omega = 4$ . The blue line represents the trajectory of the system when starting from initial conditions  $(\theta_0, -\pi)$ : it passes along the nullcline  $\frac{d\theta_E}{dt} = \varphi(\theta_E, \Theta) = 0$  (in red) and then quickly escapes to  $(\pi, \theta_1)$  once past the singular point  $(0, \theta_0)$ . The dotted purple line shows that, for any starting point  $(\Theta, \theta_E)$  where  $\Theta < \theta_0$  and  $\theta_E < \theta_0$ , the trajectory converges to the blue line

given that  $\frac{-I_E}{\lambda} > 0$ . Let now  $\Theta_0$  be a solution of (9) satisfying  $-\pi < \Theta_0 < 0$  and let us consider system (1) with initial conditions (8): it is clear that any trajectory of the system can roughly be divided into two separate chunks (see Fig. 3). Starting from point  $(\theta_0, -\pi)$ , the system immediately enters the slow motion part of the trajectory, which is adjacent to the nullcline  $\frac{d\theta_E}{dt} = \varphi(\theta_E, \Theta) = 0$ . The slowest region of motion lies in the vicinity of the singular point  $(0, \Theta_0)$ , where both  $\varphi(\theta_E, \Theta)$  and its derivative with respect to  $\theta_E$  are zero. Once the trajectory has gone beyond the singular point,  $\varphi$  turns positive again and grows quadratically in magnitude. This way  $\theta_E$  quickly reaches the value  $\theta_E = \pi$ , since it is well known that any unbounded solution of the theta neuron for positive net input explodes in finite time. At the same time  $\Theta$  increase is of order  $\sim O(\varepsilon_\Theta)$ . Now let us start by computing the time spent along the fiber which is close to the nullcline, and then direct our attention to the motion in the neighborhood of the singular point. The time needed for  $\Theta$  to grow from  $-\pi$  to  $\Theta_0$  equals

$$T_1^* = \frac{\Theta_0 + \pi}{\varepsilon_\Theta \omega}. \quad (10)$$

When  $\Theta$  reaches  $\Theta_0$ ,  $\theta_E$  is  $O(\varepsilon_\Theta)$  (recall that this is also  $O(\varepsilon^2)$ ) close to the threshold value of  $\theta_E = 0$ . In order to estimate the time that EG neuron needs to produce the first spike, i.e., to reach  $\theta_E = \pi$ , we need to examine the behavior when close to point  $(0, \Theta_0)$ . We first translate the variable  $\Theta$ , introducing  $\tilde{\Theta} = \Theta - \Theta_0$ . Using

Taylor expansion around  $\theta_E \approx 0$  and  $\Theta \approx \Theta_0$ , we can write

$$1 - \cos(\theta_E) \simeq \frac{1}{2}\theta_E^2 + O(\theta_E^4), \quad (11)$$

$$I_E + \lambda(1 + \cos(\Theta)) \simeq \sqrt{-I_E(2\lambda + I_E)}\tilde{\Theta} + O(\tilde{\Theta}^2).$$

We transform (1) to the coordinates  $(\theta_E, \tilde{\Theta})$ , taking into the account the expansion (11) and ignoring  $s_I$  which remains zero until the first gamma spike (we omit the tilde for the simplicity of notation). The resulting system is

$$\begin{aligned} \frac{d\theta_E}{dt} &= a\Theta + \frac{1}{2}\theta_E^2 + O(\theta_E^4, \Theta^2, \theta_E^2\Theta), \\ \frac{d\Theta}{dt} &= \varepsilon_\Theta\omega, \end{aligned} \quad (12)$$

with  $a = 2\sqrt{-I_E(2\lambda + I_E)}$ .

We then rescale the variables of (12) as follows:

$$\theta_E = 2x, \quad \Theta = \frac{2}{a}y, \quad \tilde{\varepsilon} = \frac{\varepsilon_\Theta\omega a}{2}.$$

In terms of the rescaled variables, with the tilde omitted, system (12) becomes

$$\begin{aligned} \frac{dx}{dt} &= f(x, y) = y + x^2 + O(x^4, y^2, x^2y), \\ \frac{dy}{dt} &= \varepsilon. \end{aligned} \quad (13)$$

This means that the nullcline of system (13) for  $\varepsilon = 0$ , defined by the parabola  $f(x, y) = 0$ , is a good approximation for the nullcline of system (1) in the neighborhood of the singular point  $(0, \Theta_0)$ , or equivalently, in coordinates  $(x, y)$ , the point  $(0, 0)$ . Thus system (13) has the same form as system (2.5) in [48] and can be restated as a Riccati equation:

$$\frac{dx}{dy} = \frac{y + x^2}{\varepsilon}. \quad (14)$$

By performing a change of variable, it can be shown that (14) is equivalent to a second-order Bessel equation. The following function is a general solution of (14):

$$x = \zeta(y) = -\sqrt{v} \frac{J_{-2/3}(2y^{3/2}/3) - c J_{2/3}(2y^{3/2}/3)}{c J_{-1/3}(2y^{3/2}/3) + J_{1/3}(2y^{3/2}/3)}, \quad (15)$$

where  $J_v$  are Bessel functions of the first kind of order  $v$ . The only solution approaching the left branch of the nullcline parabola for  $y < 0$  is the one obtained by choosing  $c = 1$ , thus we pick this value of  $c$ . The inverse of function  $\zeta(y)$ , namely  $\xi(x) = \zeta^{-1}(y)$ , defines the trajectory of  $x$  as a function of  $y$  (Fig. 3). Unfortunately, due to its highly nonlinear form, it is impossible to compute directly.

We use these results together with Theorem 2.1 and Remark 2.11 in [48] in order to derive the following estimate (the result dates back to much earlier, see for example [51]).

**Proposition 1** Let  $y_0 > 0$  and  $x_0 < 0$  satisfy  $f(x_0, y_0) = 0$ . Also fix  $\delta > 0$ . Consider a family of solutions of (13) with initial conditions  $x(0) = x_0 + O(\varepsilon)$  and  $y(0) = y_0$ . Let  $(\delta, h(\varepsilon))$  be the intersection point of this trajectory with the line  $x = \delta$ . Then, for sufficiently small  $\delta$ ,

$$h(\varepsilon) = \Omega_0 \varepsilon^{2/3} + O(\varepsilon \ln \varepsilon), \quad (16)$$

where  $\Omega_0$  stands for the smallest positive zero of the denominator in (15):

$$J_{-1/3}\left(\frac{2y^{3/2}}{3}\right) + J_{1/3}\left(\frac{2y^{3/2}}{3}\right).$$

From now on, we will use the numerical approximation

$$\Omega_0 \approx -2.34.$$

Note that the solution with initial conditions (8), transformed to the coordinates  $(x, y)$ , satisfies the assumptions of Proposition 1. Therefore, estimate (17) holds.

Now let  $T_1$  be the time of the first gamma spike, i.e., when  $\theta_E = \pi$ . From (16), it is easy to see that, after scaling back to the original variables  $(\theta_E, \Theta, \varepsilon_\Theta)$ ,  $T_1$  can be written as

$$T_1 = \frac{\Theta_0 + \pi}{\varepsilon_\Theta \omega} + \frac{C_0}{\varepsilon_\Theta^{1/3}} + O(\ln \varepsilon_\Theta), \quad (17)$$

with

$$C_0 = -\frac{2^{1/3} \Omega_0}{(\omega a)^{1/3}}. \quad (18)$$

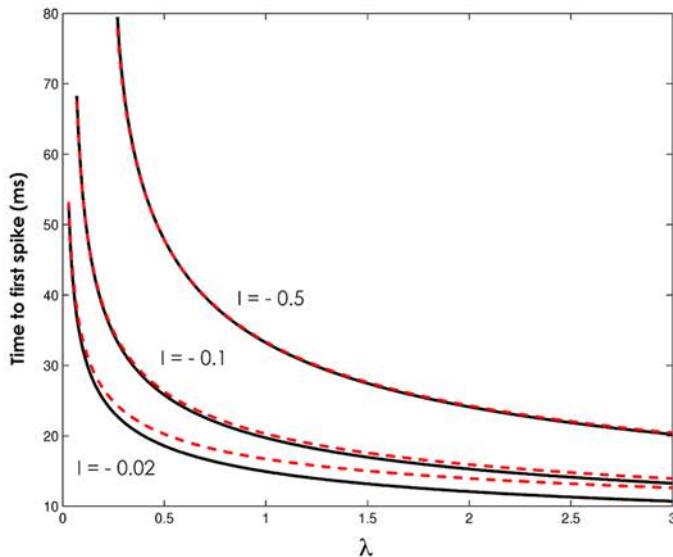
The  $O(\ln \varepsilon_\Theta)$  term in (17) and the following one, of order  $O(1)$  in  $\varepsilon_\Theta$ , happen to be zero in the theta neuron model (as well as in the QIF model) when there is no excitatory feedback from the EG cell to the theta band oscillator (see the Appendix). The next nonzero term in (17) is then of order  $O(\varepsilon_\Theta^{1/3})$ , which represents the error with respect to the time at which the true trajectory of the system reaches  $\theta_E = 2\pi$ . The value of  $\delta$  does not have to be small, on the contrary our approximation works better when  $\delta$  is such that the trajectory of the system is close to the asymptote  $\Theta = C_0 \omega \varepsilon_\Theta^{2/3}$ , as it is the case for the EG cell spike threshold  $\delta = \frac{\pi}{2}$ .

Predictions of Proposition 1 are illustrated in Figs. 3 and 4.

## 5 Subsequent Gamma Spikes, Oscillator Case

In the oscillator case, we assume that inhibition is strong enough to push the system below the SNIC bifurcation, regardless of the value of  $\Theta$ , i.e.,

$$g_{IE} > I_E + 2\lambda. \quad (19)$$



**Fig. 4** Excitable regime—time to first spike. Time to first spike as a function of coupling constant  $\lambda$ , for  $I_E = -0.5$ ,  $I_E = -0.1$  and  $I_E = -0.02$ . Black full line: simulation; red dotted line: analytic solution.  $\varepsilon_\Theta = 0.01$ ,  $\omega = 4$

If the opposite is true, the system does not encounter the bifurcation, since  $\frac{d\theta_E}{dt}$  is always greater than zero, and our analysis cannot be applied to subsequent spikes. We wish to derive an estimate on the number of EG spikes occurring along one  $\Theta$  period, which is given by the time needed for  $\Theta$  to grow from  $-\pi$  to  $\pi$ , thus equal to  $2\pi/(\varepsilon_\Theta \omega)$ . Let  $T_2, \dots, T_L$  be the subsequent gamma spikes and  $\Theta_2, \dots, \Theta_L$ , the corresponding values of  $\Theta$ . Let  $T_j^*$  be the relative time after  $T_{j-1}$  at which the total driving input to the EG neuron reaches zero from negative values:

$$I_E + \lambda(1 + \cos(\Theta_{j-1} + \varepsilon_\Theta \omega T_j^*)) - g_{IE} e^{-\varepsilon_I T_j^*} = 0. \quad (20)$$

From now on, we use the fact that  $\varepsilon_\Theta \approx \varepsilon_I^2$ , and relabel  $\varepsilon_I \equiv \varepsilon$ . Hence, we can write

$$\cos(\Theta_{j-1} + \varepsilon^2 \omega T_j^*) \approx \cos(\Theta_{j-1}) - \sin(\Theta_{j-1}) \varepsilon^2 \omega T_j^*. \quad (21)$$

We expect  $T_j^*$  to be of order  $O(\varepsilon^{-1})$  from (20),  $\cos(\Theta_{j-1})$  is then large compared to  $\sin(\Theta_{j-1}) \varepsilon^2 \omega T_j^*$ . We can thus replace (20) by a simpler formula:

$$e^{\varepsilon T_j^*} = \frac{g_{IE}}{I_E + \lambda(1 + \cos(\Theta_{j-1}))}. \quad (22)$$

Now

$$T_j^* = \frac{1}{\varepsilon} \ln \left( \frac{g_{IE}}{I_E + \lambda(1 + \cos(\Theta_{j-1}))} \right). \quad (23)$$

We denote the time interval between two successive gamma spikes by  $\Delta T_j$  and use the following estimate:

$$\Delta T_j = T_j - T_{j-1} \approx T_j^* + T_j^{**},$$

where

$$T_j^{**} = C_j \varepsilon^{-1/3}, \quad C_j = -\frac{\Omega_0}{(I_E + \lambda(1 + \cos \Theta_{j-1}))^{1/3}}. \quad (24)$$

Estimate (24) is obtained analogously as (17). We can write the modulated instantaneous Interspike Interval (ISI), i.e., the instantaneous period, as

$$\Delta T_j = \frac{1}{\varepsilon} \ln \left( \frac{g_{IE}}{I_E + \lambda(1 + \cos(\Theta_{j-1}))} \right) - \frac{1}{\varepsilon^{1/3}} \frac{\Omega_0}{(I_E + \lambda(1 + \cos \Theta_{j-1}))^{1/3}}. \quad (25)$$

We derive the intrinsic period of PING in absence of any modulation:

$$\Delta T_j^{IN} = \frac{1}{\varepsilon} \ln \left( \frac{g_{IE}}{I_E} \right) - \frac{\Omega_0}{\varepsilon^{1/3} I_E}. \quad (26)$$

After some algebra and performing a second-order Taylor expansion around  $\Theta_{j-1} \approx -\pi$ , i.e., when theta excitation is minimal, the ISI becomes

$$\Delta T_j^{\max} = \Delta T_j^{IN} - \frac{1}{\varepsilon} \frac{\lambda}{2I_E} (\Theta_{j-1} + \pi)^2 - \frac{1}{\varepsilon^{1/3}} \left[ \frac{\Omega_0 \lambda}{6I_E^{4/3}} (\Theta_{j-1} + \pi)^2 \right]. \quad (27)$$

The smallest ISI is obtained by expanding around  $\Theta_{j-1} \approx 0$ , i.e., when theta excitation is maximal:

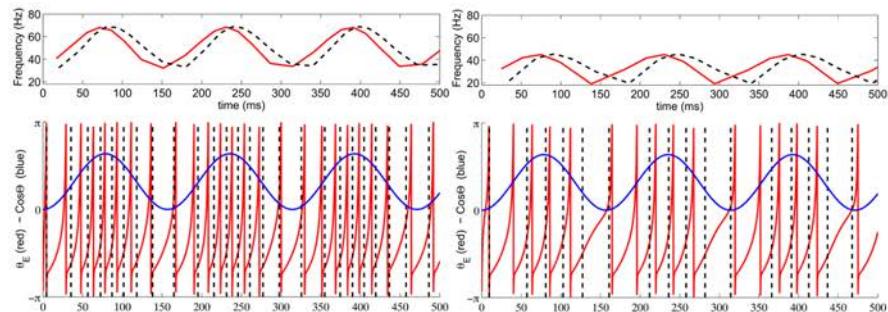
$$\begin{aligned} \Delta T_j^{\min} &= \Delta T_j^{IN} - \frac{1}{\varepsilon} \left[ \ln \left( \frac{I_E + 2\lambda}{I_E} \right) - \frac{\lambda}{2(I_E + 2\lambda)} \Theta_{j-1}^2 \right] \\ &\quad - \frac{1}{\varepsilon^{1/3}} \left[ \Omega_0 \left( \frac{1}{(I_E + 2\lambda)^{1/3}} - \frac{1}{I_E^{1/3}} + \frac{\lambda}{6(I_E + 2\lambda)^{4/3}} \Theta_{j-1}^2 \right) \right], \end{aligned} \quad (28)$$

from (27) and (28) we can estimate respectively the lowest and highest gamma frequencies attained during theta modulation. We then derive an expression for the interval of time  $\Delta T$  between the first to the last gamma spike within half a  $\Theta$ -period:

$$\begin{aligned} \Delta T &= \frac{1}{\varepsilon} \sum_{i=1}^{M_0} \ln \left( \frac{g_{IE}}{I_E + \lambda(1 + \cos(\Theta_{j-1}))} \right) \\ &\quad + \frac{1}{\varepsilon^{1/3}} \sum_{i=1}^{M_0} \frac{-\Omega_0}{(I_E + \lambda(1 + \cos(\Theta_{j-1})))^{1/3}}, \end{aligned} \quad (29)$$

where  $M_0$  is the number of gamma spikes as  $\Theta$  varies between  $-\pi$  and 0. Writing

$$\frac{1}{M_0} \sum_{i=1}^{M_0} \ln \left( \frac{g_{IE}}{I_E + \lambda(1 + \cos(\Theta_{j-1}))} \right) \approx \frac{1}{\pi} \int_{-\pi}^0 \ln \left( \frac{g_{IE}}{I_E + \lambda(1 + \cos(\Theta))} \right) d\Theta$$



**Fig. 5** Oscillatory regime—dynamics. Two cases in which formula (31) gives a correct prediction of the number of gamma spikes and a fair estimate of spike times. *Upper panel:* instantaneous firing frequency of gamma cell obtained from simulation (full red line) and from Eq. (25) (black dotted line). *Lower panel:* the simulation of EG cell membrane potential is shown in red while black dotted lines represent firing times predicted by our analysis; the blue curve shows theta modulation ( $1 + \cos \Theta$ ). Left:  $I_E = 0.5$ ,  $\varepsilon = 0.1$ ,  $\lambda = 0.8$ ,  $g_{IE} = 6$ ,  $\omega = 4$ . Right:  $I_E = 0.1$ ,  $\varepsilon = 0.1$ ,  $\lambda = 0.5$ ,  $g_{IE} = 6$ ,  $\omega = 4$

and

$$\frac{1}{M_0} \sum_{i=1}^{M_0} \frac{-\Omega_0}{(I_E + \lambda(1 + \cos(\Theta_{j-1})))^{1/3}} \approx \frac{1}{\pi} \int_{-\pi}^0 \frac{-\Omega_0}{(I_E + \lambda(1 + \cos(\Theta)))^{1/3}} d\Theta$$

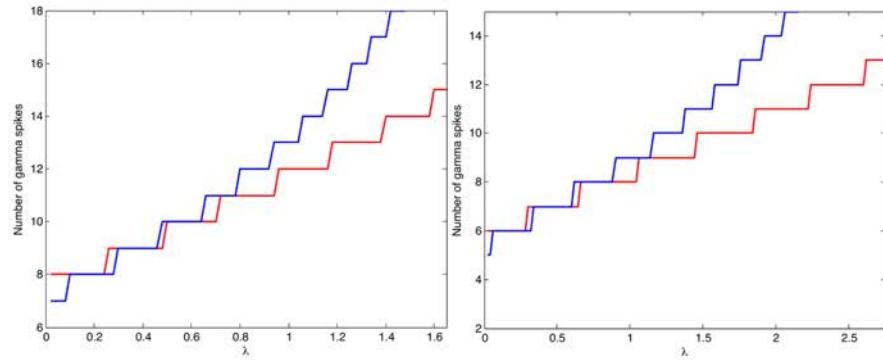
we obtain

$$\begin{aligned} \frac{M_0}{\pi} &\left( \varepsilon \int_{-\pi}^0 \ln \left( \frac{g_{IE}}{I_E + \lambda(1 + \cos(\Theta))} \right) d\Theta \right. \\ &\left. + \varepsilon^{5/3} \int_{-\pi}^0 \frac{-\Omega_0}{(I_E + \lambda(1 + \cos(\Theta)))^{1/3}} d\Theta \right) \approx \frac{\pi}{\omega}. \end{aligned} \quad (30)$$

As  $M_0$  provides an estimate for the number of gamma spikes as  $\Theta$  grows from  $-\pi$  to 0, the total number of gamma spikes  $M$  is the largest integer such that

$$\begin{aligned} \frac{M}{2\pi} &\left( \varepsilon \int_{-\pi}^{\pi} \ln \left( \frac{g_{IE}}{I_E + \lambda(1 + \cos(\Theta))} \right) d\Theta \right. \\ &\left. + \varepsilon^{5/3} \int_{-\pi}^{\pi} \frac{-\Omega_0}{(I_E + \lambda(1 + \cos(\Theta)))^{1/3}} d\Theta \right) < \frac{2\pi}{\omega}. \end{aligned} \quad (31)$$

This formula works well, especially when inhibition is sufficiently strong. Figure 5 shows two cases where the formula gives the exact prediction of the number of gamma spikes and a good approximation of spike times. It is worth to mention that in the oscillatory case there is no phase reset at the end of a theta cycle, meaning that the initial conditions are never the same at the beginning of a theta oscillation. As a consequence, the result in (31) does not hold as a rigorous solution but as an average estimate, and the exact number of spikes can still vary over different trials. In Fig. 6, we show the direct comparison between the predictions of the formula and the simulation, as a function of  $\lambda$ . Note that for  $\lambda$  very small the estimate of the formula is too big. There we would need to include more terms in the  $\varepsilon$  expansion to get a



**Fig. 6** Oscillatory regime—number of spikes. Number of spikes in the simulation (blue) and the prediction of the formula as a function of  $\lambda$  varying from 0 to  $(g_{IE} - I_E)/2$ ,  $\omega = 4$ . In the left panel  $I_E = 0.7$ ,  $\varepsilon = 0.1$ ,  $g_{IE} = 4$ . In the right panel  $I_E = 0.5$ ,  $\varepsilon = 0.1$ ,  $g_{IE} = 6$

more accurate prediction. When  $\lambda$  is large, for fixed  $g_{IE}$ , the positive input is such that inhibition is not sufficient to periodically time the spikes. As a consequence, the estimate of formula (31) becomes too small.

## 6 Subsequent Gamma Spikes, Excitable Case

### 6.1 Second Gamma Spike

Let  $T_2^*$ , be defined by

$$I_E + \lambda(1 + \cos(\Theta(T_1 + T_2^*))) - g_{IE}e^{-\varepsilon T_2^*} = 0. \quad (32)$$

Note that

$$\Theta(T_1 + T_2^*) = \Theta_1 + \omega\varepsilon^2 T_2^* = \Theta_0 + \omega\varepsilon^2 T_2^* + O(\varepsilon^{4/3}) + O(\varepsilon^2 \ln \varepsilon^2). \quad (33)$$

but, similarly to the oscillator case, we expect  $T_2^*$  to be of order  $O(\varepsilon^{-1})$  from (32). This allows us to neglect the last two terms on the RHS of (33). Hence, by (32) and (33), we have

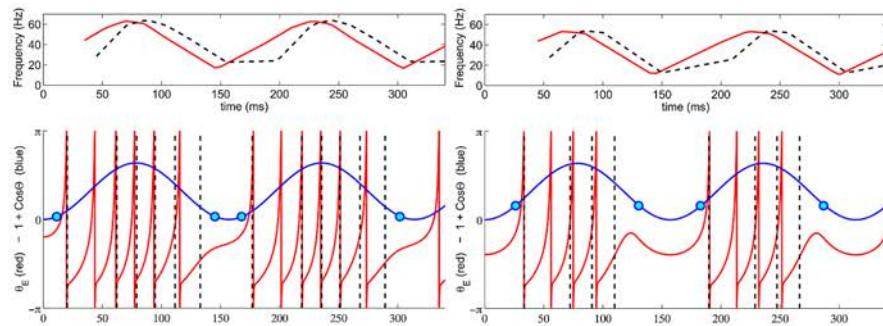
$$e^{\varepsilon T_2^*} = \frac{g_{IE}}{\lambda(-\sin(\Theta_0)\varepsilon^2\omega T_2^* + O(\varepsilon^{4/3}))}. \quad (34)$$

Further, we write

$$\varepsilon T_2^* = -\ln \varepsilon - \ln(-\ln \varepsilon) + A, \quad (35)$$

and substitute into (34) getting

$$\begin{aligned} -\frac{1}{\varepsilon \ln \varepsilon} e^A &= \frac{g_{IE}}{-\lambda \sin(\Theta_0)\omega \varepsilon (-\ln \varepsilon - \ln(-\ln \varepsilon) + A)} \\ &= \left(-\frac{1}{\varepsilon \ln \varepsilon}\right) \left(\frac{g_{IE}}{-\lambda \sin(\Theta_0)\omega}\right) \frac{1}{(1 - (-\ln(-\ln \varepsilon)) + A)/\ln \varepsilon} \end{aligned}$$



**Fig. 7** Excitable regime—dynamics. Three cases for which formula (31) gives a correct prediction of the number of gamma spikes. Plot colors as in Fig. 4.  $\Theta_0$  and  $2\pi - \Theta_0$ , i.e. the theta phases where the first and last Hopf bifurcation approximately take place, are shown in cyan.  $\omega = 4$ ,  $\varepsilon = 0.1$  and  $g_{IE} = 6$ . Left panel:  $I_E = -0.1$ ,  $\lambda = 1$ . Right panel:  $I_E = -0.5$ ,  $\lambda = 1$

$$\approx \left( -\frac{1}{\varepsilon \ln \varepsilon} \right) \left( \frac{g_{IE}}{-\lambda \sin(\Theta_0) \omega} \right) \left( 1 - \frac{\ln(-\ln(\varepsilon))}{\ln \varepsilon} + \frac{A}{\ln \varepsilon} \right); \quad (36)$$

$$e^A \approx -\frac{g_{IE}}{\lambda \sin(\Theta_0) \omega} + O\left( -\frac{\ln(-\ln(\varepsilon))}{\ln \varepsilon} \right). \quad (37)$$

It then follows from (35) that keeping only the leading terms:

$$T_2 \approx -\frac{\ln \varepsilon}{\varepsilon} + T_1 \quad \text{and} \quad \Theta_2 \approx -\omega \varepsilon \ln \varepsilon + \Theta_1. \quad (38)$$

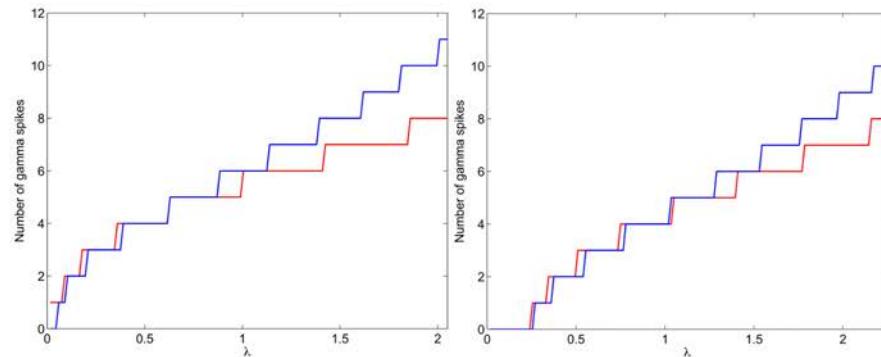
## 6.2 Subsequent Gamma Spikes

We can write a variant of estimate (31) for the excitable case:

$$\begin{aligned} & \frac{M}{2\Theta_0} \left( \varepsilon \int_{\Theta_0}^{2\pi - \Theta_0} \ln \left( \frac{g_{IE}}{I_E + \lambda(1 + \cos(\Theta))} \right) d\Theta \right. \\ & \left. + \varepsilon^{5/3} \int_{\Theta_0}^{2\pi - \Theta_0} \frac{1}{(I_E + \lambda(1 + \cos(\Theta)))^{1/3}} d\Theta \right) < \frac{2\Theta_0}{\omega}, \end{aligned} \quad (39)$$

where now the extremes in the integrals are chosen to be the times of the first and last gamma spikes (i.e., the times when the EG neuron crosses the SNIC bifurcation respectively from below and above), assuming that these would be approximately symmetric with respect to the  $\Theta$  cycle.

This formula works adequately for large inhibition and relatively small (negative)  $I_E$ . Otherwise, due to the intricate interplay between the growth of  $\Theta$  and the decay of  $s$  almost to 0 (witnessed in the computation of  $T_2$ ), it is not sufficient to have just the lowest terms of the  $\varepsilon$  expansion of  $\Delta T_j$ . Figure 7 shows two cases where the formula gives the exact prediction of the number of gamma spikes. In Fig. 8, we show the direct comparison between the predictions of the formula and the simulation, as a function of  $\lambda$ . For  $\lambda$  large inhibition is too weak to time the spikes and the estimate of the formula becomes too small.



**Fig. 8** Excitable regime—number of spikes. Number of spikes in the simulation (blue) and the prediction of the formula as a function of  $\lambda$ . *Left panel:*  $I_E = -0.1$ ,  $g_{IE} = 6$ ,  $\varepsilon = 0.1$ ,  $\omega = 4$ . *Right panel:*  $I_E = -0.5$ ,  $g_{IE} = 6$ ,  $\varepsilon = 0.1$ ,  $\omega = 4$

## 7 Conclusions and Future Directions

In this paper, we investigated how a continuous, strong, low frequency (1–10 Hz) modulation determines the spiking properties of a simplified PING oscillator. This work has been particularly motivated by recent investigation on the role of theta-gamma interactions in processing speech signals [52]. Syllabic input are in fact known to possess a quasiperiodic structure matching theta frequency [24]. Within this framework, theta-modulated gamma spikes need to be aligned to the onset and the offset of linguistically relevant chunks [23]. It is then crucial to understand the timing of gamma spikes and the way they are influenced by theta input, since theta is supposed to detect the presence of long timescale syllabic content. It remains to be unveiled whether the scaling we analytically determined here is produced in more realistic models for speech processing [52] currently under development. Indeed, this result could also be used for other purposes: investigating how theta fluctuations modulate gamma firing in the hippocampus; determining the impact of alpha oscillations on higher frequencies (including gamma), which are thought to carry bottom-up information in visual perception. Indeed timing of first spike is assumed to be particularly relevant in visual cortex, since it has been shown that it would facilitate the neural encoding of stimuli [53].

To explore the dynamics of the system, we split the problem into two parameter regimes: In the first, the frequency of gamma spikes is only modulated by theta phase, while in the second the gamma cell would only fire if forced by theta input. In the former regime, by restating the problem in form of a Mathieu differential equation and looking at the first zero of the Mathieu function solving the initial value problem, we were able to find the time to first gamma spike. In the latter, we separate the dynamics into three time scales, one characterizing EG neuron dynamics in absence of any external input, one for theta dynamics, and one for synaptic inhibition, and we approximate the time to first spike by using an extension of the geometric singular perturbation theory based on the application of the blow-up method [46, 48].

Computations align with the intuition (arising from the fact that  $\theta_E$  is a type I neuron) that time to first spike decreases in both cases with coupling strength  $\lambda$  and

constant driving current  $I_E$ . Interestingly, in the excitable case, we found that time to first spike depends approximately on  $\lambda^{-1/6}$ , which implies that it saturates rapidly as  $\lambda$  grows. As a second notable result,  $T_1$  scales as  $c_0/\varepsilon_\Theta + c_1/\varepsilon_\Theta^{1/3} + O(\ln(\varepsilon_\Theta))$ ,  $\varepsilon_\Theta$  being the speed of theta cycle. Building on these results, we subsequently computed the time to successive spikes in both regimes, where inhibitory synaptic decay time becomes an important factor. For both regimes were able to compute approximate spike times and predict the exact number of spikes per theta cycle (and instantaneous frequency of firing as a direct consequence) in a range of parameter values that leads to firing within the gamma frequency band.

In the present work, we analyzed a simple system in which coupling was limited to a feedforward theta-gamma connection. It would be a natural next step to extend the analysis to bidirectional coupling by including a feedback from gamma spikes to the  $\Theta$  oscillator. A second assumption we made in constructing our system stated that the gamma circuit internal delay between excitatory and inhibitory spikes was negligible, meaning that both cells would fire at exactly the same time. To make the model more biologically appealing, one could relax this hypothesis by introducing a synaptic delay after an excitatory spike and study the correspondent system (i.e., a full PING). For relatively short delays, we would expect the results obtained in this paper to hold at least qualitatively. Throughout this paper, we considered gamma to be a simplified PING generator, on the other hand it still remains an open question whether the same characteristics of theta-gamma modulation we explored here would still be found in a different gamma generator, e.g., an Interneuron Network Gamma (ING) network [54] that can still be implemented with Type I neurons as in the case of this work.

### Competing Interests

The authors declare that they have no competing interests.

### Authors' Contributions

LF, MK, AH, BG designed research; LF, MK performed research; LF, MK, AH, BG wrote the manuscript.

**Acknowledgements** LF, AH are supported by the European Research Council; AH and BG are partially supported by ANR “Neurobot” and “Dopanic”; BG is supported by CNRS and INSERM and LABEX “IEC”; MK supported by INRIA.

### Appendix

We show that the term of order  $O(\ln \varepsilon_\Theta)$  in expansion (17) is zero in our model. The subsequent term, of order  $O(1)$  is also zero, but we do not include the result here since computations are long and heavy. The interested reader could derive this result from [51].

In the absence of any synaptic input from the inhibitory neuron, we restate system (1) to reduce the complexity of notation:

$$\begin{aligned}\frac{dx}{dt} &= \varphi(x, y) \equiv (1 - \cos(x)) + (I_E + \lambda(1 + \cos(y)))(1 + \cos(y)), \\ \frac{dy}{dt} &= \varepsilon\psi(x, y) \equiv \varepsilon\omega.\end{aligned}\quad (40)$$

Then, following [51], we write the expansion

$$T_1 = \frac{\Theta_0 + \pi}{\varepsilon_\Theta \omega} + \frac{C_0}{\varepsilon_\Theta^{1/3}} + D_0 \ln \varepsilon_\Theta + E_0 + O(\varepsilon_\Theta^{1/3}), \quad (41)$$

which approximates the time to the first spike in the excitable case.

Coefficient  $D_0$  in (41) might be written as

$$\begin{aligned}D_0 &= \frac{1}{3} \frac{1}{\psi(S)} \frac{6\varphi_{xx}(S)\psi_x(S) - 2\varphi_{xxx}(S)\psi(S)}{3\varphi_{xx}^2(S)} \\ &\quad - \frac{1}{6} |\psi(S)| \frac{\psi_x(S)\sqrt{2/|\varphi_{xx}(S)\varphi_y(S)|}}{\psi^2(S)\sqrt{|\varphi_{xx}(S)/(2\varphi_y(S))|}} \operatorname{sign} \varphi_y(S),\end{aligned}\quad (42)$$

where  $S$  stands for the coordinates of the singular point  $S = (0, \Theta_0)$  and subscripts indicate the derivatives, i.e.,  $\varphi_x(S)$  is the first derivative of  $\varphi$  with respect to  $x$ , taken at  $(0, \Theta_0)$ . It is easy to verify that any derivative of  $\varphi$  with respect to  $x$  of order  $n$ , for  $n$  odd, is equal to zero at  $S$ . Furthermore, since  $\psi(x, y)$  is constant in system (40),  $\psi_x(S)$  is clearly zero. Hence,  $D_0 = 0$ .

## References

1. Buzsaki G, Draguhn A: **Neuronal oscillations in cortical networks.** *Science* 2004, **304**:1926-1929.
2. Wang XJ: **Neurophysiological and computational principles of cortical rhythms in cognition.** *Physiol Rev* 2010, **90**:1195-1268.
3. Womelsdorf T, Schöffelen J, Oostenveld R, Singer W, Desimone R, Engel AK, Fries P: **Modulation of neuronal interactions through neuronal synchronization.** *Science* 2007, **316**:1609-1612.
4. Young CK, Eggermont JJ: **Coupling of mesoscopic brain oscillations: recent advances in analytical and theoretical perspectives.** *Prog Neurobiol* 2009, **89**(1):61-78.
5. Schöffelen J, Oostenveld R, Fries P: **Neuronal coherence as a mechanism of effective corticospinal interaction.** *Sci Signal Transduct Knowl Environ* 2005, **308**(5718):111.
6. Huerta PT, Lisman JE: **Bidirectional synaptic plasticity induced by a single burst during cholinergic theta oscillation in CA1 in vitro.** *Neuron* 1995, **15**(5):1053-1063.
7. Fell J, Klaver P, Lehnertz K, Grunwald T, Schaller C, Elger C, Fernández G: **Human memory formation is accompanied by rhinal-hippocampal coupling and decoupling.** *Nat Neurosci* 2001, **4**:1259-1264.
8. Jensen O, Kaiser J, Lachaux JP: **Human gamma-frequency oscillations associated with attention and memory.** *Trends Neurosci* 2007, **30**(7):317-324.
9. Fries P, Reynolds J, Rorie A, Desimone R: **Modulation of oscillatory neuronal synchronization by selective visual attention.** *Science* 2001, **291**(5508):1560-1563.
10. Gray CM, König P, Engel A, Singer W: **Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties.** *Nature* 1989, **338**(6213):334-337.

11. Bragin A, Jandó G, Nádasdy Z, Hetke J, Wise K, Buzsáki G: **Gamma (40–100 Hz) oscillation in the hippocampus of the behaving rat.** *J Neurosci* 1995, **15**:47–60.
12. Jensen O, Colgin LL: **Cross-frequency coupling between neuronal oscillations.** *Trends Cogn Sci* 2007, **11**(7):267–269.
13. Arnal LH, Morillon B, Kell CA, Giraud AL: **Dual neural routing of visual facilitation in speech processing.** *J Neurosci* 2009, **29**(43):13445–13453.
14. Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE: **An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex.** *J Neurophysiol* 2005, **94**(3):1904–1911.
15. Canolty RT, Edwards E, Dalal SS, Soltani M, Nagarajan SS, Kirsch HE, Berger MS, Barbaro NM, Knight RT: **High gamma power is phase-locked to theta oscillations in human neocortex.** *Science* 2006, **313**(5793):1626.
16. Morillon B, Liégeois-Chauvel C, Arnal LH, Bénar CG, Giraud AL: **Asymmetric function of theta and gamma activity in syllable processing: an intra-cortical study.** *Front Psychol* 2012, **3**:248.
17. Belluscio MA, Mizuseki K, Schmidt R, Kemper R, Buzsaki G: **Cross-frequency phase-phase coupling between  $\theta$  and  $\gamma$  oscillations in the hippocampus.** *J Neurosci* 2012, **32**(2):423–435.
18. Palva S, Palva JM: **The functional roles of alpha-band phase synchronization in local and large-scale cortical networks.** *Front Psychol* 2011, **2**:204.
19. Scheffer-Teixeira R, Belchior H, Caixeta FV, Souza BC, Ribeiro S, Tort ABL: **Theta phase modulates multiple layer-specific oscillations in the CA1 region.** *Cereb Cortex* 2012, **22**(10):2404–2414.
20. Wulff P, Ponomarenko AA, Bartos M, Korotkova TM, Fuchs EC, Bähner F, Both M, Tort ABL, Kopell NJ, Wisden W, Monyer H: **Hippocampal theta rhythm and its coupling with gamma oscillations require fast inhibition onto parvalbumin-positive interneurons.** *Proc Natl Acad Sci USA* 2009, **106**(9):3561–3566.
21. Buzsaki G, Buhl DL, Harris KD, Csicsvari J, Czéh B, Morozov A: **Hippocampal network patterns of activity in the mouse.** *Math Eng Ind* 2003, **116**:201–211.
22. Ghitza O: **Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm.** *Front Psychol* 2011, **2**:130.
23. Giraud AL, Poeppel D: **Cortical oscillations and speech processing: emerging computational principles and operations.** *Nat Neurosci* 2012, **15**(4):511–517.
24. Shamir M, Ghitza O, Epstein S, Kopell NJ: **Representation of time-varying stimuli by a network exhibiting oscillations on a faster time scale.** *PLoS Comput Biol* 2009, **5**(5):e1000370.
25. Bressloff PC, Coombes S: **Dynamics of strongly-coupled spiking neurons.** *Neural Comput* 2000, **12**:91–129.
26. Brunel N, Hakim V: **Fast global oscillations in networks of integrate-and-fire neurons with low firing rates.** *Neural Comput* 1999, **11**(7):1621–1671.
27. Ermentrout GB, Chow CC: **Modeling neural oscillations.** *Physiol Behav* 2002, **77**(4–5):629–633.
28. Ko TW, Ermentrout G: **Phase-response curves of coupled oscillators.** *Phys Rev E* 2009, **79**:1–6.
29. Kopell N, Ermentrout G: **Mechanisms of phase-locking and frequency control in pairs of coupled neural oscillators.** In *Handbook of Dynamical Systems. Volume 2*. Amsterdam: Elsevier; 2002:3–54.
30. Ostojic S, Brunel N, Hakim V: **Synchronization properties of networks of electrically coupled neurons in the presence of noise and heterogeneities.** *J Comput Neurosci* 2009, **26**(3):369–392.
31. Tiesinga PH, Sejnowski TJ: **Mechanisms for phase shifting in cortical networks and their role in communication through coherence.** *Front Human Neurosci* 2010, **4**:196.
32. Kilpatrick ZP, Ermentrout B: **Sparse gamma rhythms arising through clustering in adapting neuronal networks.** *PLoS Comput Biol* 2011, **7**(11):e1002281.
33. Tiesinga P, Sejnowski TJ: **Cortical enlightenment: are attentional gamma oscillations driven by ING or PING?** *Neuron* 2009, **63**(6):727–732.
34. Ermentrout G, Kopell N: **Parabolic bursting in an excitable system coupled with a slow oscillation.** *SIAM J Appl Math* 1986, **46**(2):233–253.
35. Lewis TJ, Rinzel J: **Dynamics of spiking neurons connected by both inhibitory and electrical coupling.** *SIAM J Appl Math* 2003, **14**:283–309.
36. Yu N, Kuske R, Li YX: **Stochastic phase dynamics and noise-induced mixed-mode oscillations in coupled oscillators.** *Chaos* 2008, **18**:015112.
37. Ledoux E, Brunel N: **Dynamics of networks of excitatory and inhibitory neurons in response to time-dependent inputs.** *Front Comput Neurosci* 2011, **5**:25.
38. Vierling-Claassen D, Kopell N: **The dynamics of a periodically forced cortical microcircuit, with an application to schizophrenia.** *SIAM J Appl Dyn Syst* 2009, **8**(2):710.

39. Tort ABL, Rotstein HG, Dugladze T, Gloveli T, Kopell NJ: **On the formation of gamma-coherent cell assemblies by oriens lacunosum-moleculare interneurons in the hippocampus.** *Proc Natl Acad Sci USA* 2007, **104**(33):13490-13495.
40. Börgers C, Kopell NJ: **Effects of noisy drive on rhythms in networks of excitatory and inhibitory neurons.** *Neural Comput* 2005, **17**(3):557-608.
41. Kopell NJ, Börgers C, Pervouchine D, Malerba P, Tort ABL: **Gamma and theta rhythms in biophysical models of hippocampal circuits.** In *Hippocampal Microcircuits*. New York: Springer; 2010:423-457.
42. Roopun AK, Kramer M, Carracedo LM, Kaiser M, Davies CH, Traub RD, Kopell NJ, Whittington M: **Temporal interactions between cortical rhythms.** *Front Neurosci* 2008, **2**(2):145-154.
43. Schroeder CE, Lakatos P: **Low-frequency neuronal oscillations as instruments of sensory selection.** *Trends Neurosci* 2009, **32**:9-18.
44. Ruby L: **Applications of the Mathieu equation.** *Am J Phys* 1996, **64**:39.
45. Fenichel N: **Geometric singular perturbation theory.** *J Differ Equ* 1979, **31**:53-98.
46. van Gils SA, Krupa M, Szmolyan P: **Asymptotic expansions using blow-up.** *Z Angew Math Phys* 2005, **56**(3):369-397.
47. Krupa M, Szmolyan P: **Relaxation oscillations and canard explosion.** *J Differ Equ* 2001, **174**:312-368.
48. Krupa M, Szmolyan P: **Extending geometric singular perturbation theory to nonhyperbolic points—fold and canard points in two dimensions.** *SIAM J Math Anal* 2001, **33**:286-314.
49. Ermentrout B: **Type I membranes, phase resetting curves, and synchrony.** *Neural Comput* 1996, **8**:979-1001.
50. Tonneller A, Belmabrouk H, Martinez D: **Event-driven simulations of nonlinear integrate-and-fire neurons.** *Neural Comput* 2007, **19**(12):3226-3238.
51. Mishchenko EF, Rozov NK: *Differential Equations with Small Parameters and Relaxation Oscillations*. New York: Plenum Press; 1980.
52. Hyafil A, Gutkin B, Giraud AL: **A theoretical exploration of speech/neural oscillation alignment for speech parsing.** In *Front. Hum. Neurosci. Conference Abstract: XI International Conference on Cognitive Neuroscience (ICON XI)*; 2011.
53. Thorpe S, Delorme A, Van Rullen R: **Spike-based strategies for rapid processing.** *Neural Netw* 2001, **14**(6-7):715-725.
54. Wang XJ, Buzsaki G: **Gamma oscillation by synaptic inhibition in a hippocampal interneuronal network model.** *J Neurosci* 1996, **16**:6402-6413.

### 11.3 The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex



#### ARTICLE

Received 19 May 2014 | Accepted 14 Jul 2014 | Published 2 Sep 2014

DOI: 10.1038/ncomms5694

OPEN

## The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex

L. Fontolan<sup>1,\*</sup>, B. Morillon<sup>2,\*</sup>, C. Liegeois-Chauvel<sup>3</sup> & Anne-Lise Giraud<sup>1</sup>

The fact that feed-forward and top-down propagation of sensory information use distinct frequency bands is an appealing assumption for which evidence remains scarce. Here we obtain human depth recordings from two auditory cortical regions in both hemispheres, while subjects listen to sentences, and show that information travels in each direction using separate frequency channels. Bottom-up and top-down propagation dominates in  $\gamma$ - and  $\delta$ - $\beta$  ( $<40$  Hz) bands, respectively. The predominance of low frequencies for top-down information transfer is confirmed by cross-regional frequency coupling, which indicates that the power of  $\gamma$ -activity in A1 is modulated by the phase of  $\delta$ - $\beta$  activity sampled from association auditory cortex (AAC). This cross-regional coupling effect is absent in the opposite direction. Finally, we show that information transfer does not proceed continuously but by time windows where bottom-up or top-down processing alternatively dominates. These findings suggest that the brain uses both frequency- and time-division multiplexing to optimize directional information transfer.

<sup>1</sup>Department of Neuroscience, University of Geneva, Biotech Campus, 9, Chemin des Mines, Geneva 1211, Switzerland. <sup>2</sup>Department of Psychiatry, Columbia University Medical Center, New York, New York 10032, USA. <sup>3</sup>INSERM U1106—Institut de Neurosciences des Systèmes, Université Aix-Marseille, Marseille 13005, France. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to A.-L.G. (email: anne-lise.giraud@unige.ch).

**A** popular view of brain functioning is that the central neural system minimizes its reaction to environmental stimuli by predicting probable events and inferring their most probable causes<sup>1–3</sup>. This mechanism ensures that reactions are appropriate, that is, maximal for unexpected events and minimal to frequent ones, irrespective of their magnitude. Functionally, predictive coding is a possible realization of the anticipatory function of the brain<sup>4,5</sup>. It assumes that the mismatch between descending and ascending information is assessed at each processing level, possibly at the cortical column scale<sup>6</sup>, so that only the error signal is further propagated. This theory is seducing because it offers a parsimonious computational mechanism to flexibly minimize stimulus-driven information, depending on the stage where stimulus features are anticipated, from sensory to high-level action representations<sup>7–9</sup>. The predictive coding framework potentially accommodates a number of well-known psychophysical and macroscopic neurophysiological phenomena, such as priming, repetition suppression and mismatch negativity<sup>3,10–12</sup>. At the biophysical and mechanistic levels, however, there are virtually no data showing how predictive coding could operate. Only a few theoretical proposals attempt to describe how predictions and prediction errors are computed and how information is being transferred in each direction<sup>6</sup>.

One of those stipulates that ascending and descending information could be conveyed via distinct frequency channels, the  $\gamma$ - and  $\beta$ -channels for up- and downgoing information, respectively<sup>7,13</sup>. This would imply that the brain uses multiplexing<sup>14,15</sup> as a means to transmit signals of different nature and content in parallel and opposite directions. In particular, top-down (T-D)  $\beta$ -activity could provide a modulatory gain on lower-tier  $\gamma$ -activity<sup>16,17</sup>. In the context of predictive coding a frequency dissociation between bottom-up (B-U) and T-D information transfer, with slower rates for T-D mechanisms, could be accounted for by the fact that T-D-propagated signals (predictions) follow from the linear accumulation of prediction errors<sup>6</sup>.

Experimental evidence of a spectral dissociation for up- and downgoing information remains scarce and not unequivocal. When probing two hierarchical regions of the monkey visual cortex during a spatial attention task, the  $\gamma$ -up/ $\beta$ -down scheme was only partly confirmed<sup>18</sup>. Moreover, even if spectral multiplexing would keep B-U and T-D information apart, a scheme with only two modulation bands ( $\beta$  and  $\gamma$ ) and one carrier (high- $\gamma$ ) might be underspecified<sup>19</sup>.

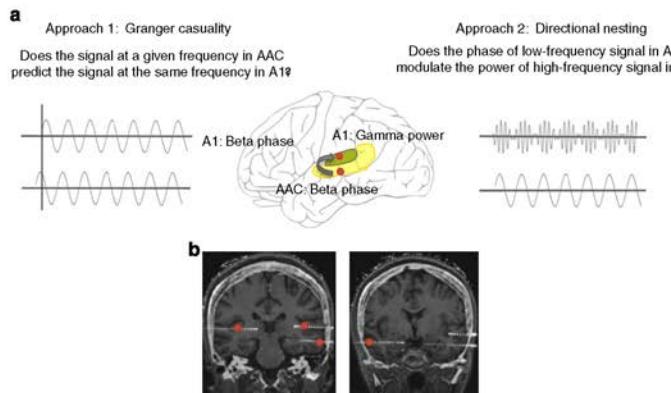
Here we explored whether the human brain propagates B-U and T-D signals using distinct frequency bands. We used human depth electrode recordings made at locations corresponding to two successive steps in speech processing<sup>20</sup> (Fig. 1). We examined the directionality of cross-regional interactions using non-parametric Granger causality (GC)<sup>21,22</sup> to establish the dominant direction of information flow within specific frequency bands, and cross-regional phase-amplitude coupling to address whether high-frequency power in one region was modulated as a function of the phase of low-frequency activity in the other one (see Fig. 1a). By restricting cross-regional coupling results to those frequency domains where causality is significant, we show that B-U and T-D information transfer dominates in high- and low frequency-domains, respectively, and proceed in a discontinuous fashion, suggesting both frequency and time division are used to unmix up- and downgoing information in the brain.

## Results

**Characterization of intra-cortical responses to speech.** We present results from three epileptic subjects who underwent a

simple experiment during which they passively listened to spoken sentences. The protocol was part of a larger one involving more speech material (syllables and words)<sup>23</sup> aiming at investigating speech-specific responses throughout the temporal cortex. We focused on the sentences data set and analysed intracortical activity from depth macroelectrodes located in primary (A1) and association auditory cortex (AAC) regions (Fig. 1b) in response to passive listening of 110 repetitions of 2.5-s long sentences. We assumed that spoken sentences had sufficient duration and complexity to engage B-U and T-D information transfer. One of the three subjects (S1) was implanted bilaterally in the two temporal lobes at identical functional locations (Fig. 1b). The two other subjects were implanted in the left (S2) or the right (S3) temporal lobe, for a total of two subjects per hemisphere and location. The limited number of subjects is inherent to the method as depth electrodes are rarely inserted in humans' auditory cortex, and more rarely at equivalent functional sites in the left and right hemispheres. We analysed signals from the electrode contacts showing the most typical evoked electrophysiological landmarks of A1 and AAC (typical latencies and auditory-evoked response shapes; see Methods). The corticograms obtained for all subjects at each location confirmed distinct response patterns in A1 versus AAC regions (Fig. 2a–d, upper panels). Cross-correlating the wideband stimulus envelope<sup>24,25</sup> with the cortical responses (see Methods) emphasized the functional distinction across hierarchical levels (Fig. 2a–d, lower panels). In A1  $\gamma$ -activity correlated with the speech acoustic structure, whereas, consistent with a more integrated function of higher-level regions,  $\gamma$ -activity in AAC (left dominant) was largely induced and independent from fast acoustic modulations. Frequency-specific interactions between the acoustic speech signal and the intracortical electroencephalography (iEEG) were also observed in the  $\delta/\theta$ -band in all regions, and in the low  $\beta$ -band (around 12–14 Hz) in all regions except in the right AAC. Finally, in accordance with a weaker specialization for speech of right temporal regions, stimulus/brain cross-correlations were overall stronger in the left than in the right A1 (Fig. 2a,b, lower panels), and  $\gamma$ -induced responses in AAC were drastically left dominant (Fig. 2c,d, upper panels).

**Cross-regional GC.** We examined directed functional connectivity between A1 and AAC across the whole iEEG spectrum. We compared propagation directions by computing non-parametric GC in the frequency domain<sup>21,22</sup> on a trial-by-trial basis for each subject (see Methods). We further averaged the GC time-frequency (TF) patterns across time, trials and sentences. In both the B-U (A1 causing AAC activity) and the T-D (AAC causing A1) analyses, we found several GC peaks distributed between 1 and 140 Hz (Supplementary Fig. 1 and Table 1 for details in the 1–20 Hz range). We observed that GC values were overall larger at low than at high frequencies. This global decrease in GC values might be related to the power law decay in the amplitude spectrum of brain activity<sup>26</sup>, as GC is sensitive to asymmetries in the power spectra. Critically, we found a dominance of T-D GC in the low part of the spectrum (<40 Hz) and of B-U GC at high frequencies (>40 Hz). This effect was confirmed using two complementary statistical approaches (Fig. 3a,b versus Fig. 3c), and the spectral division of B-U and T-D was consistent across subjects (and hemispheres). Although B-U dominance above 40 Hz is in line with the hypothesis that the brain mostly uses the  $\gamma$ -channel to propagate sensory information forward, T-D GC dominance was not limited to the  $\beta$ -range<sup>6,7,13</sup>, but broadly covered the whole  $\delta$ – $\beta$ -range. In the left hemisphere, we additionally found several discrete GC peaks in the B-U direction within the 1–20 Hz range. The B-U and T-D GC peaks did not align across subjects (Fig. 3, Supplementary Fig. 1, insets and Table 1), yet in each individual



**Figure 1 | Experimental approach and electrodes position.** (a) Experimental approach and hypotheses. We explored the processing of speech in auditory cortex through two distinct tests: GC, which allows for testing causal relationship between different regions within the same frequency band; and directional phase-amplitude coupling that examines phase-power dependencies both across brain areas and across frequencies.  $\beta$ - and  $\gamma$ -frequencies were of particular interest in view of our working hypotheses. (b) Example of electrode positioning. In S1, electrodes were positioned at equivalent locations on each hemisphere in A1 and auditory association cortex (AAC). The electrode contacts used along the shaft were selected based on their anatomical location and functional responses (typical shape and latencies of evoked responses<sup>66</sup>).

B-U and T-D directions appeared spectrally non-overlapping (Supplementary Fig. 1, insets): B-U GC peaks aligned to the troughs of T-D GC peaks and vice versa. Such a frequency splitting might indicate that specific sub-ranges within the  $\delta$ - $\beta$  domain ( $<40$  Hz) specialize in directional information transfer, even though T-D overall dominates in this frequency range. Altogether, these results are consistent with the hypothesis that the brain uses distinct frequency channels to propagate feed-forward and feedback information. However, the picture arising from the GC data appears more complex than a simple  $\gamma$ -up/ $\beta$ -down scheme and also involves lower frequencies<sup>6,7,13</sup>.

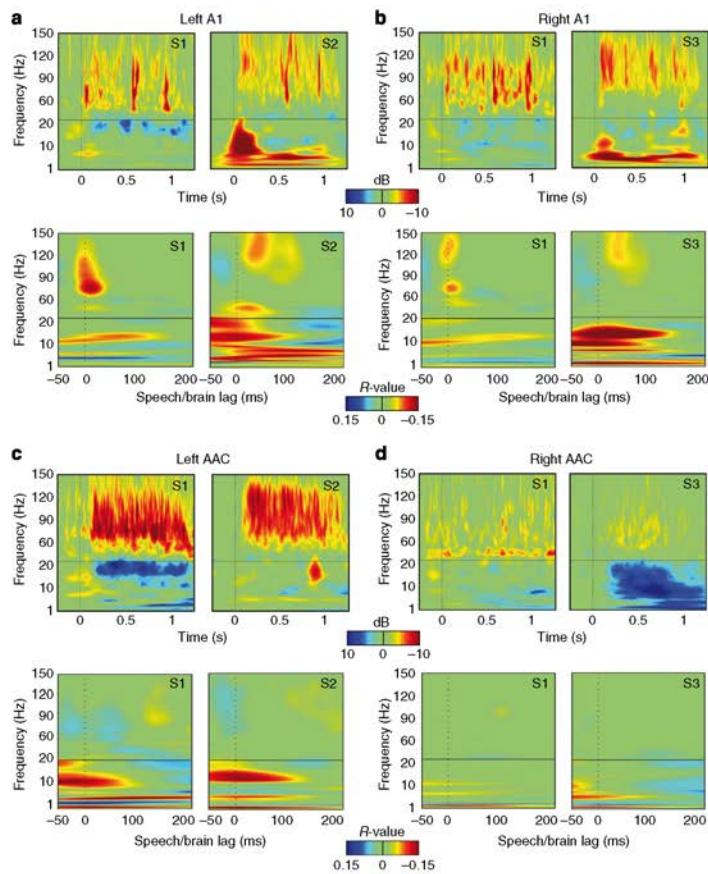
**Cross-regional phase-amplitude coupling.** GC indicated the predominant direction of information transfer, but did not provide information about the influence ascending or descending signals may locally exert on neural activity. We therefore explored whether low-frequency T-D rhythms influenced distant (for example, AAC influence on A1)  $\gamma$ -power changes, reasoning that efficient information transfer should modulate the timing and/or the amount of population spiking at target level<sup>27,28</sup>. In particular, we tested whether the GC peaks observed in the low-frequency range (1–20 Hz) were associated with distant  $\gamma$ -power changes. To characterize the influence one region exerted on the region hierarchically below or above, we examined cross-regional nonlinear coupling across frequency bands, as the excitability of neuronal populations in sensory systems is shaped by low-frequency oscillations through their phase<sup>29,30</sup>. To do so, we computed circular-to-linear correlations that quantify how the phase of low frequencies sampled in one region co-varies with the amplitude of higher frequencies in the other region. We confirmed that cross-regional effects were stronger than local ones (Fig. 4), that is, within-region phase-amplitude coupling (see Supplementary Fig. 2 and see Methods). Statistical significance of phase-amplitude coupling was assessed using a non-parametric cluster analysis<sup>31</sup> (see Methods). For each significant cross-regional phase-amplitude coupling cluster, we subsequently confirmed that there was a corresponding GC peak at the

frequency of the modulating phase (Table 1). In Fig. 4, vertical shaded bars indicate the overlap of phase-amplitude coupling and GC peaks. Overall this ensures that the observed modulations of  $\gamma$ -power were driven by the phase of distant lower frequencies.

Consistent with the notion that T-D information propagated in the low-frequency range has an influence on local  $\gamma$ -activity, we found a modulation of  $\gamma$ -power in the left A1 as a function of the phase of  $\theta$ -activity in the left AAC in both subjects (Fig. 4a and Supplementary Fig. 2). This pattern was similar in the right hemisphere, even though it survived statistical correction only in S3 (Fig. 4b). To each T-D (AAC-phase/A1-power) phase-amplitude coupling cluster corresponded a T-D GC peak at the phase frequency (Fig. 4, shaded red bars; see Table 1), allowing to conjecture that the T-D influence of AAC on A1 observed with GC in the low-frequency range was associated with  $\gamma$ -power modulation at target level. We also detected a significant modulation of  $\gamma$ -power in AAC as a function of the phase of  $\delta$  (1–3 Hz) activity measured in A1. As for T-D effects, B-U GC peaks (Table 1) aligned with each of these clusters (Fig. 4, shaded blue dotted bars). Finally, we observed a left dominance of this effect at  $\delta$ -rate, both with GC (Fig. 3c) and phase-amplitude coupling (Supplementary Fig. 3) measures, suggesting that B-U flow was stronger at very low frequencies (1–2 Hz) in the left hemisphere.

Altogether, GC and phase-amplitude coupling measures concurred to show a frequency division for B-U and T-D information transfer, whereby local  $\gamma$ -activity was globally modulated as a function of distant  $\delta$ -phase in the B-U direction (Fig. 4, blue clusters) and as a function of distant  $\delta$ - $\beta$  phase in the T-D direction (Fig. 4, red clusters). These findings suggest that the multiplexing of B-U and T-D information transfer operates, at least in part, by varying the modulation frequency of local  $\gamma$ -activity.

**Time division in B-U and T-D causality.** Directional multiplexing by spectral division enables continuous information transfer in B-U and T-D directions simultaneously. To assess whether information transfer was indeed continuous in both



**Figure 2 | Individual regional speech-induced corticograms and speech/brain cross-correlation.** Individual trial-averaged time-frequency representations (corticograms) of neural response to sentences (upper panels) and trial-averaged within-frequency speech/brain amplitude cross-correlograms (lower panels), computed in left (a) and right (b) A1, and left (c) and right (d) AAC. The dotted lines indicate sentence onset (upper panels) and zero-lagged speech/brain cross-correlation (lower panels).

**Table 1 | Low frequency peaks in Granger causality.**

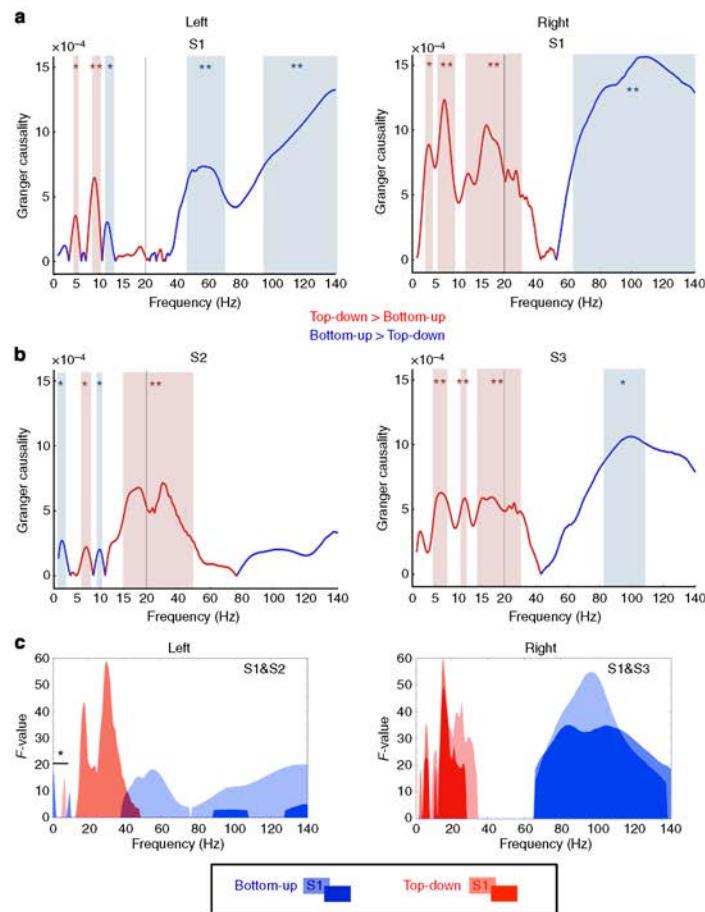
Frequency peaks	Top-down				Bottom-up			
	Peak 1		Peak 2		Peak 1		Peak 2	
	Hz	P-value	Hz	P-value	Hz	P-value	Hz	P-value
S1 Left	5	0.05	9	0.01	1	0.05	7	n.s.
S2 Left	7	n.s.	13	n.s.	2	0.05	10	n.s.
S1 Right	7	0.01	16	0.05	2	0.05	5	n.s.
S3 Right	7	0.01	12	n.s.	3	n.s.	9	n.s.

n.s., non-significant.

Low-frequency (1–20 Hz) peaks (2 maximum) for each data set and causal direction (top-down or bottom-up; excerpted from Supplementary Fig. 1. Peaks are sorted in ascending frequency, shown under the Hz columns, with the corresponding significance level (see Methods) shown under P-value columns (statistics are FDR corrected).

directions (and related frequency ranges), we computed GC at any instant (1 ms resolution) during the processing of auditory sentences to obtain T-D and B-U GC TF representations.

Contrary to what we expected, we did not observe lasting periods when GC dominates in one direction or the other. Rather, TF GC was organized as an alternation of frequency-specific bins,



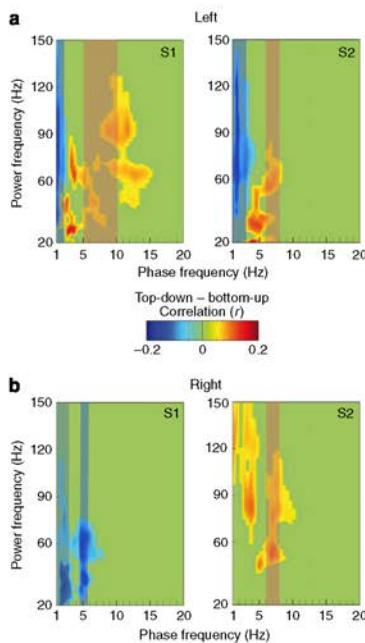
**Figure 3 | Spectral asymmetry between B-U and T-D using GC.** Spectral differences (1–140 Hz) between T-D (AAC to A1) and B-U (A1 to AAC) causal influences in left and right auditory cortex of subject 1 (a) and subjects 2 and 3 (b). Red and blue lines show, respectively, T-D and B-U predominance, averaged over time, trials and sentences. Statistically significant differences between T-D and B-U are highlighted (shaded bars; FDR correction; \* $q < 0.05$ , \*\* $q < 0.01$ ). (c) F-values obtained from one-way analysis of variance (ANOVA) analysis for each subject, testing the difference between T-D (red) and B-U (blue) causal directions. Data from S1 are shown in semi-transparent colours, and data from S2 and S3 in full colours. Only significant values are shown ( $P \leq 0.05$ , Bonferroni corrected). The star further indicates an interaction (two-way ANOVA, see Methods) in the 1–6 Hz range, where B-U GC dominates in the left hemisphere ( $P \leq 0.05$ , Bonferroni corrected).

suggesting that information transfer proceeds by alternating periods of dominant B-U and dominant T-D (Fig. 5a), cold colours for B-U and hot colours for T-D). We tested for a periodicity in the alternation of T-D and B-U GC by computing the Fourier spectrum of T-D minus B-U TF matrix (Fig. 5b,c; false discovery rate (FDR) correction, see Methods). Although B-U and T-D flow did not use the same frequency ranges, we found a common low rate temporal arrangement. In all subjects, a significant temporal modulation at  $\delta$ -rate (1–3 Hz) was observed (Fig. 5b,c; see permutation tests in Methods). The results hence suggest an alternation of dominant T-D and B-U information transfer approximately every 300–500 ms in both

flow directions. It is important to note that the temporal structure of GC changes may also reflect that GC is based on a linear model of temporal dependencies, which cannot account for the nonlinear dependencies we have established in terms of cross-regional phase-amplitude coupling. It is hence unclear whether this alternation would hold for inter-regional cross-frequency coupling effects.

## Discussion

In telecommunications, frequency-division multiplexing refers to the use of separate carrier frequencies to transmit distinct



**Figure 4 | Cross-regional dominance of phase-amplitude coupling.** Circular-to-linear correlations computed between low-frequency phase (1–20 Hz) of one region (A1 or AAC) and higher-frequency power (20–150 Hz) of the other region in left (a) and right (b) hemispheres. B-U (A1-phase modulating AAC power) and T-D (AAC-phase modulating A1 power) phase-amplitude coupling values were contrasted, with blue and red clusters indicating B-U and T-D dominances, respectively. In addition, we controlled that inter-regional cross-frequency dependencies dominated over local ones, by controlling for local phase, that is, performing the contrast (T-D–local-AAC)–(B-U–local-A1). Only significant ( $P \leq 0.01$ , cluster corrected) contrasts are reported. Shaded bars represent GC peaks (see Table 1) overlapping with phase-amplitude coupling results at the frequency of the modulating phase.

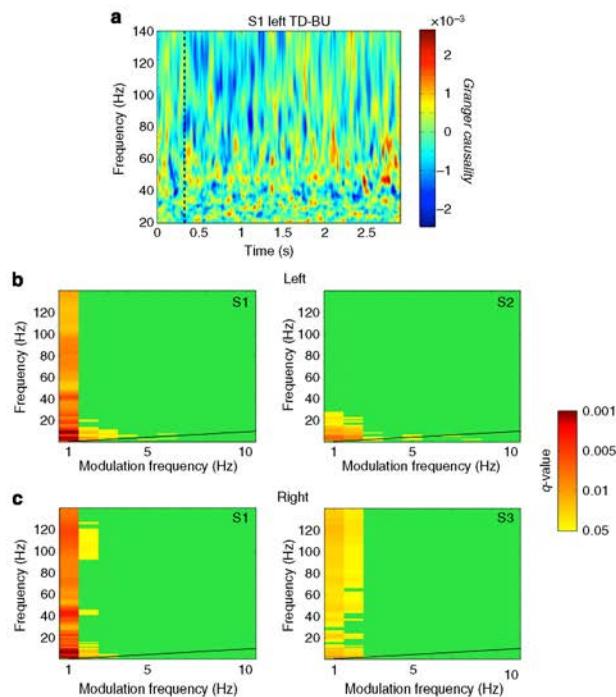
modulated signals on a single physical support<sup>32</sup>. In auditory processing, the notion of multiplexing is increasingly referred to in a different acceptation, that is, the description of parallel information processing at several timescales<sup>14,15</sup>. Here we explored the notion that the brain uses directional multiplexing, whereby B-U and T-D information are propagated using distinct modulation frequencies and/or different carrier frequencies. This idea arises from neurophysiological studies showing that the low- $\beta$ -range (around 14 Hz) was mostly related to endogenous T-D process<sup>33–36</sup>. From a theoretical viewpoint, the idea that the brain constantly compares incoming input with internal representations<sup>15</sup>, calls for information processing at different timescales<sup>6,7,13</sup> and information unmixing.

The current results indicate a frequency dissociation in information transfer along the auditory cortical hierarchy:  $\delta$ - $\beta$  frequencies dominated in the T-D direction and  $\gamma$ -frequencies in the B-U direction. These findings support theoretical proposals<sup>6,13</sup> and complement other results—so far only partly conclusive—obtained in humans and monkeys<sup>18,33–37</sup>. By

combining GC and nonlinear cross-regional phase-amplitude coupling, we further show that T-D processes operated by modulating fast neural activity at  $\delta$ - $\beta$  rates in the target area, that is, the  $\gamma$ -power in A1 being modulated as a function of the phase of low frequencies in AAC. Furthermore, we observed several T-D GC peaks in the  $\delta$ - $\beta$  (4–30 Hz) frequency range, and the frequency of these peaks varied across individuals (and hemispheres). Overall, we detected more GC peaks than directional phase-amplitude coupling clusters, suggesting either that not all the detected GC peaks translated into phase-power modulations between source and target locations, for instance due to the presence of phase–phase coupling<sup>38</sup>, or that the circular-to-linear correlation method was not sensitive enough to detect them all. Importantly, however, to each single phase-amplitude coupling cluster corresponded a GC peak. This cross-validates the results and ensures that  $\gamma$ -modulations by the phase sampled from another region truly reflects distant modulations. Combining GC and phase-amplitude coupling constitutes an exploratory alternative to dynamic causal model (DCM)<sup>39</sup>. Chen *et al.*<sup>40</sup> used DCM to assess amplitude–amplitude cross-frequency coupling between high- and low-visual areas during perception of human faces. They also found qualitative evidence for functional asymmetry coupling, where the effects of low frequencies on high frequencies were greater in the backward direction relative to the forward direction. However, although DCM allows for an exploration of linear and nonlinear interactions within a single model, our approach may be more flexible in discovering non-hypothesized neurophysiological phenomena, such as those we detected in the very low-frequency range (see ref. 41 for a comparison of DCM and GC).

Our findings confirmed that T-D neural flow uses lower frequency ranges than B-U, but also point to a more complex picture than the previously hypothesized  $\gamma$ -up  $\beta$ -down scheme<sup>6,18,33–37</sup>. In left auditory regions, we detected GC B-U peaks in the  $\delta$ -frequency range (1–3 Hz), indicating that very low frequencies were also involved in B-U transfer (Table 1). Cross-frequency coupling did not only confirm this effect but further showed that the phase of  $\delta$ -activity sampled in left A1 was associated with a modulation of high  $\gamma$ -power (80–100 Hz) in left AAC (Fig. 4). The fact that this effect dominated in the left hemisphere (Supplementary Fig. 3) could reflect (i) that  $\gamma$ -activity in AAC was more pronounced in left than right AAC (Fig. 2), (ii) that the low-frequency phase-locking of responses (Supplementary Fig. 4) and stimulus/brain correlations (Fig. 2) were stronger in left than right A1, (iii) or both. At any rate, the left dominance in  $\delta$ / $\gamma$ -coupling presumably reflects some aspects of the functional specialization of left auditory regions in speech processing<sup>42</sup>. Importantly, the spectral division of B-U and T-D information transfer appears partly flexible, depending on cognitive demands and/or on the interaction of stimulus rhythms with local oscillatory properties.

The current results do not only show a spectral division but also suggest a time division of labour between B-U and T-D processes. The analysis of temporal modulations of GC TF representations (Fig. 5) revealed the presence of significant slow fluctuations (1–3 Hz) of GC in the  $\gamma$ -frequency range, suggesting that B-U and T-D information sequentially dominated over periods of  $\sim 300$ –500 ms. What could determine the regular alternation of B-U and T-D-dominant periods at this slow rate remains unclear at this point. It could result from time constants that are specific to speech. It has been shown that predicting forthcoming speech involves a syllable-based mechanism<sup>43</sup>, which is roughly compatible with 300 ms predictive segments. Alternatively, slow modulations could be entirely driven by endogenous  $\delta$ - or  $\alpha$ -rhythms, whose phase (i) determines whether a stimulus is going to be detected at the sensory level<sup>44–46</sup> and (ii)



**Figure 5 | Temporal modulations of GC.** (a) Difference between T-D and B-U Granger causal influence plotted as a function of time and frequency (shown for S1). GC time-frequency representations do not show evidence of continuous causal influences across distinct frequency bands, for example, B-U  $\gamma$  and T-D  $\delta$ - $\beta$ . Conversely we observe alternations in both time and frequency, suggestive of a discontinuous pattern of information transmission. GC modulation spectrum in left (b) and right (c) auditory cortex, resulting from Fourier-transforms of GC time-frequency data. Only significant ( $q \leq 0.01$ , FDR corrected) modulations are reported. Black lines correspond to the diagonal (that is, modulation frequency = modulated frequency).

indexes the dynamics of sequential information processing at higher stages<sup>47</sup>. In this respect, it would be interesting to obtain similar data from the visual modality with less temporally structured stimuli, and assess to what extent the timing of GC is stimulus driven, or emerges from properties of the brain organization.

The main advantage of spectral multiplexing is to prevent interferences during multiple and continuous information transfer using the same physical support. In the cortex, BU and TD information travel via separate vectors<sup>13</sup> but information unmixing, by, for example, multiplexing, is required in relays where ascending and descending information converge and are integrated (that is, superficial layers). The modulation of neuronal spiking (here approximated by local high  $\gamma$ -activity<sup>28</sup>) at distinct,  $\delta$ - $\beta$  versus  $\gamma$ , rates could be an efficient means to achieve it. Alternatively, interferences could be avoided by using a single information channel, with time windows during which B-U or T-D dominates. Although the observed spectral division between B-U and T-D information flows (Fig. 3) supports the former scenario, temporal alternation of ascending and descending information (GC modulations; Fig. 5) supports the latter scenario. Although frequency and time division are not conceptually incompatible, the use of both mechanisms for directional information transfer appears computationally

redundant. The presence of effects in the  $\delta$  (1–3 Hz) range for both spectral (Figs 3 and 4) and time (Fig. 5) division is fairly compatible with the predictive coding framework. This reflects the fact that both B-U and T-D tend to fluctuate at  $\delta$ -rate (Fig. 5), while being predominantly oriented towards the B-U direction (Fig. 3 and Supplementary Fig. 3). Such an asymmetry in information flow is in line with the proposal that T-D message passing results from the accumulation of B-U evidence, as this process could translate in the co-occurrence of a continuous B-U accumulation of prediction errors and a discontinuous T-D prediction flow. More generally, the present results are consistent with predictive coding models in the sense that T-D predictions of auditory input rest on a nonlinear mapping from higher-level representations, as shown by the nonlinear cross-regional phase-amplitude coupling results (Fig. 4).

T-D effects were mainly associated with modulations of  $\gamma$ -activity at rates ranging from 5 to 30 Hz. The  $\beta$ -rhythm could have a local (cortical) origin that is compatible with a function in T-D control. An interesting model of  $\beta$ -generation from *in vitro* slice preparations suggests that low  $\beta$ -activity could result from the local concatenation of two independent higher-frequency rhythms ( $\gamma$ ) generated in superficial and deep layers, respectively<sup>48</sup>. *In vitro* experimental observations<sup>48,49</sup> indicate that there is an alternation between either the two independent higher-

frequency rhythms or the  $\beta$  one, depending on the level of excitation and on synaptic plasticity. We previously speculated that this generation mode could constitute a possible mechanical switch between a state where information flows freely upward ( $\gamma$  in deep and superficial layers) and a state where information transfer is redirected, first confined within a cortical level and then directed downward<sup>7</sup>. From an information-processing perspective, the latter state could serve to match ascending and descending information, and reduce the discrepancies between incoming inputs and internal representations. Remarkably, the biophysical mechanism of  $\beta$ -generation proposed by Roopun *et al.*<sup>49</sup>, involving the concatenation of slow and fast rhythms, implies the temporal alternation of B-U dominant phases and T-D dominant phases, consistent with what we observed here.

An alternative account for the frequency range where T-D effects operate could be related to the  $\alpha$ -rhythm physiology, as T-D phase-amplitude coupling effects were largely distributed around 10 Hz.  $\alpha$ -Rhythm is the most conspicuous and widespread of the brain rhythms owing presumably to its thalamo-cortical origin<sup>50</sup>, and is an important effector of attentional processes<sup>51,52</sup>. Although it could be involved in descending mechanisms, its general role in attention and sensory gating is hardly compatible with one in specific information and representations transfer. However,  $\alpha$ -rhythm displays bistability with occasional splitting in high- $\delta$  and low- $\beta$  components<sup>53–55</sup>. It is unclear how this splitting operates, but if it was associated with a change from a widespread generation mode to a more local one, it could also possibly underpin the effects we observe here. Alternatively, low  $\beta$ -oscillations could be generated by the interaction between a feedforward and a feedback  $\alpha$ -input, as suggested by another model<sup>56</sup>.

Thanks to unique depth intracortical human recordings collected in two hierarchical regions of speech processing bilaterally in one subject and unilaterally in other two subjects, we show a spectral division of the B-U and T-D processing flow. Although our findings confirm a  $\gamma$ -up scheme, they do not support a  $\beta$ -down one. They show that a larger band involving  $\delta$ - $\beta$  frequencies are involved in T-D information transfer, and further suggest that local  $\gamma$ -activity is modulated by the phase of lower-frequency distant oscillatory activity. We additionally showed that directional information transfer does not proceed continuously, but alternates at a 1- to 3-Hz rate. These data suggest that speech processing uses both distinct modulation frequencies and temporal windows to transfer information in B-U and T-D direction. The reason why B-U and T-D unmixing appear implemented in a redundant manner, and the extent to which the time constants we observed here are specific to speech processing or could instead generalize, remain open questions.

## Methods

**Subjects.** Three French female subjects participated in this study. They suffered from drug-resistant partial epilepsy and were implanted for presurgical investigation with chronic depth electrodes in: right and left auditory cortex (S1, 45 years old), left auditory cortex (S2, 30 years old) and right auditory cortex (S3, 35 years old). The electrodes of interest were located in Heschl's gyrus (primary auditory cortex, A1) and laterally in the superior temporal gyrus in a region we refer to as AAC, as well as in other cortical structures that were not relevant for the study (see functional validation of electrode position below). The subjects provided informed consent to the protocol, which was approved of by the institutional review board of the French Institute of Health. Neuropsychological assessment indicated that they had intact language functions. Brainstem-evoked potentials and pure tone audiograms carried out before iEEG indicated intact cochlear and brainstem auditory functions. Analysis of iEEG indicated that the epileptic zones were located outside the regions examined here.

**Electrophysiological recordings.** *Stimuli and data acquisition.* The subjects listened to 110 repetitions of two 2.5-s-long sentences in French, uttered by a French female whose voice had a fundamental frequency of 201 Hz. Stimuli were presented monaurally to both ears, in a pseudo-randomized order with an interstimulus interval of 4,135 s, and only the contra-lateral response was taken into account.

The stimuli were recorded digitally at a sampling frequency of 44.1 kHz and delivered to the subjects at 22 kHz with a 75-dB sound pressure level headset using E-prime software (Psychology Software Tools Inc., Pittsburgh, PA, USA).

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 EEG 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  $\mu$ V).

**Anatomo-functional definition of contacts position.** The stereotactic method was based on the co-registration of the subjects' magnetic resonance imaging (MRI) with the stereotactic angiogram, to prevent injury to brain vessels. Multi-lead 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. 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 (see Fig. 1b for electrode position in S1). Auditory-evoked potentials measured in response to pure tones were used to functionally delineate A1 and AAC, and select the right electrodes. Auditory-evoked potentials were averaged over trials, after epoching (200–635 ms) and taking the 150–50 ms pre-stimulus time period as baseline. All contacts that elicited no significant responses were discarded. In a second step, A1 was functionally defined based on the presence of early P20/N30 components. These responses were located in the medial and intermediate part of Heschl's gyrus. Third, for each functional area (A1 and AAC in left and/or right hemispheres), the most responsive contact was selected for subsequent analyses.

**Preprocessing.** Data analysis was performed with EEGLab v.8 (sccn.ucsd.edu/eeqlab) for data extraction, Fieldtrip (<http://www.ru.nl/donders/fieldtrip>) and Fast\_tf (<http://cogimage.dsi.cnr.it/logicieli/>), for TF decomposition. For GC, we modified and adapted the scripts used in ref. 57. Data were epoched into segments, including a baseline period (328 ms for S1, 1000 ms for S2 and S3) before stimulus onset and an after stimulus period (452 ms for all subjects). Epochs including signals that deviated from the average response of all the trials were discarded, by computing the correlation between each single trial and the average response, and then rejecting the 15% of trials with the lowest Pearson's correlation value. We set this conservative rejection criterion and validated the approach on the basis of visual inspection of the signals. The data set analysed was hence free of suspicious electrical activity related to epilepsy, for example, interictal spikes.

We used a bipolar montage for data analyses, meaning that electrical activity from all contacts was subtracted from a common reference signal that corresponded to the average response of the least-responsive (mesial or lateral) contact of each electrode. This resulted in attenuating global noise (50 Hz ambient electric field) in a similar way for all contacts.

**TF analyses and power spectrum.** A TF continuous wavelet transform was applied to each epoch using a family of complex Morlet wavelets ( $m = 7$ ), resulting in an estimate of power and phase at each time point and each frequency, with a 0.5-Hz resolution below 20 and 1 Hz above. The TF resolution of the wavelets was frequency dependent (at 7 Hz  $\sigma = 150$  ms, 1 Hz; at 35 Hz  $\sigma = 30$  ms, 5 Hz). We restricted the analysis to frequencies between 1 and 150 Hz, spanning the whole range of relevant brain rhythms (up to high  $\gamma$ -activity). Figure 2 shows the typical cortical responses (increase or decrease in signal power relative to baseline in decimal logarithmic units (dB) at each time and frequency data point) for all subjects and brain areas.

**Speech/brain cross-correlations.** *Sentences characterization.* For each sentence, we estimated the wideband envelope of the speech waveform<sup>24,25</sup>. The raw speech waveform was band-pass filtered into 32 frequency bands, encompassing 80–8,500 Hz with a logarithmical spacing, modelling the cochlear frequency decomposition. The absolute value of the Hilbert transform of each band-passed signal constitutes an estimate of the envelope for that frequency band, and their sum an estimate of the wideband envelope. Finally, we computed the power in each frequency band at each time point, with a millisecond resolution, similar to the iEEG data, that is, between 1 and 150 Hz, with a 0.5 Hz resolution below 20 Hz and 1 Hz above, by applying a TF wavelet transform, using a family of complex Morlet wavelets ( $m = 7$ ).

*Speech/brain cross-correlation computation.* We cross-correlated over time for each trial, sentence and region the oscillatory power estimates of the neural data at each frequency (1–150 Hz) with the corresponding frequency of the acoustic signal, between –50 and 200 ms, relatively to the acoustic input (with brain response following acoustic signal for positive cross-correlation values). This procedure results in an estimate of stimulus/response correlation at each frequency and for multiple time delays. At each time delay, correlation equals 1 if the two signals are perfectly identical when taking into account this time delay, and 0 if the two signals are totally unrelated. Data were subsequently averaged over trials and sentences.

**GC analysis.** GC is classically used to assess causal influence among two time series<sup>22</sup>. The basic assumption is that a time series  $X(t)$  linearly causes another time series  $Y(t)$  if the future trend of the latter is better predicted by looking at the past of  $X$  and  $Y$  than by looking at the past of  $Y$  alone. For stationary processes, the computation of GC relied on multivariate autoregressive models to estimate the prediction error in the two conditions<sup>22</sup>. For non-stationary time series, such as oscillating neural signals, GC spectra can be obtained in a non-parametric manner by computing Geweke's frequency domain version of GC without going through the multivariate autoregressive model fitting<sup>35,59</sup>. We therefore used a spectral density matrix factorization technique on complex cross-spectra, obtained from the continuous wavelet transform of recorded iEEG time series<sup>58</sup>. Both parametric and non-parametric GC have been previously used in neuroscience to assess linear directional influence between two communicating brain areas in local field potential<sup>60</sup>, EEG<sup>57</sup> and functional MRI data<sup>61</sup>.

GC was computed on a trial-by-trial basis for each subject using the method proposed by Dhamala *et al.*<sup>58</sup>, and then averaged across time, trials and sentences (Fig. 3 and Supplementary Fig. 1). Statistics were then computed by generating 1,000 permutations of iEEG data, in which left and right electrodes, as well as trials, were randomized. This procedure permits to rule out that the observed effects arose from noise or specific methodology, as the exact same data and algorithms were used to compute the permuted trials. For each realization, we computed the mean GC across trials and the corresponding s.d. The original GC spectra were then standardized to obtain a vector of Z-values, one for each frequency.

T-D and B-U influences can be measured simultaneously<sup>18,62</sup>. The information flow was considered T-D when GC from AAC to A1 exceeded GC from A1 to AAC, and B-U in the other case.

We tested for significant frequency peaks separately for each T-D and B-U GC direction (Supplementary Fig. 1 and Table 1) together with frequency ranges where T-D and B-U GC spectra were significantly different (Fig. 3a,b). For the first analysis, we directly compared the Z-transformed vectors obtained from GC spectra to a zero-mean normal distribution, and corrected for multiple comparisons with the FDR method at a one-tailed  $q$ -value of  $q \leq 0.05$ . For the second analysis, we first computed the difference in Z-values between T-D and B-U Granger spectra at each frequency point, and then compared it with the zero-mean normal distribution thresholding at a two-tailed  $q$ -value of  $q \leq 0.05$  (or  $q \leq 0.01$ , see Fig. 3), FDR corrected.

To further explore effects of flow direction, we applied a one-way analysis of variance test on time-averaged GC spectra for all data sets at each frequency point (Fig. 3c). Red (blue) areas correspond to frequencies where the T-D (B-U) mean is significantly higher than the B-U (T-D) mean. Values were thresholded at  $P \leq 0.05$  (Bonferroni corrected for multiple comparisons). We also performed a two-way analysis of variance to investigate together the effects of hemisphere (left/right) and flow direction (T-D/B-U). The four data sets were tested pairwise (contrasting SI left and right hemispheres, and S2 with S3) and the values were thresholded at  $P \leq 0.05$  (Bonferroni corrected for multiple comparisons). A significant interaction was detected at low frequencies (1–6 Hz).

To assess the temporal alternation of directional GC peaks, we first subtracted T-D and B-U TF matrices at each trial (see example in Fig. 5a). We then Fourier transformed this data at each frequency band (1 Hz resolution) and averaged across trials and sentences to obtain the modulation spectrum for each data set. For statistical testing, we computed 1,000 permutations of GC data by shuffling trials and electrodes, z-scored the data using the mean and variance obtained from permutations and used FDR correction for multiple comparisons (Fig. 5b,c).

**Cross-regional phase-amplitude coupling.** Cross-frequency coupling dependencies were studied under the phase-amplitude coupling framework. The rationale for using phase-amplitude coupling is that cross-frequency interaction could provide dynamic gating of information. Cross-regional phase-amplitude coupling would in turn reveal direct nonlinear interactions between distant sites<sup>63</sup>. We used phase and squared power values (amplitude) to approximate circular and Gaussian distribution, respectively. We subsequently computed the circular-to-linear correlation<sup>64</sup>, between each 1–20 Hz phase and 20–150 Hz amplitude frequencies. Correlations were computed across trials, sentences and time dimensions altogether, resulting in an estimate of the amount  $r$  of correlation between two frequencies, under a phase-amplitude dependency. To compute inter-regional dominance, we contrasted T-D and B-U phase-amplitude coupling analyses, while controlling for their dominance over local effects. We controlled for local phase (T-D—local AAC)—(B-U—local A1) to ensure that the amplitude modulations detected in one region are significantly more strongly related to the low-frequency phase sampled in the distant region than to local low-frequency phase.

Significant phase-amplitude coupling was based on corrected  $P$ -values using non-parametric permutation tests to generate null distributions of the maximum cluster size<sup>21</sup>. This implicitly adjusts for searching over multiple frequencies. The null distribution was obtained by computing 1,000 times the circular-to-linear correlation from a random mix of data taken equitably from the four types of phase/power relations we investigated (T-D, B-U, local A1 and local AAC). Clusters were defined as contiguous  $r$ -values above 0.045. Clusters of the 99th percentile (corresponding to  $P \leq 0.01$ ) were considered significant and are reported throughout the manuscript.

To further explore effects of flow direction and hemisphere (Supplementary Fig. 3), we contrasted the left and right phase-amplitude coupling patterns

observed in Fig. 4 (T-D minus B-U corrected for local phase). To highlight only left-dominant results, we applied to the left-right contrast a mask corresponding to the left hemispheric phase-amplitude coupling patterns. Statistics were computed similarly than before, except that we took a random mix of data taken equitably from the eight types of phase/power relations we investigated (left/right, T-D/B-U, local A1/AAC). In this approach, clusters were defined as contiguous  $r$ -values above 0.07 and clusters of the 99th percentile (corresponding to  $P \leq 0.01$ ) were considered significant.

**Phase-locking value.** For each region of interest, we evaluated the evoked modulation spectrum. We computed the phase-locking value<sup>65</sup> across trials for each time point and 1–150 Hz frequency. To obtain the modulation spectrum, we averaged the resulting phase-locking values over time and sentences.

## References

1. Friston, K. A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **360**, 815–836 (2005).
2. Knill, D. C. & Pouget, A. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* **27**, 712–719 (2004).
3. Winkler, I. & Czigler, I. Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. *Int. J. Psychophysiol.* **83**, 132–143 (2012).
4. Mumford, D. On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol. Cybern.* **66**, 241–251 (1992).
5. Rao, R. P. & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2**, 79–87 (1999).
6. Bastos, A. M. *et al.* Canonical microcircuits for predictive coding. *Neuron* **76**, 695–711 (2012).
7. Arnal, L. H. & Giraud, A. L. Cortical oscillations and sensory predictions. *Trends Cogn. Sci.* **16**, 390–398 (2012).
8. van Wassenhove, V., Grant, K. W. & Poeppel, D. Visual speech speeds up the neural processing of auditory speech. *Proc. Natl Acad. Sci. USA* **102**, 1181–1186 (2005).
9. Rauss, K., Schwartz, S. & Pourtois, G. Top-down effects on early visual processing in humans: a predictive coding framework. *Neurosci. Biobehav. Rev.* **35**, 1237–1253 (2011).
10. Costa-Faidella, J., Baldeuweg, T., Grimm, S. & Escera, C. Interactions between “what” and “when” in the auditory system: temporal predictability enhances repetition suppression. *J. Neurosci.* **31**, 18590–18597 (2011).
11. Wacongne, C. *et al.* Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc. Natl Acad. Sci. USA* **108**, 20754–20759 (2011).
12. Recasens, M., Grimm, S., Wollbrink, A., Panter, C. & Escera, C. Encoding of nested levels of acoustic regularity in hierarchically organized areas of the human auditory cortex. *Hum. Brain Mapp.* doi: 10.1002/hbm.22582 (2014).
13. Wang, X. J. Neurophysiological and computational principles of cortical rhythms in cognition. *Physiol. Rev.* **90**, 1195–1268 (2010).
14. Walker, K. M., Bizley, J. K., King, A. J. & Schnupp, J. W. Multiplexed and robust representations of sound features in auditory cortex. *J. Neurosci.* **31**, 14565–14576 (2011).
15. Panzeri, S., Brunel, N., Logothetis, N. K. & Kayser, C. Sensory neural codes using multiplexed temporal scales. *Trends Neurosci.* **33**, 111–120 (2010).
16. Cannon, J. *et al.* Neurosystems: brain rhythms and cognitive processing. *Eur. J. Neurosci.* **39**, 705–719 (2014).
17. Sauseng, P., Klimesch, W., Gruber, W. R. & Birbaumer, N. Cross-frequency phase synchronization: a brain mechanism of memory matching and attention. *Neuroimage* **40**, 308–317 (2008).
18. Bosman, C. A. *et al.* Attentional stimulus selection through selective synchronization between monkey visual areas. *Neuron* **75**, 875–888 (2012).
19. Ainsworth, M. *et al.* Rates and rhythms: a synergistic view of frequency and temporal coding in neuronal networks. *Neuron* **75**, 572–583 (2012).
20. Gourevitch, B., Le Bouquin Jeannes, R., Faucon, G. & Liegeois-Chauvel, C. Temporal envelope processing in the human auditory cortex: response and interconnections of auditory cortical areas. *Hear. Res.* **237**, 1–18 (2008).
21. Granger, C. W. J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**, 424–438 (1969).
22. Bressler, S. L. & Seth, A. K. Wiener-Granger causality: a well established methodology. *Neuroimage* **58**, 323–329 (2011).
23. Morillon, B., Liegeois-Chauvel, C., Arnal, L. H., Benar, C. G. & Giraud, A. L. Asymmetric function of theta and gamma activity in syllable processing: an intra-cortical study. *Front. Psychol.* **3**, 248 (2012).
24. Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A. & Ghazanfar, A. A. The natural statistics of audiovisual speech. *PLoS Comput. Biol.* **5**, e1000136 (2009).
25. Smith, Z. M., Delgutte, B. & Oxenham, A. J. Chimaeric sounds reveal dichotomies in auditory perception. *Nature* **416**, 87–90 (2002).

26. Buzsaki, G. & Draguhn, A. Neuronal oscillations in cortical networks. *Science* **304**, 1926–1929 (2004).
27. Buzsaki, G., Anastassiou, C. A. & Koch, C. The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci.* **13**, 407–420 (2012).
28. Ray, S. & Maunsell, J. H. Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol.* **9**, e1000610 (2011).
29. Schroeder, C. E. & Lakatos, P. Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* **32**, 9–18 (2009).
30. Lakatos, P. *et al.* An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J. Neurophysiol.* **94**, 1904–1911 (2005).
31. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).
32. Stern, H. P. E., Mahmoud, S. A. & Stern, L. E. *Communication Systems: Analysis and Design* (Pearson Prentice Hall, 2004).
33. Buffalo, E. A., Fries, P., Landman, R., Buschman, T. J. & Desimone, R. Laminar differences in gamma and alpha coherence in the ventral stream. *Proc. Natl Acad. Sci. USA* **108**, 11262–11267 (2011).
34. Buschman, T. J. & Miller, E. K. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* **315**, 1860–1862 (2007).
35. Arnal, L. H., Wyart, V. & Giraud, A. L. Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nat. Neurosci.* **14**, 797–801 (2011).
36. Bartolo, R., Prado, L. & Merchant, H. Information processing in the primate basal ganglia during sensory-guided and internally driven rhythmic tapping. *J. Neurosci.* **34**, 3910–3923 (2014).
37. Gregoriou, G. G., Gotts, S. J., Zhou, H. & Desimone, R. High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science* **324**, 1207–1210 (2009).
38. Belluscio, M., Mizuseki, K., Schmidt, R., Kemper, R. & Buzsaki, G. Cross-frequency phase-phase coupling between theta and gamma oscillations in the hippocampus. *J. Neurosci.* **32**, 423–435 (2012).
39. Kiebel, S. J., Garrido, M. I., Moran, R., Chen, C. C. & Friston, K. J. Dynamic causal modeling for EEG and MEG. *Hum. Brain Mapp.* **30**, 1866–1876 (2009).
40. Chen, C. C., Henson, R. N., Stephan, K. E., Kilner, J. M. & Friston, K. J. Forward and backward connections in the brain: a DCM study of functional asymmetries. *Neuroimage* **45**, 453–462 (2009).
41. Friston, K., Moran, R. & Seth, A. K. Analysing connectivity with Granger causality and dynamic causal modelling. *Curr. Opin. Neurobiol.* **23**, 172–178 (2013).
42. Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402 (2007).
43. Gagnepain, P., Henson, R. N. & Davis, M. H. Temporal predictive codes for spoken words in auditory cortex. *Curr. Biol.* **22**, 615–621 (2012).
44. Ng, B. S., Schroeder, T. & Kayser, C. A precluding but not ensuring role of entrained low-frequency oscillations for auditory perception. *J. Neurosci.* **32**, 12268–12276 (2012).
45. Henry, M. J. & Obleser, J. Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc. Natl Acad. Sci. USA* **109**, 20095–20100 (2012).
46. Arnal, L. H., Doelling, K. B. & Poeppel, D. Delta-beta coupled oscillations underlie temporal prediction accuracy. *Cereb. Cortex* doi: 10.1093/cercor/bhu103 (2014).
47. Wyart, V., de Gardelle, V., Scholl, J. & Summerfield, C. Rhythmic fluctuations in evidence accumulation during decision making in the human brain. *Neuron* **76**, 847–858 (2012).
48. Roopun, A. K. *et al.* Period concatenation underlies interactions between gamma and beta rhythms in neocortex. *Front. Cell. Neurosci.* **2**, 1 (2008).
49. Roopun, A. K. *et al.* Temporal interactions between cortical rhythms. *Front. Neurosci.* **2**, 145–154 (2008).
50. Bollimunta, A., Mo, J., Schroeder, C. E. & Ding, M. Neuronal mechanisms and attentional modulation of corticothalamic alpha oscillations. *J. Neurosci.* **31**, 4935–4943 (2011).
51. Jensen, O. & Bonnefond, M. Prefrontal alpha- and beta-band oscillations are involved in rule selection. *Trends Cogn. Sci.* **17**, 10–12 (2013).
52. Sadaghiani, S. *et al.* Intrinsic connectivity networks, alpha oscillations, and tonic alertness: a simultaneous electroencephalography/functional magnetic resonance imaging study. *J. Neurosci.* **30**, 10243–10250 (2010).
53. Freyer, F. *et al.* Biophysical mechanisms of multistability in resting-state cortical rhythms. *J. Neurosci.* **31**, 6353–6361 (2011).
54. Freyer, F., Aquino, K., Robinson, P. A., Ritter, P. & Breakspear, M. Bistability and non-Gaussian fluctuations in spontaneous cortical activity. *J. Neurosci.* **29**, 8512–8524 (2009).
55. Vierling-Claassen, D., Cardin, J. A., Moore, C. I. & Jones, S. R. Computational modeling of distinct neocortical oscillations driven by cell-type selective optogenetic drive: separable resonant circuits controlled by low-threshold spiking and fast-spiking interneurons. *Front. Hum. Neurosci.* **4**, 198 (2010).
56. Jones, S. R. *et al.* Quantitative analysis and biophysically realistic neural modeling of the MEG mu rhythm: rhythmodogenesis and modulation of sensory-evoked responses. *J. Neurophysiol.* **102**, 3554–3572 (2009).
57. Brovelli, A. Statistical analysis of single-trial Granger causality spectra. *Comput. Math. Methods Med.* **2012**, 697610 (2012).
58. Dhamala, M., Rangarajan, G. & Ding, M. Analyzing information flow in brain networks with nonparametric Granger causality. *Neuroimage* **41**, 354–362 (2008).
59. Barnett, L. & Seth, A. K. The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference. *J. Neurosci. Methods* **223**, 50–68 (2014).
60. Brovelli, A. *et al.* Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality. *Proc. Natl Acad. Sci. USA* **101**, 9849–9854 (2004).
61. Roebroeck, A., Formisano, E. & Goebel, R. Mapping directed influence over the brain using Granger causality and fMRI. *Neuroimage* **25**, 230–242 (2005).
62. Chen, Y., Bressler, S. L. & Ding, M. Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data. *J. Neurosci. Methods* **150**, 228–237 (2006).
63. van der Meij, R., Kahana, M. & Maris, E. Phase-amplitude coupling in human electrocorticography is spatially distributed and phase diverse. *J. Neurosci.* **32**, 111–123 (2012).
64. Berens, P. CircStat: a MATLAB toolbox for circular statistics. *J. Stat. Softw.* **31**, 1–21 (2009).
65. Lachaux, J. P., Rodriguez, E., Martinerie, J. & Varela, F. J. Measuring phase synchrony in brain signals. *Hum. Brain Mapp.* **8**, 194–208 (1999).
66. Liegeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P. & Chauvel, P. Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. *Electroencephalogr. Clin. Neurophysiol.* **92**, 204–214 (1994).

### Acknowledgements

This work was supported by the ERC (grant 260347), Fyssen, Philippe and Bettencourt-Schueller Foundations, and the CNRS. We thank Andrea Brovelli for methods support and Itsaso Olasagasti for her comments on our manuscript.

### Author contributions

C.L. carried out the experimental work; L.F., B.M. and A.-L.G. analysed the data and wrote the paper.

### Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications/>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://www.nature.com/reprintsandpermissions/>

**How to cite this article:** Fontolan, L. *et al.* The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. *Nat. Commun.* **5**:4694 doi: 10.1038/ncomms5694 (2014).

 This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>