Article

Cortical oscillations and speech processing: emerging computational principles and operations

GIRAUD MAMESSIER, Anne-Lise, POEPPEL, David

Abstract

Neuronal oscillations are ubiquitous in the brain and may contribute to cognition in several ways: for example, by segregating information and organizing spike timing. Recent data show that delta, theta and gamma oscillations are specifically engaged by the multi-timescale, quasi-rhythmic properties of speech and can track its dynamics. We argue that they are foundational in speech and language processing, 'packaging' incoming information into units of the appropriate temporal granularity. Such stimulus-brain alignment arguably results from auditory and motor tuning throughout the evolution of speech and language and constitutes a natural model system allowing auditory research to make a unique contribution to the issue of how neural oscillatory activity affects human cognition.

Reference

GIRAUD MAMESSIER, Anne-Lise, POEPPEL, David. Cortical oscillations and speech processing: emerging computational principles and operations. *Nature neuroscience*, 2012, vol. 15, no. 4, p. 511-7

DOI : 10.1038/nn.3063
PMID : 22426255

nature neuroscience

Cortical oscillations and speech processing: emerging computational principles and operations

Anne-Lise Giraud[1] & David Poeppel[2]

Neuronal oscillations are ubiquitous in the brain and may contribute to cognition in several ways: for example, by segregating information and organizing spike timing. Recent data show that delta, theta and gamma oscillations are specifically engaged by the multi-timescale, quasi-rhythmic properties of speech and can track its dynamics. We argue that they are foundational in speech and language processing, 'packaging' incoming information into units of the appropriate temporal granularity. Such stimulus-brain alignment arguably results from auditory and motor tuning throughout the evolution of speech and language and constitutes a natural model system allowing auditory research to make a unique contribution to the issue of how neural oscillatory activity affects human cognition.

During the evolution of human speech, the articulatory motor system has presumably structured its output to match those rhythms the auditory system can best apprehend[1]. Similarly, the auditory system has likely become tuned to the complex acoustic signal produced by combined jaw and articulator rhythmic movements[2]. Both auditory and motor systems must, furthermore, build on the existing biophysical constraints provided by the neuronal infrastructure. The present article proposes a perspective whereby neuronal oscillations in auditory cortex constitute a critical component of auditory-articulatory alignment and provide a first step deciphering continuous speech information.

Acoustic, neurophysiological and psycholinguistic analyses of connected speech demonstrate that there exist organizational principles and perceptual units of analysis at very different time scales[3]. Short-duration cues and information with a high modulation frequency, typically in ~30–50 Hz range and associated with an important part of the signal fine structure, correlate with attributes at the phonemic scale, such as formant transitions (for example, /ba/ versus /da/), the coding of voicing (for example, /ba/ versus /pa/), and other features. Almost an order of magnitude slower, the acoustic envelope of naturalistic speech closely correlates with syllabic rate and has a canonical time signature as well, the modulation spectrum typically peaking between 4 and 7 Hz. The accretion of signal input into lexical and phrasal units, perceptual groupings that carry, for example, the intonation contour of an utterance, occurs at yet a lower modulation rate, roughly 1–2 Hz. Although the temporal modulations on these three scales are aperiodic, they are sufficiently rhythmic to elicit robust regularities in the time domain, even in single utterances.

The rich frequency composition of speech has motivated much research on the neural foundations of speech perception. Although spectral information must be analyzed for successful processing, temporal modulations at low and high rates within each frequency band are critical. Spectral impoverishment of speech can be tolerated to a remarkable degree[4,5], whereas temporal manipulations cause marked failures of perception[6]. The framework we propose here hence focuses on bottom-up temporal analysis of speech.

We advance the hypothesis that a critical ingredient for parsing and decoding connected speech lies in the infrastructure provided by neuronal oscillations, neuronal population behavior especially well suited to deal with time-domain phenomena. Adopting and adapting ideas originating in previous work[3,7,8], we argue for a principled relation between the time scales present in speech and the time constants underlying neuronal cortical oscillations that is both a reflection of and the means by which the brain converts speech rhythms into linguistic segments. In this hypothesis, the low gamma (25–35 Hz), theta (4–8 Hz) and delta (1–3 Hz) bands provide a link between neurophysiology, neural computation, acoustics and psycholinguistics. The close correspondences between (sub)phonemic, syllabic and phrasal processing, on the one side, and gamma, theta and delta oscillations, on the other, suggest potential mechanisms for how the brain deals with the 'temporal administrivia' that underpin speech perception. Restricting our scope to the theta and gamma bands, the neurophysiological model we propose parallels a phenomenological model[8] that stipulates phase-locking and nested theta-gamma oscillations (to explain counterintuitive behavioral findings), suggesting that the brain can decode extremely impoverished speech provided that the syllabic rhythm is maintained[9]. We discuss new experimental evidence illustrating the operations and computations implicated in the context of this oscillatory framework. We also propose that oscillation-based decoding generalizes to other auditory stimuli and sensory modalities.

The central conjecture: oscillations determine speech analysis

We propose a cascade of processes that transform continuous speech into a discrete code, invariant to speech rate, reflecting certain essential temporal features of sublexical units (**Fig. 1**). This model achieves segmentation of connected speech at two timescales, which should permit the readout of discrete phonemic and syllabic units. We hypothesize that intrinsic oscillations in auditory cortex (A1 and A2, or Brodmann areas 41 and 42) interact with the neuronal (spiking) activity generated by an incoming speech signal. Subsequent to the

[1]Inserm U960, Département d'Etudes Cognitives, Ecole Normale Supérieure, Paris, France. [2]Department of Psychology, New York University, New York, New York, USA. Correspondence should be addressed to A.-L.G. (annelise.giraud@gmail.com).

Published online 18 March 2012; doi:10.1038/nn.3063

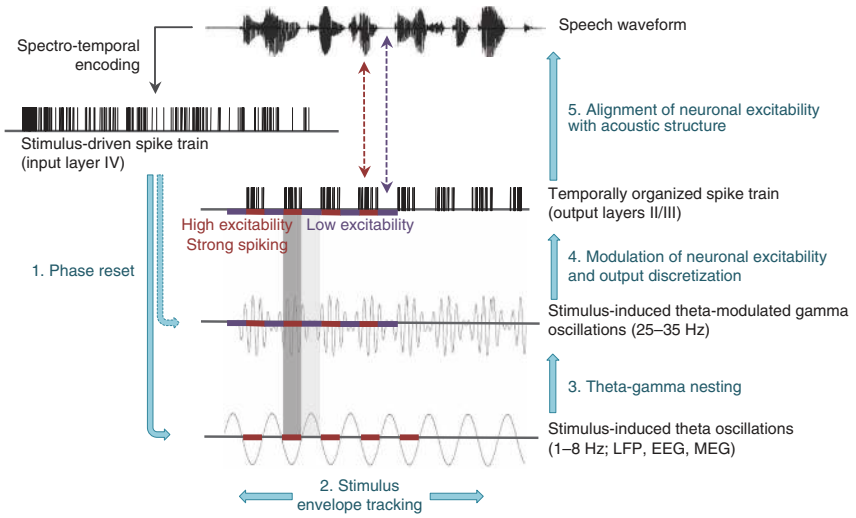

Figure 1 A theory of early oscillation-based operations in speech perception. Five operations allow connected speech to be parsed by cortical theta and gamma oscillations. We assume a high-resolution spectro-temporal representation of speech in primary auditory cortex. We represent a typical spike train in layer IV cortical neurons. Most of these neurons phase-lock to speech amplitude modulations. Response onset elicits a reset of theta oscillations in superficial layers (step 1) where auditory cortex output is generated. After reset, theta oscillations track the speech envelope (step 2). Theta reset induces a transient pause in gamma activity and a subsequent reset of gamma oscillations. Theta and gamma generators that are weakly coupled at rest become more strongly coupled and nested (step 3). Gamma power controls the excitability of neurons generating the feedforward signal from A1 to higher order areas (step 4). Neuronal excitability phase aligns to speech modulations (step 5): gamma tends to be strong when the energy in the signal is weak.

encoding of the spectro-temporal properties of a speech stimulus, the salient points ('edges') in the input signal cause phase resetting of the intrinsic oscillations in auditory cortex, in the theta and likely the gamma band (step 1). The activity in the theta band, in particular, is modulated to entrain to and track the envelope of the stimulus (step 2). The theta and gamma bands, which concurrently process stimulus information, lie in a nesting relation such that the phase of theta shapes the properties (amplitude, and possibly phase) of gamma (step 3). The activity in the gamma band has a tightly coupled relation to spike trains, regulating spike patterns (step 4). Finally, neuronal excitability is modulated such that acoustic structure of the input is aligned with neuronal excitability (step 5; **Fig. 1**). By this hypothesis, the theta and gamma oscillations act (i) by discretizing (sampling) the input spike trains to generate elementary units of the appropriate temporal granularity for subsequent processing and (ii) by creating packages of spike trains and excitability cycles. In summary, speech onsets trigger cycles of neuronal encoding at embedded syllabic and phonemic scales.

Phase resetting, speech envelope tracking and nesting

In human auditory cortex, sustained oscillatory activity can be detected at rest in discrete frequency bands, mostly in the delta–theta, alpha and low gamma domains[7] (**Fig. 2a**). When auditory cortex is stimulated by speech, resting oscillatory activity gives way to a temporally structured activity (**Fig. 2b**). The neuronal response profile is remarkably similar to the spectro-temporal structure of the speech envelope in the same 1–140 Hz frequency range (**Fig. 2c**,**d**). Cortical activity, however, does not track speech modulations equally over the whole 1–140 Hz frequency range, but preferentially in the theta and low and high gamma domains (**Fig. 2e**). In the example data set, we observe maxima of speech–brain coherence around 4 and 30–70 Hz. Thus, speech temporally organizes (resets) oscillatory activity that is already visible at rest, but only in specific frequency domains corresponding to the sampling rates optimal for phonemic and syllabic sampling.

Auditory cortex responds strongly to stimuli with complex temporal modulations in the amplitude (AM) and frequency (FM) domains[10,11], such as speech. Recent experiments using both noninvasive methodologies such as magnetoencephalography (MEG) and electroencephalography (EEG) and intracranial methodologies such as electrocorticography and stereotactic EEG have supported both phase resetting of cortical oscillations[12] and phase tracking of speech envelopes by the latter[12–15]. An important generalization has emerged: when envelope tracking fails, speech intelligibility is compromised. For example, in studies using rate manipulation by means of speech compression[13,15] or envelope manipulation by means of filtering[12], when theta band activity ceases to follow the speech envelope, intelligibility sharply degrades. Theta phase resetting and entrainment and phase tracking hence appear as two critical operations in parsing continuous speech. Whether gamma oscillations are directly phase reset by the stimulus or only through a theta phase reset is not yet well understood. Although established in the hippocampus, the mechanisms of theta and gamma generation, and their functional interaction, have still not been demonstrated in the auditory cortex. Investigations in animals, including slice work, could clarify whether theta and gamma rhythms are independently generated and reset, and how they interact during continuous auditory stimulation.

By preferentially tracking modulations within the delta–theta and gamma bands, auditory cortex 'discards' modulations situated in the beta (15–20 Hz) range. Thus speech is analyzed at two, and perhaps more, discontinuous time scales, with integration windows of ~150 ms and above versus ~30 ms and below. We assume this discontinuity to be a possible means by which speech is analyzed in parallel at syllabic and phonemic rates. Yet in left auditory cortex, analyses at slow and fast rates are not independent. The fractionation of modulation tracking over two discontinuous scales permits oscillatory nesting, the process by which the phase of slow cortical oscillations controls higher rate oscillations; namely, their power or phase[16]. Through theta-gamma nesting, concurrent syllabic and phonemic analyses can remain hierarchically bound. Nesting is manifest and can be functionally relevant only if there is a minimum ratio across frequencies. In the theta-gamma nesting pattern that emerges in the human primary auditory cortex in response to speech (**Fig. 2f**), there is a frequency ratio of about 4, suggesting that about 4 cycles of the higher frequency occur during one cycle of the lower one. Whether nesting is preserved when speech is accelerated, up to which ratio, and how its potential failure affects speech comprehension are important missing elements of the puzzle. Partial evidence has been obtained from human intracortical recordings[15], but more work is needed.

Spike patterning and discretization

Schroeder and colleagues have proposed that spiking is hierarchically controlled by cortical oscillations[17]. Oscillations are typically recorded

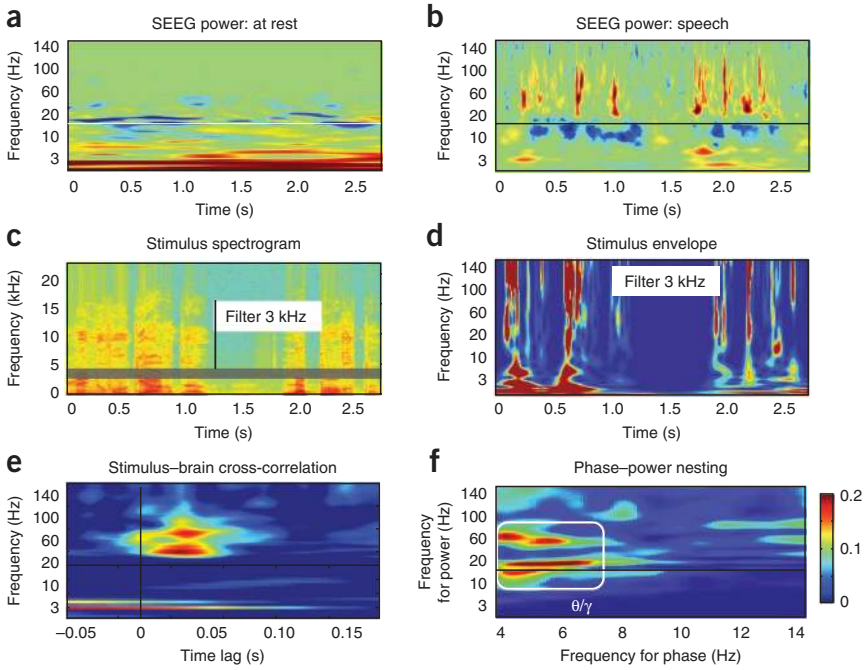

Figure 2 Speech–brain interaction from human intracortical recordings of primary auditory cortex. (**a**) Time–frequency representation of cortical activity at rest. (**b**) Time–frequency representation of cortical activity in response to the French spoken sentence "Le nouveau garde la porte." (**c**) Stimulus spectrogram, which shows spectro-temporal modulations and formant structure. (**d**) An example modulation spectrum extracted from a band centered around 3 kHz (bandwidth 0.5 kHz). To cross-correlate speech with the brain response, the broadband speech spectrum (1–5 kHz) was split into frequency bands (32 channels) from which the temporal envelope in the 1–140 Hz modulation range is extracted. In the band shown, modulations cover this entire range. (**e**) Auditory cortex power strongly correlates with speech modulations in two frequency bands, theta and gamma. The theta band aligns to speech with zero time lag; the gamma band reflects speech modulations after a 40-ms time lag. (**f**) An index of inter-trial phase consistency, which reflects frequency-specific locking between stimulus and brain. The cross-correlation between index and stimulus is an indicator of how oscillations phase-track speech amplitude modulations. White box, theta-gamma frequency nesting. These data provide experimental confirmation from human auditory cortex for the three first proposed operations (steps 1 to 3 in **Fig. 1**). SEEG, stereotactic EEG. Data courtesy of C. Liégeois-Chauvel, analyzed by B. Morillon, Y. Beigneux, L. Arnal, C. Bénar, C. Liégeois-Chauvel and A.-L.G.

as local field potential (LFP) signals from superficial and deep cortical layers (**Fig. 3a**). By contrast, stimulus-driven spikes are stronger in the intermediate layer IV (ref. 18), where thalamo-cortical fibers are densest[19]. Layer IV pyramidal cells in turn contact layer II/III pyramidal cells, whose axons reach layer IV of the next hierarchical stage[20]. This simplified input–output network is largely modulated by interneurons that are thought to be at the origin of oscillatory activity (**Fig. 3b**). The pyramidal-interneuron gamma (PING) network[21] is a state-of-the art model of brain oscillations that generates clustered spikes at a gamma rate. Neurons receiving input from a PING network exhibit a low firing probability for about 15 ms and a high one for the next 15 ms (compare **Fig. 1**), obviously approximate values as low gamma activity varies at rest in humans between less than 30 to about 40 Hz. This low-gamma intrinsic activity, called weak gamma, becomes stronger during auditory stimulation, as each individual neuron becomes more likely to fire at each cycle. The hypothesis that output spiking is temporally structured by stimulus-induced oscillatory activity is both anatomically and functionally plausible, and there is growing evidence that cognitive operations depend on spike timing and the alignment of spikes with the phase of oscillations[22]. There is, however, no direct evidence that oscillations affect spiking in those superficial neurons that provide input to the next hierarchical stage. This specific conjecture could be addressed by targeted experimentation in animals, including detailed analyses of microcolumn anatomical and functional connectivity. Critically, a comparison of spike timing in layers IV and II/III in early auditory cortical regions during continuous speech is required to establish the hypothesized input–output transformation.

We argue that a functional consequence of the modulation of vertical circuits by gamma and theta oscillators is the organization of spike timing[23] and the ensuing discretization of the cortical output[17]. Assuming 'continuous' spiking input to auditory cortex, the output signal is chunked by periodic modulation of the firing likelihood. Although the discretization process is presumably not the

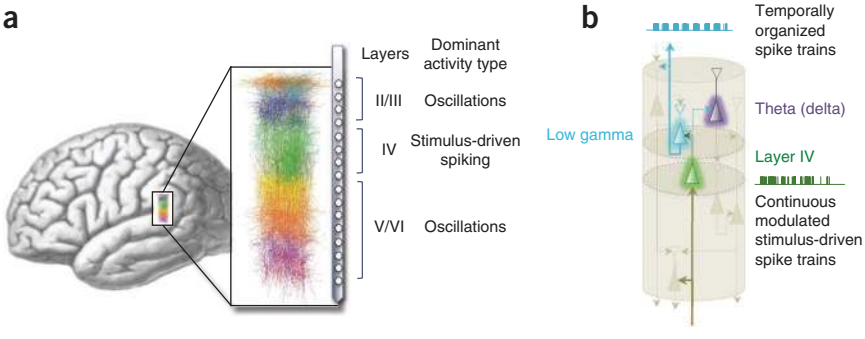

Figure 3 Generation of oscillations in a cortical column. (**a**) Schematic distribution of oscillatory and stimulus-driven spiking activity in a cortical column (courtesy M. Oberlaender, Max Planck Institute, Florida; modified from ref. 49). Oscillatory activity is typically detected in the superficial (II/III) and deep (V/VI) cortical layers, whereas stimulus-driven spiking is strongest in layer IV (right). (**b**) Cortical column networks that could underpin the operations depicted in **Figure 1**. We assume two populations of pyramidal neurons in superficial layers, one involved in low gamma generation, the other in theta generation. These populations are connected through an excitatory connection from theta to gamma (details in **Fig. 4**). Under the cumulative influence of theta and gamma oscillations, the spike train–reflecting activity in input layer IV is transformed in a discontinuous spike train in the superficial layer, which will be read out by the next hierarchical stage.

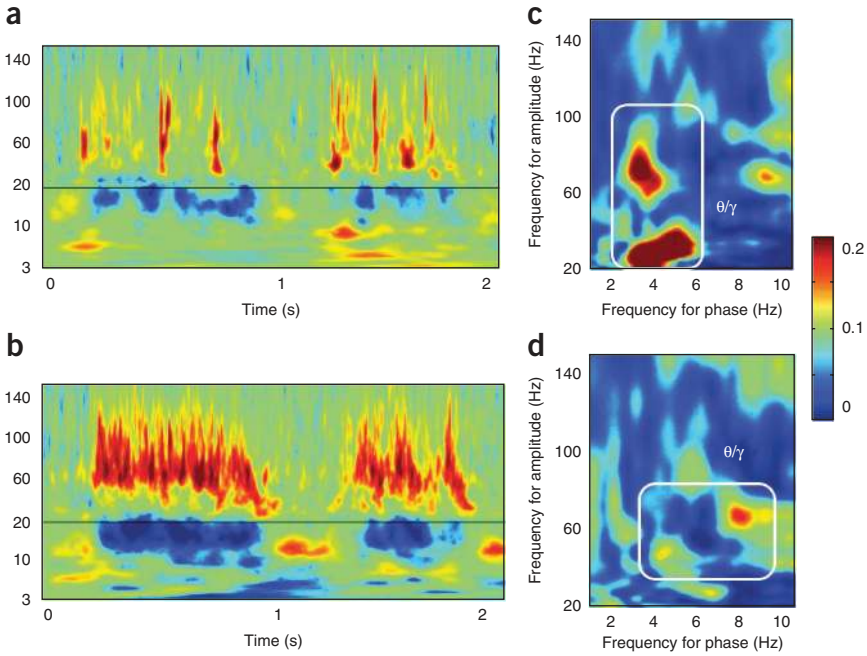

Figure 4 Comparison of neural responses in auditory primary and association (Brodmann area 22) cortices. (**a**,**b**) Time–frequency representations obtained from recordings made with stereotactic EEG in humans (see also **Fig. 2**) in response to a spoken sentence. (**c**,**d**) Theta phase–gamma power nesting. Although gamma power is stronger in association (lower panels) than in primary (upper panels) auditory cortex, it only tracks fast stimulus modulations in the primary region. Yet theta-gamma nesting (white box) is detectable in both areas, suggesting that gamma activity is controlled by the stimulus in primary auditory cortex but controlled by theta activity in the association area. Note that theta tracking is also slower in the association area, supporting the notion of downsampling when progressing in the auditory cortical hierarchy.

only—nor perhaps the essential—computational transformation at a given auditory cortical processing step, we argue that it is fundamental at early stages (A1 and A2).

One feature of signal discretization is to create, at the population level, alternations of time periods for sensory information integration and transmission. Metaphorically, neurons in superficial layers of A1 could count the number of spikes occurring in layer IV cells over half a gamma cycle and convey this averaged number to the next processing stage, which could express it directly by, for instance, emitting an analogous spike number. This results in a temporally structured output pattern, whereby both gamma phase and spike rate are relevant coding cues[23]. For such a scheme to work, a slow rhythm must integrate gamma-discretized information to perform second-level statistics. With respect to speech, psychophysical data[9] suggest integration over ~120 ms, which falls into the theta rhythm. Thus, a gamma-based code could be read out and integrated through a theta-based mechanism. In our proposal, spike discretization serves to present the stimulus in discrete chunks (segments) from which many different types of computations can be performed. Ultimately, the process permits phonological abstraction, generating discrete representations that make contact with spatially distributed phonemic and syllable representations underlying recognition[24,25]. Critical testing of these hypotheses requires demonstrating downsampling when progressing in the hierarchy—for example, from Brodmann areas 41 and 42 to Brodmann area 22. Both electrocorticography and stereotactic EEG recordings in humans indicate that Brodmann area 22 tracks speech modulations at theta and delta but no longer at gamma rates. Even though gamma activity is robustly detectable in Brodmann area 22 when a subject listens to connected speech, this activity is de-correlated from speech modulations, yet it remains controlled by theta activity (theta-gamma nesting, **Fig. 4**).

Alignment of neuronal excitability with speech modulations

We assume that gamma oscillations control neuronal excitability in superficial cortical layers, where the main output signals are emitted toward higher processing stages. How the periodicity in output neuronal excitability aligns with the stimulus is an open question. A logical proposal is that phases of high neuronal excitability in superficial layers coincide with the time periods when the most energetic parts of the speech signal reach layer IV. To address this issue, we developed a biophysical model of coupled theta and gamma oscillations adapted from previous theoretical work[26,27], in which theta-oscillating pyramidal-interneuron networks control gamma oscillating dyads (PING; **Fig. 5a**) through an excitatory connection.

Speech modulations (5–10 Hz) elicit a discharge in theta neurons, which then track the modulations in speech even when they are not fully periodic, or faster than the intrinsic theta rate (**Fig. 5b**). The excitation of PING by pyramidal-interneuron theta (PINT) networks sets a period of excitability that lasts about three or four gamma cycles[27], which is approximately the minimum duration of a syllable. Whereas the rate of theta follows the speech envelope rate, the rate of gamma does not change depending on input rate (in part owing to the restriction of speech modulations mostly to slow patterns, <10 Hz). The release of excitation from PINT neurons between syllables resets gamma oscillations, which enables time-locking of the gamma and output cells to the next syllable onset. The resulting response of output cells is discontinuous. For each gamma cycle, output neurons may fire or not fire, which constitutes a binary code reflecting the shape of the speech envelope. This model does not need to assume a direct reset of gamma oscillations by the stimulus. Detailed laminar analysis of spike timing[28] in awake humans using new recording methods while listening to speech could clarify whether the response in superficial layers indeed provides a discrete code, and more detailed modeling work should investigate how efficient such a code might be.

Asymmetric sampling

It has been suggested that the two rates at which the incoming signal is 'sampled' are at least in part laterally distributed[3]. This hypothesis shares many attributes with Zatorre and colleagues' spectral-temporal asymmetry model[29]. The main assumption is that gamma sampling dominates in left auditory cortex, underpinning neural computations on a 12.5–25 ms timescale, whereas theta sampling is assumed to dominate in right auditory cortex. Many functional magnetic resonance imaging (fMRI) experiments have tested the idea, largely supporting the conjecture that temporal processing at different timescales is associated with hemispherically asymmetric activation[30–32]. More data have been acquired with other experimental approaches (for example, EEG, MEG, combined EEG and fMRI, near infrared spectroscopy), and such asymmetries seem to be already present at rest in adult humans[7,33] and during auditory processing

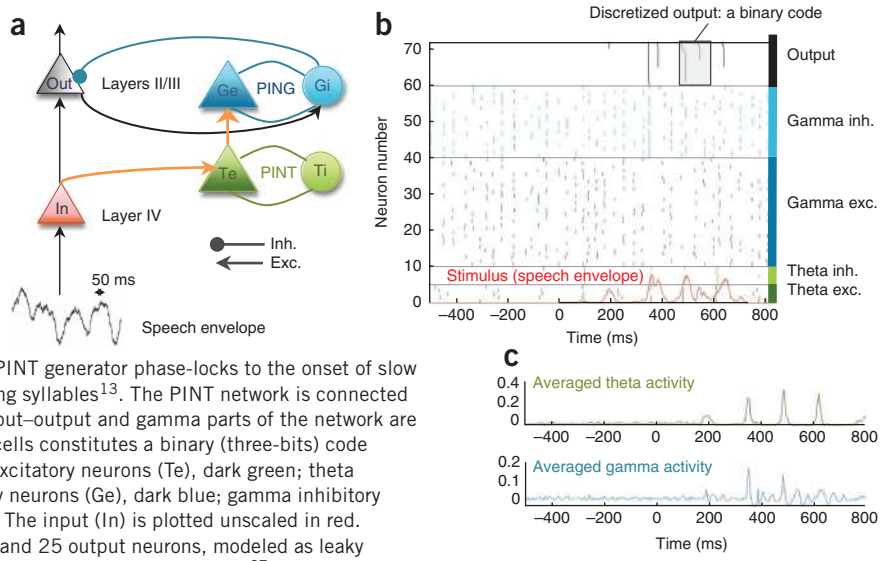

Figure 5 A biophysical model of coupled theta and gamma oscillations. (**a**) The model uses pyramidal-interneuron gamma (PING) and pyramidal-interneuron theta (PINT) networks, whereby oscillations at both frequencies are generated by the interaction between a pyramidal excitatory (exc.) population and an inhibitory (inh.) population. (**b**) Rastergram of the simulated network in response to an English sentence filtered through precortical auditory pathways[50]; the input corresponds to one channel centered on 1.5 kHz. The network exhibits intrinsic gamma and theta activities before the onset of the sentence, and gamma oscillations are modulated by theta rhythms[26]. The PINT generator phase-locks to the onset of slow modulations (5–10 Hz) in the speech signal, signaling syllables[13]. The PINT network is connected to the PING one by an excitatory connection. The input–output and gamma parts of the network are similar to those in ref. 27. The response of outputs cells constitutes a binary (three-bits) code reflecting the shape of the speech envelope. Theta excitatory neurons (Te), dark green; theta inhibitory neurons (Ti), light green; gamma excitatory neurons (Ge), dark blue; gamma inhibitory neurons (Gi), light blue; output neurons (Out), black. The input (In) is plotted unscaled in red. The network is composed of 5 Te, 5 Ti, 60 Ge, 20 Gi and 25 output neurons, modeled as leaky integrate-and-fire neurons, with Ge and output neurons having an extra m-current[27]. Synaptic release includes both synaptic rise and decay time constants. (**c**) Averaged oscillatory activity: theta activity phase-locks to the stimulus, and gamma activity follows speech envelope and theta activity. Model development and simulations by A. Hyafil, B. Gutkin, L. Fontolan, O. Ghitza and A.-L.G.

in infants[34]. Asymmetric oscillatory properties in auditory cortex can be related to cytoarchitectonic differences, showing more large pyramidal cells in cortical layer III in left than right auditory cortex and larger cortical columns and interpatch distances[35]. Given that gamma activity seems to originate in superficial layers[36] (**Fig. 3**), it is possible that differences in cytoarchitectonic organization are involved in an asymmetric oscillatory regime.

An important computational issue pertains to the process of integrating gamma-parsed segments into longer, syllable-length units. Because of sampling asymmetry, steady-state speech signals like vowels are better analyzed by right than left auditory cortex. Although this has been experimentally confirmed[14], it does not necessarily reflect a better linguistic analysis. Rather, analysis at slow rates (long time constants) allows for a more accurate spectral analysis that is essential for paralinguistic processes such as speaker identification. Analyses of vowels at short time scales by a gamma-dominant sampling in left auditory cortex is presumably sufficient for vowel identification in the context of speech processing. Although theta activity is detected at rest in both temporal cortices, its location in the left, over auditory association regions, is compatible with an integrative rather than simply a sampling function.

Dysfunctional oscillatory processes

Compelling evidence for our view would be to show that knocking out oscillatory mechanisms entails specific speech and language impairments. This is tricky because it is at present impossible to establish causal links between susceptibility genes involved in heritable language pathologies and cortical oscillations. Yet dyslexia, autism and specific language impairment are presumably good candidates to test this hypothesis, as they share structural and functional anomalies of the perisylvian region and even susceptibility genes[37]. These genes are involved in synaptogenesis, neuronal migration or ion channel formation and could hence influence oscillatory neuronal behavior. In dyslexia, more readily than in the two other pathologies, direct links between auditory oscillatory activity and reading disability can be envisaged. When the dyslexia-linked genes *KIAA0319* or *DCDC2* are deleted

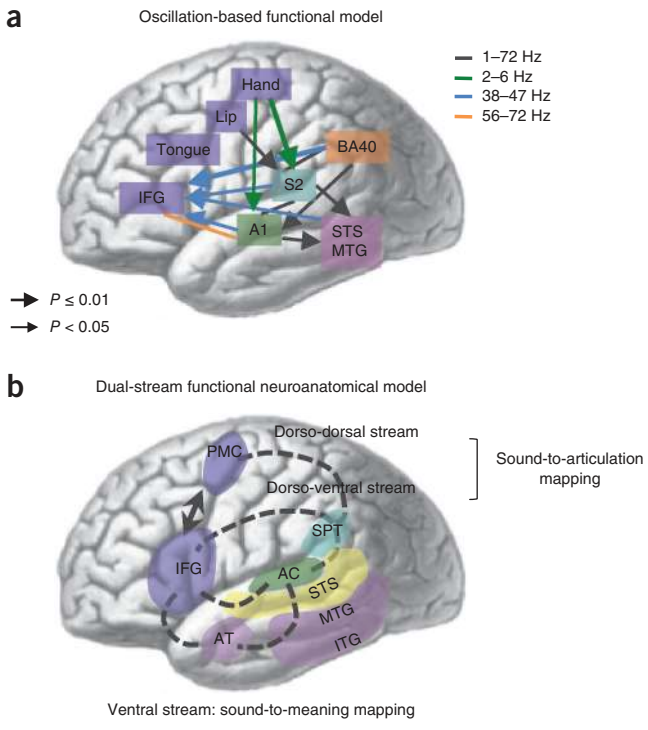

Figure 6 Functional anatomy of the speech processing network. (**a**) Anatomy as adapted from ref. 33, deriving largely from mapping of cortical oscillations. Stronger correlations between regions (thick arrows, $P \leq 0.01$) reflect stronger coupling between oscillatory activity in tested frequency bands and the BOLD response. (**b**) Anatomy as adapted from ref. 44, deriving largely from imaging and lesion–deficit data. Dotted lines illustrate the putative connectivity in the dorsal and ventral processing streams. Regions in the same color indicate areas implicated in oscillatory[33] (**a**) or imaging and lesion[44] (**b**) analyses. A1, primary auditory cortex; S2, secondary somatosensory cortex; BA40, Brodmann area 40 (supramarginal gyrus); STS, superior temporal sulcus; MTG, middle temporal gyrus; IFG, inferior frontal gyrus; PMC, premotor cortex; AT, anterior temporal cortex; AC, auditory cortex; SPT, sylvian parieto-temporal area; ITG, inferior temporal gyrus.

in animal models, neuronal migration is particularly disturbed in deep and superficial cortical layers where oscillations are generated[38,39]. Temporal sampling mediated by cortical oscillations has recently been proposed to be a central mechanism in several aspects of dyslexia[40]. This proposal emphasizes a deficit involving theta oscillations, impairing low temporal modulations tracking syllable coding and even multisensory processing. We argue in a complementary way for the possibility of a gamma oscillation deficit yielding an auditory phonemic deficit.

If people with dyslexia parse speech at a frequency slightly higher or lower than the usual low gamma rate, their phonemic representations could exhibit an idiosyncratic format. Phonemic units would be either undersampled or oversampled, without necessarily inducing major perceptual deficits[41,42]. This anomaly would selectively complicate the grapheme-to-phoneme matching, leaving speech perception and production unaffected. The phonological impairment could take different forms, with a stronger impact on the acoustic side for undersampling (insufficient acoustic detail per time unit) and on the memory side for oversampling (too many frames to be integrated per time unit).

Using auditory steady state responses, we observed that the left-dominant response around 30 Hz present in subjects with normal reading ability is absent in those with dyslexia, suggesting that the ability of their left auditory cortex to parse speech at the appropriate phonemic rate was altered. Those with dyslexia had a strong response at this frequency in right auditory cortex and therefore an abnormal asymmetry. The magnitude of the anomalous asymmetry correlated with behavioral measures in phonology (such as non-word repetition and rapid automatic naming). We also found that readers with dyslexia had a stronger resonance than controls in both left and right auditory cortices at frequencies between 50 and 80 Hz, suggesting that the deficit in these subjects was accompanied with phonemic oversampling. This oversampling positively correlated with a phonological memory deficit[43].

Although important, the observation that oscillatory anomalies co-occur with atypical phonological representations remains correlational. Causal evidence that auditory sampling is determined by cortical columnar organization could be obtained from knockout animal models comparing neuronal activity (multi-unit activity and LFP) to continuous auditory stimuli in sites with various degrees of columnar disorganization. Such animal work can, however, only indirectly address a specific relation to speech and language.

Cortical oscillations and language functional organization

Intrinsic asymmetries in cortical oscillations are observed not only in auditory cortex. The phenomenon is even more marked in motor cortex, specifically in tongue, lip and hand regions, where theta and low and high gamma activity appear strongly left dominant[33]. Resting oscillatory activity is also stronger in a left than right inferior parietal region, Brodmann area 40. Using graph theoretical analyses on combined EEG and fMRI data, we delineated a core network where oscillations are asymmetric at rest, including A1 (Brodmann areas 41 and 42), the somatosensory cortex, the articulatory motor cortex and Brodmann area 40. Of note, we did not observe left-dominant oscillatory asymmetry in the posterior superior temporal region (Brodmann area 22; Wernicke's area) and in the inferior frontal cortex (Brodmann areas 44 and 45; Broca's region). This is peculiar, as these two regions have a strong asymmetric function during language processing. The finding suggests either that oscillations do not contribute to the function of these regions during linguistic processing or that oscillatory activity is absent at rest but acquired during processing. We confirmed the latter, showing that Wernicke's and Broca's regions 'inherit' oscillatory asymmetries during linguistic processing from the core network. The oscillation-based topographic distribution of the speech processing network (**Fig. 6a**) accords well with standard descriptions of the functional anatomy of speech processing[44] (**Fig. 6b**).

How specific is an oscillation-based parsing model?

Neuronal oscillations, especially in the ranges discussed here, are ubiquitous, and the time scales we implicate for perceptual analysis are demonstrable in other cases as well, including that of vision[45]. This suggests that the hypotheses we put forth have the potential to generalize across domains. However, our immediate concern is more narrow: the model we propose is specific to speech processing insofar as speech modulations are produced by quasiperiodic cortical motor commands whose time constants more likely match those of the auditory cortex than do those of other acoustic stimuli. This may be the case because most of the human neocortex works on preferred frequency channels—for example, gamma—or more specifically because auditory cortex is tuned throughout development by periodic efferent input from the premotor cortex in anticipation of spoken speech[46]. Tuning between left premotor and auditory cortices in the low gamma—that is, phonemic—range can be visualized, for instance, using auditory steady state responses[43]. Such optimal stimulus–brain alignment is hard to identify in other cognitive contexts, although music and the analysis of conspecific signals may be considered candidates. In that sense, speech is only a good model. An even more compelling case is audio-visual speech, where stimulus periodicity is generated and apprehended by two independent sensory streams and then cross-modally unified using specific discretization and integration schemes[47]. More generally, though, every process relying on proactive behavior—for example, active sensing[48]—presumably relies on similar mechanisms whereby the sensory intake of continuously varying stimuli is framed by the temporal characteristics of an associated motor behavior.

In this article, we have articulated a set of hypotheses to investigate the relation between the perception of connected speech and neurobiological mechanisms. We developed a model at anatomic (**Figs. 3** and **6**), physiological (**Figs. 1**, **2** and **4**) and computational (**Fig. 5**) levels. At the center of the research program lies the assumption that cortical oscillations provide ways to temporally organize the incoming speech signal. The main emerging principles are that two prerequisites for constructing intelligible representations of the speech stream are phase-locking between stimulus and cortex in (at least) two discrete time domains and the hierarchical coupling of related cortical oscillations during speech processing.

ACKNOWLEDGMENTS

We are deeply grateful to C. Liegeois-Chauvel for providing stereotactic EEG data and C.-G. Bénar for related methodological support. In A.-L.G.'s team we thank Y. Beigneux, B. Morillon and L. Arnal, who analyzed these data; A. Hyafil, C. Kapdebon and L. Fontolan, who carried out the computational modeling work; and K. Lehongre and D. Roussillon, who conducted the experiments in human subjects. In D.P.'s team, we thank H. Luo, M. Howard and G. Cogan for pioneering work and many discussions of these issues. We also thank our colleagues O. Ghitza, S. Greenberg, B. Gutkin, V. Wyart, C. Lorenzi, F. Ramus and C. Schroeder for motivating and discussing various aspects of this research. This work is supported by the Centre National de la Recherche Scientifique of France and the European Research Council (A.-L.G.), and US National Institutes of Health grant 2R01 DC05660 (D.P.).

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/natureneuroscience/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Heimbauer, L.A., Beran, M.J. & Owren, M.J. A chimpanzee recognizes synthetic speech with significantly reduced acoustic cues to phonetic content. *Curr. Biol.* **21**, 1210–1214 (2011).
2. Liberman, A.M. & Mattingly, I.G. The motor theory of speech perception revised. *Cognition* **21**, 1–36 (1985).
3. Poeppel, D. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun.* **41**, 245–255 (2003).
4. Shannon, R.V. *et al.* Speech recognition with primarily temporal cues. *Science* **270**, 303–304 (1995).
5. Lorenzi, C. *et al.* Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proc. Natl. Acad. Sci. USA* **103**, 18866–18869 (2006).
6. Adank, P. & Janse, E. Perceptual learning of time-compressed and natural fast speech. *J. Acoust. Soc. Am.* **126**, 2649–2659 (2009).
7. Giraud, A.L. *et al.* Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* **56**, 1127–1134 (2007).
8. Ghitza, O. Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front. Psychol.* **2**, 130 (2011).
9. Ghitza, O. & Greenberg, S. 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 (2009).
10. Liégeois-Chauvel, C. *et al.* Temporal envelope processing in the human left and right auditory cortices. *Cereb. Cortex* **14**, 731–740 (2004).
11. Ding, N. & Simon, J.Z. Neural representations of complex temporal modulations in the human auditory cortex. *J. Neurophysiol.* **102**, 2731–2743 (2009).
12. Luo, H. & Poeppel, D. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* **54**, 1001–1010 (2007).
13. Ahissar, E. *et al.* Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl. Acad. Sci. USA* **98**, 13367–13372 (2001).
14. Abrams, D.A. *et al.* Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *J. Neurosci.* **28**, 3958–3965 (2008).
15. Nourski, K.V. *et al.* Temporal envelope of time-compressed speech represented in the human auditory cortex. *J. Neurosci.* **29**, 15564–15574 (2009).
16. Canolty, R.T. & Knight, R.T. The functional role of cross-frequency coupling. *Trends Cogn. Sci.* **14**, 506–515 (2010).
17. Schroeder, C.E. & Lakatos, P. Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* **32**, 9–18 (2009).
18. Atencio, C.A., Sharpee, T.O. & Schreiner, C.E. Cooperative nonlinearities in auditory cortical neurons. *Neuron* **58**, 956–966 (2008).
19. Sakata, S. & Harris, K.D. Laminar structure of spontaneous and sensory-evoked population activity in auditory cortex. *Neuron* **64**, 404–418 (2009).
20. Wang, X.J. Neurophysiological and computational principles of cortical rhythms in cognition. *Physiol. Rev.* **90**, 1195–1268 (2010).
21. Börgers, C., Epstein, S. & Kopell, N.J. Background gamma rhythmicity and attention in cortical local circuits: a computational study. *Proc. Natl. Acad. Sci. USA* **102**, 7002–7007 (2005).
22. Fries, P., Nikolic, D. & Singer, W. The gamma cycle. *Trends Neurosci.* **30**, 309–316 (2007).
23. Kayser, C., Logothetis, N.K. & Panzeri, S. Millisecond encoding precision of auditory cortex neurons. *Proc. Natl. Acad. Sci. USA* **107**, 16976–16981 (2010).
24. Chang, E.F. *et al.* Categorical speech representation in human superior temporal gyrus. *Nat. Neurosci.* **13**, 1428–1432 (2010).
25. Rauschecker, J.P. & Scott, S.K. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* **12**, 718–724 (2009).
26. Kopell, N. *et al.* Gamma and theta rhythms in biophysical models of hippocampal circuits. in *Hippocampal Microcircuits: A Computational Modeller's Resource Book* (eds. Cutsuridis, V., Graham, B.P., Cobb, S. & Vida, I.) Ch 15 (Springer, 2011).
27. Shamir, M. *et al.* Representation of time-varying stimuli by a network exhibiting oscillations on a faster time scale. *PLoS Comput. Biol.* **5**, e1000370 (2009).
28. Atencio, C.A. & Schreiner, C.E. Columnar connectivity and laminar processing in cat primary auditory cortex. *PLoS ONE* **5**, e9521 (2010).
29. Zatorre, R.J., Belin, P. & Penhune, V.B. Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.* **6**, 37–46 (2002).
30. Boemio, A. *et al.* Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat. Neurosci.* **8**, 389–395 (2005).
31. Jamison, H.L. *et al.* Hemispheric specialization for processing auditory nonspeech stimuli. *Cereb. Cortex* **16**, 1266–1275 (2006).
32. Obleser, J., Eisner, F. & Kotz, S.A. Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J. Neurosci.* **28**, 8116–8123 (2008).
33. Morillon, B. *et al.* Neurophysiological origin of human brain asymmetry for speech and language. *Proc. Natl. Acad. Sci. USA* **107**, 18688–18693 (2010).
34. Telkemeyer, S. *et al.* Sensitivity of newborn auditory cortex to the temporal structure of sounds. *J. Neurosci.* **29**, 14726–14733 (2009).
35. Hutsler, J. & Galuske, R.A. Hemispheric asymmetries in cerebral cortical networks. *Trends Neurosci.* **26**, 429–435 (2003).
36. Gireesh, E.D. & Plenz, D. Neuronal avalanches organize as nested theta- and beta/gamma-oscillations during development of cortical layer 2/3. *Proc. Natl. Acad. Sci. USA* **105**, 7576–7581 (2008).
37. Pagnamenta, A.T. *et al.* Characterization of a family with rare deletions in CNTNAP5 and DOCK4 suggests novel risk loci for autism and dyslexia. *Biol. Psychiatry* **68**, 320–328 (2010).
38. Peschansky, V.J. *et al.* The effect of variation in expression of the candidate dyslexia susceptibility gene homolog Kiaa0319 on neuronal migration and dendritic morphology in the rat. *Cereb. Cortex* **20**, 884–897 (2010).
39. Wang, Y. *et al.* Dcdc2 knockout mice display exacerbated developmental disruptions following knockdown of doublecortin. *Neuroscience* **190**, 398–408 (2011).
40. Goswami, U. A temporal sampling framework for developmental dyslexia. *Trends Cogn. Sci.* **15**, 3–10 (2011).
41. Ramus, F. & Szenkovits, G. What phonological deficit? *Q. J. Exp. Psychol. (Hove)* **61**, 129–141 (2008).
42. Ziegler, J.C. *et al.* Speech-perception-in-noise deficits in dyslexia. *Dev. Sci.* **12**, 732–745 (2009).
43. Lehongre, K. *et al.* Altered low-gamma sampling in auditory cortex accounts for the three main facets of dyslexia. *Neuron* **72**, 1080–1090 (2011).
44. Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402 (2007).
45. Holcombe, A.O. Seeing slow and seeing fast: two limits on perception. *Trends Cogn. Sci.* **13**, 216–221 (2009).
46. Eliades, S.J. & Wang, X. Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature* **453**, 1102–1106 (2008).
47. Chandrasekaran, C. *et al.* Monkeys and humans share a common computation for face/voice integration. *PLoS Comput. Biol.* **7**, e1002165 (2011).
48. Schroeder, C.E. *et al.* Dynamics of active sensing and perceptual selection. *Curr. Opin. Neurobiol.* **20**, 172–176 (2010).
49. Oberlaender, M. *et al.* Cell type-specific three-dimensional structure of thalamocortical circuits in a column of rat vibrissal cortex. *Cereb. Cortex* doi:10.1093/cercor/bhr317 (16 November 2011).
50. Chi, T., Ru, P. & Shamma, S.A. Multiresolution spectrotemporal analysis of complex sounds. *J. Acoust. Soc. Am.* **118**, 887–906 (2005).