

Preprint copy of:

Neural Entrainment and Attentional Selection in the Listening Brain

Jonas Obleser^{1,3,*} and Christoph Kayser^{2,3,*}

The streams of sounds we typically attend to abound in acoustic regularities. Neural entrainment is seen as an important mechanism that the listening brain exploits to attune to these regularities and to enhance the representation of attended sounds. We delineate the neurophysiology underlying this mechanism and review entrainment alongside its more pragmatic signature, often called ‘speech tracking’. The latter has become a popular analytical approach to trace the reflection of acoustic and linguistic information at different levels of granularity, from neurophysiology to neuroimaging. As we discuss, the concept of entrainment offers both a putative neurophysiological mechanism for selective listening and a versatile window onto the neural basis of hearing and speech comprehension.

What Do We Mean by Entrainment?

Imagine yourself immersed in the babble of multiple talkers around you. In this common situation, your goal will likely be to listen to, and to communicate with, one of these talkers. For normal-hearing individuals, this task can pose a challenge, but it rarely prevents us from communicating. For the hearing-impaired listener, however, as well as for any artificial intelligence system, it can border on being intractable. Over the past decade, cognitive neuroscience has been concerned with one apparent feature (or, more contentiously, a key mechanism) used by the brain to resolve this multiple-talker disarray: ‘neural **entrainment**’ (see [Glossary](#)).

Put in the simplest terms, neural entrainment refers to the temporal alignment of an observed neural process with the regularities in an exogenously occurring stimulus, such as speech, where the sensory regularities may arise from syllabic or word-level boundaries or other articulatory cues ([Figure 1A](#)).

Often the underlying neural process is supposed to arise from a truly rhythmic neural generator, the phase of which resets relative to the occurrence of specific sensory events [[1](#)]. The question of whether the neural process is indeed oscillatory in a physical sense matters for the precise mechanistic interpretations of this concept ([Box 1](#)). However, as we discuss, it may be less important for some practical applications. In the literature, the term neural entrainment is sometimes used synonymously with ‘**phase coding**’ or ‘**speech tracking**’, but as we emphasize, these refer to distinct aspects of the underlying neural or phenomenological processes.

How relevant is the broader concept of neural entrainment to the functioning of the auditory brain and to us making sense of the auditory world? We first look closer at the evidence for whether and how the timing of brain activity becomes phase aligned with sensory input. Second, we ask, from the vantage point of the listening brain, what adaptive advantage does neural entrainment offer; for example, when attending to one of multiple talkers? Finally, we look at recent developments in applying the concept of entrained activity to augment comprehension and cognitive function in general.

We conclude that entrainment in most general terms comprises both a putative neurophysiological mechanism ([Box 1](#)) and a theoretical concept that bridges micro- and macroscopic spatial scales of brain activity. Not least, however, it represents a versatile tool to trace increasingly sophisticated representations of acoustic and linguistic information in the listening brain.

A Neural Phase Code for Sensory Information

In 2007, Luo and Poeppel published a landmark study [[2](#)] in which they showed that a set of spoken sentences could be distinguished from each other based solely on the time series of the phase of a narrow-band signal extracted from human auditory brain activity. The word was out: dynamic brain

Highlights

Electrophysiological studies suggest that entrainment of rhythmic activity can be a key mechanism to control neural sensory gain.

Selective entrainment reflects exogenous stimulus attributes and the endogenous selection of task-relevant attended information.

The neural tracking of multiple acoustic and linguistic features sheds light on their differential encoding and how they shape perception.

Entrainment as a concept bridges across temporal and spatial scales of brain activity.

¹Department of Psychology, University of Lübeck, 23562 Lübeck, Germany

²Department for Cognitive Neuroscience and Cognitive Interaction Technology, Center of Excellence, Bielefeld University, 33615 Bielefeld, Germany

³The authors contributed equally

*Correspondence:
jonas.obleser@uni-luebeck.de,
christoph.kayser@uni-bielefeld.de

activity at the timescale of around 4–8 Hz (**theta** band) temporally aligns (i.e., becomes phase entrained or **phase resets**) to the dynamic regularities in speech. Since then, this measurable form of phase coding of speech information in theta-band activity has provided a window onto the neural representation of speech. It also paved the way toward a more general understanding of how competing acoustic streams are represented by the listening brain.

Moreover, a swift development toward understanding and exploiting noninvasively recorded **magnetoencephalography (MEG)** or **electroencephalography (EEG)** signals with much greater refinement was put in motion. The idea that temporally entrained, possibly oscillatory or at least pseudorhythmic, neural activity might align with systematic regularities in the environment (as depicted in [Figure 1A](#)) has been a highly consequential narrative in auditory neuroscience – despite the omnipresent questions of whether the observed activity arises from a neural oscillator and whether the oscillatory nature is critical for the presumed function ([Box 1](#)).

Part of this success story lies in technical developments. First, recording and analysis of continuous local activity obtained directly from within a brain area [e.g., **local field potentials**, **electrocorticography (ECoG)** signals] has become more prominent as the properties and the precise neural origins of these signals are better understood [3]. Second, the relation of these only invasively obtained signals to their noninvasive counterparts in MEG/EEG signals has become clearer [4], allowing insights to bridge across mesoscopic and macroscopic scales [5]. Third, the analysis of all of these signals has grown considerably in sophistication due to advances in signal processing and improvements in the quantification of statistical associations in high-dimensional datasets (e.g., [6]).

On a more practical note, this discovery has provided an important lesson: to study the neural encoding of complex sounds such as speech, it is not necessary to segment these into isolated chunks (e.g., words, syllables) and assume that each evokes orderly distinct and measurable evoked responses [7]. Rather, the timing of the entrained neural signal (i.e., its phase) proved sufficient to decode acoustic and linguistic information from brain activity. Thus, studying entrainment allows researchers to characterize multiple and distinct aspects of sound encoding and perception based on real-life stimuli in a single experiment.

How Can Entrainment Facilitate Selective Encoding in Auditory Cortex?

To understand the putative functional roles of entrainment in auditory perception, we need to assess their underlying local electrophysiological correlates. Why and how is it that entrainment may facilitate selective encoding in auditory cortex?

First and importantly, ecological sounds such as speech abound in temporal regularities that are amenable to the concept of entrainment, with certain timescales of acoustic-energy modulation being more relevant than others for conveying conspecific vocalizations such as speech ([8]; cf. [9]). Temporal modulation rates below 10 cycles per second as well as spectral modulation rates below ~2 cycles per kilohertz [10] carry most information about speech, but also about many natural and non-human animal sounds [11]. Notably, temporally structured neural activity at these timescales (the so-called **delta**, ~0.5–4 Hz, and **theta**, 4–8 Hz, bands) is routinely seen in alert and anaesthetized animals as well as in the human brain, as shown by recordings of local field potentials and ECoG and MEG/EEG data [12–16]. In the brain, this spectral energy is present both during silence (i.e., in spontaneous activity) and during acoustic stimulation, suggesting that the underlying neural processes may, at least in some instances, reflect an endogenous oscillatory signal. Still, it remains debated to what degree such activity is directly reflective of a genuinely rhythmic (i.e., oscillatory) neural process [17,18]. In many circumstances, it is more likely to constitute a **pseudorhythmic signal** brought about by focusing on a specific timescale using data filtering, which may, for example, arise from a sequence of responses evoked by the regularities in the acoustic input ([Box 1](#)).

In spite of this ambiguity, there is clear evidence that links slow network activity mechanistically to the overall function of the auditory brain. The phase of such low-frequency activity systematically predicts

Glossary

Alpha, delta, theta, gamma bands: typical timescales associated with prominent rhythmic activity in the brain, often defined as: alpha, 8–12 Hz; delta, 1–4 Hz; theta, 4–8 Hz; and gamma, above 30 Hz (see ‘Oscillatory activity’).

Attention, selective attention: the mechanism by which the brain selects one of several stimuli and enhances or prioritizes its processing over that of others. This is in distinction to divided attention, where the aim is to near-simultaneously capture the information provided by two or more stimuli.

Electrocorticography (ECoG): a measure of brain activity obtained using invasive electrophysiological recordings, usually obtained in human patients in the context of planned, epilepsy-related brain surgery; likely to pick up primarily local field potentials.

Electroencephalography (EEG): a noninvasive recording of electrical brain activity from the scalp.

Entrainment, phase entrainment: the phase of one (e.g., neural) possibly oscillating system becomes coupled to the phase of another possibly oscillating system (e.g., sensory input). See [Box 1](#) for a more detailed definition. The phase entrainment may, for example, be induced by a phase reset.

Evoked versus induced activity: a key distinction in electrophysiology, denoting in the most general terms the distinction of ongoing, endogenous (‘induced’) versus stimulus-evoked, exogenous (‘evoked’) activity. A second common interpretation of these terms puts emphasis on ‘induced’ referring to stimulus-related but not strictly stimulus-locked activity, while ‘evoked’ refers to a strictly stimulus-locked response.

Gain, gain function: the mathematical relationship describing the quantitative relation between a neuron’s membrane potential (reflecting the sum of the synaptic inputs) and its firing rate (reflecting its output).

Local field potential: the mesoscopic commensurate electrical activity produced by all synaptic and somatic potentials in the vicinity of the electrode.

Magnetoencephalography (MEG): noninvasive recording of the magnetic field induced by electrical

modulations in high-frequency (**gamma band**) **oscillations** and the firing rates of individual neurons [12,19–21] (Figure 1B). Importantly, note that this association between the phase of low-frequency network activity and the behavior of individual neurons may be independent of whether a genuine endogenous oscillatory process underlies the observed phase dynamics. Hence, issues of endogenous versus evoked network activity notwithstanding, phase coding possibly reflects a generic process that ties slow network activity with sensory encoding.

Relevant for a functional interpretation, this local spike–network coupling is specific to the stimulus [22]. That is, it carries information about the nature of individual sounds. Such a spike–network relation constitutes a phase-coding mechanism at the level of individual neurons and theoretical studies suggest that this mechanism may contribute to the noise-resilient encoding of acoustic information [23].

Is Entrainment under Cognitive Control?

Think of a listener actively aiming to isolate or ‘track’ one talker among many or against a backdrop of environmental sounds. Here it becomes relevant whether this neuron-to-network coupling is reflexive and whether it is under cognitive control. Thus, if the listener adjusts her goals (i.e., switches **attention** to another talker), does this bring about observable adjustments to this coupling? If this is the case, a key mechanism would arise for controlling which acoustic information is selectively processed: an adjustment of the relative timing of the entrained activity could shape how neurons respond to individual sounds among a complex acoustic scene. For instance, inducing a stable phase relation between a series of syllables and the low-frequency activity in auditory cortex could enhance the fidelity with which specific syllables are encoded (Figure 1).

What could be the direct mechanisms that link **phase entrainment** to the representation of acoustic information? One suggestion has been that the phase of slow rhythmic activity directly reflects the **gain** (e.g., the excitability) of local neurons (Figure 1C,D); thereby, a neuron selectively amplifies the acoustic information received during one phase of the entrained network activity over the information received during another [1,24]. In this scenario, the phase of entrained activity may shape the selectivity of what information neurons represent. Alternatively, the phase of the entrained activity may determine the overall level of neural firing without altering the specificity of how information is encoded (Figure 1D). Still, such an unspecific bias may offer computational advantages; for example, by increasing the influence on downstream regions.

A combined electrophysiological and modeling study in rodents provided evidence for both: while delta phase in the **primary auditory cortex (A1)** reflected changes in stimulus-unrelated bias, the phase of theta/alpha activity reflected changes in sensory gain [19] (Figure 1C,D). This suggests that entrained activity at different timescales reflects functionally distinct mechanisms. Critically, this study did not clearly demonstrate that the observed delta-band activity arises from an endogenous oscillatory process, and hence it may reflect neural entrainment in the broad sense.

Conceptually, during selective listening, local auditory cortical activity may entrain to an attended sound frequency, if this contains some rhythmicity in its mode of presentation, and implement the selective amplification of the relevant information. Experimental data support a direct link between specific cell types in A1, the phase alignment of delta-band activity, and perceptual benefits for discriminating sounds depending on their timing relative to this delta-band activity [25]. If the underlying neural activity arises from an endogenous oscillatory process, the rhythmicity of this process directly predicts the temporally specific amplification or attenuation of upcoming sounds over several hundreds of milliseconds. If the underlying activity reflects series of evoked responses, this forward-directed attenuation versus amplification of future sounds will be restricted to a shorter time window, corresponding to the typical autocorrelation time of the local **evoked activity**.

How Feature Specific Is Entrainment in Auditory Cortex?

Support for an active role of entrainment comes from studies demonstrating the locally specific entrainment of A1 neurons tuned to the attended sound frequency [26]. For example, A1 neurons

brain activity outside the scalp (see ‘Electroencephalography’).

Oscillation, oscillatory activity: a perfectly rhythmic (sinusoidal) signal with a defined timescale (frequency), amplitude, and phase. The phase expresses the position of peaks (troughs) relative to some reference.

Phase coding: used either to denote the phase entrainment of a continuous brain signal or to denote the consistent timing of action potentials relative to the phase of the local field potential.

Phase reset: a neural or other oscillator can exhibit a discontinuity (‘jump’) in its instantaneous phase, often in response to a resetting event (e.g., the occurrence of an exogenous stimulus).

Power entrainment: the power envelope (i.e., the slow changes in the power of an oscillatory signal) becomes coupled to the dynamics of another system (e.g., sensory input).

Primary auditory cortex (A1): a functionally and anatomically definable area, which forms the first cortical target of afferent auditory nerve fibers and which in the human brain is located along Heschl’s gyrus.

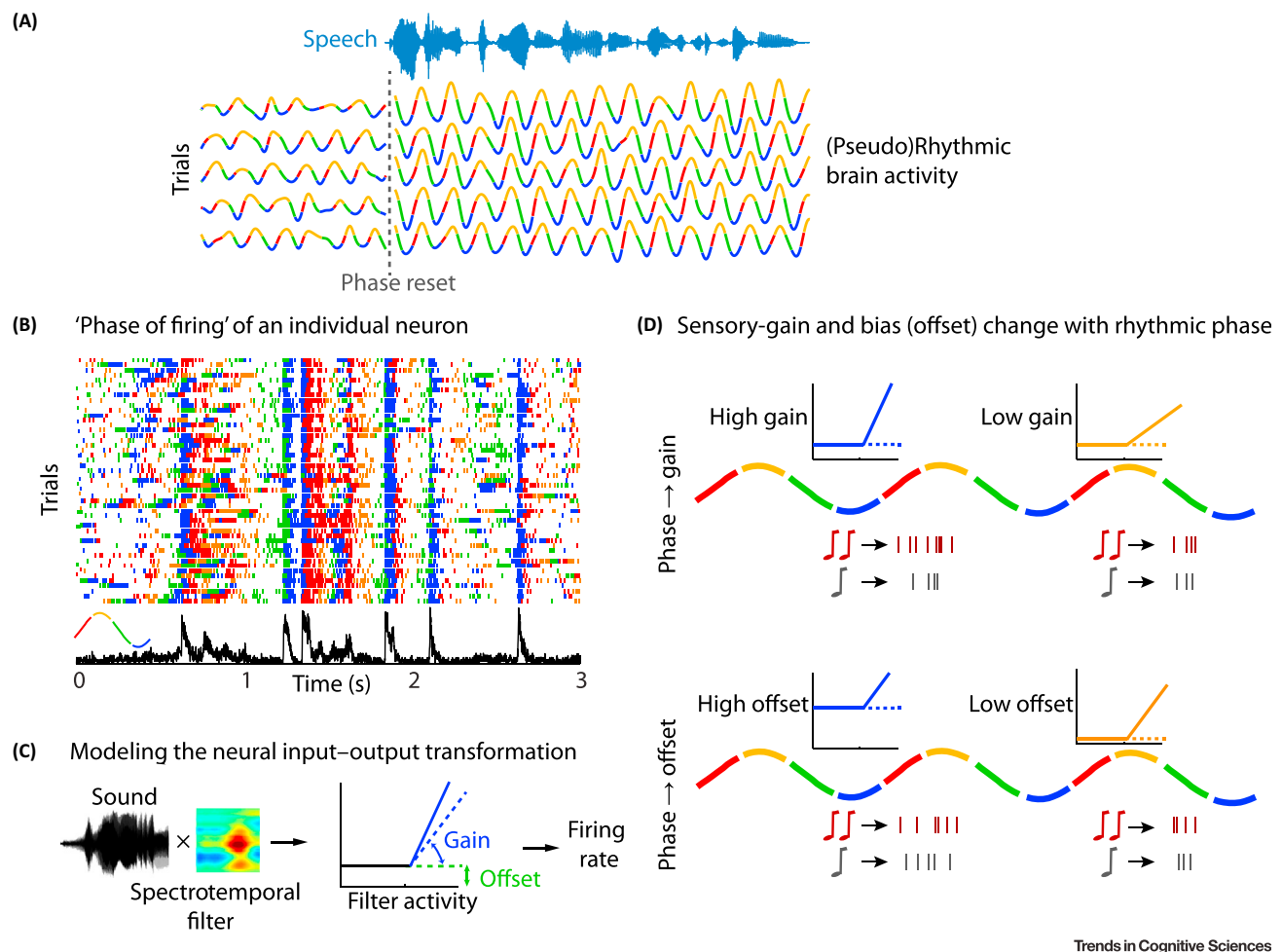
Pseudorhythmic signal: a temporally structured signal that evolves around a dominant timescale, but which is not strictly oscillatory; for example, obtained by band-pass filtering of a continuous wide-band signal.

Speech tracking, neural tracking: a loose term subsuming the quantitative and qualitative descriptions of a significant relation between brain activity and a sensory stimulus such as speech; quantified, for example, using phase entrainment or power entrainment.

Superior temporal gyrus (STG): a large part of the temporal lobe, known to contain primarily auditory regions; often used in distinction from A1.

System identification: a family of diverse mathematical techniques that all seek to describe the input–output function of dynamical systems.

Tonotopic organization: The functional organization of (primary) auditory regions, by which neurons preferentially responding to a specific sound frequency are organized spatially along a frequency gradient.



Trends in Cognitive Sciences

Figure 1. Entrainment and Neural Phase Coding in the Auditory Cortex.

(A) Schematic illustrating the entrainment of an endogenous signal, such as a delta-band field potential, to a stream of sounds. The phase entrainment between brain and stimulus is here visualized by the consistent alignment of phase (color coded) across trials to the stimulus, reflecting, for example, the alignment of brain activity to syllabic, word-level, or prosodic cues. This is also known as phase coding of the stimulus by the specific temporal pattern of the phase along time. Mechanistically, this could, for example, be implemented by the phase reset of an endogenous oscillatory signal by the stimulus. (B) The firing of auditory cortex neurons is often systematically timed (i.e., aligned) to the ongoing network activity. The figure shows the action potentials of one neuron recorded in the rat primary auditory cortex (A1) in response to many repeats (rows) of the same stimulus, color coded according to the phase of the local field potential in the delta band (sinusoidal inset). The shown neuron is mostly active around the (preferred) blue–red phase. Adapted, with permission, from [19]. (C) Computational model (a so-called linear–nonlinear model) for the response of individual neurons, based on a filter characterizing the spectral and temporal sensitivity profile of each neuron, and an output function linking the filter activity to the observed firing rate. The output function is described by a bias (offset) and a slope (gain). Theoretically, the bias and gain could be constant at all times; alternatively, they could vary (e.g., with the state of entrained local network activity), implementing the selective amplification of appropriately timed stimuli. (D) Experimental evidence shows that actual observed neural responses in A1 are best explained by neural encoding models that systematically vary the gain of the output function relative to the theta/alpha phase and the bias relative to the delta phase of A1 field potentials [19]. This provides two mechanisms by which the response to a sequence of sounds is shaped, affecting the level of overall activity and the specificity at which acoustic information is encoded.

selective to low frequencies align with a regular acoustic stream specifically when low frequencies are attended, and vice versa for **selective attention** to high frequencies [27,28]. Moreover, when stimulated with not strictly rhythmic, speech-like stimuli, such as consonant–vowel sequences, high-frequency regions align their preferred phase to the transient while low-frequency regions align to the vowel. This effectively induces a phase-opposing entrainment across A1 [29,30].

Hence, it seems that entrainment *per se* operates in a highly specific manner along the **tonotopic organization** of A1 and may reflect an adaptation to the regularities of ecologically relevant sounds. Thus, locally, entrainment may be a core ingredient of selective attention that effectively acts as a filter enhancing the neural encoding of specific sounds defined by their relative timing and spectral features [31] (Figure 2). It remains to be seen whether such phase-opposing patterns reflect a fixed frequency separation or can adapt to the statistical properties of momentarily relevant sounds.

These results lend themselves to speculation that the state of entrained activity should have a direct perceptual consequence. Neuroimaging studies have shown that the power and phase of specific EEG signals, even when these are not induced by strictly regular stimuli, correlate with the perceptual performance by which human listeners detect or discriminate sounds, with similar results being found for other sensory modalities [32]. Thus, rhythmic or entrained brain activity may be one mechanism by which perception is shaped into what some consider a 'rhythmic mode of perception' [33–37].

At the same time, these insights about the sound-frequency-specific and temporally opposed entrainment of A1 activity complicate the interpretation of MEG/EEG data (Box 2). We cannot attribute the spatially unselective entrainment quantified in these measurements to a localized neural origin. This lack of spatial specificity raises the question of what precisely the entrainment observed in MEG/EEG studies reflects. Future work, bridging these spatial scales and more refined modeling of neuroimaging data, will be required to close this gap – which is considerable, as the next section illustrates.

What Are the Neural Generators of Entrained Activity?

Above and beyond the notion of entrainment via phase coding as a mechanism for selecting or prioritizing attended stimuli, the precise physiological and neurostructural origins of the observed entrainment have proved hard to unravel. It is critical to understand whether entrainment arises from the coupling between the external stimulus and a genuine intrinsic neural oscillator (so-called 'induced oscillations') or whether the observed entrainment reflects a sequence of evoked responses (Box 1). That is, the apparently regular pattern of neural activity may arise from an intrinsic oscillator that is not driven by a periodic stimulus or may arise solely as a result of a (pseudo)rhythmic stimulus. The answer may differ between brain regions and timescales. Modeling studies and *in vivo* data suggest that one prominent pattern of rhythmic brain activity implied in visual attention, parieto-occipital alpha-band activity, arises from a genuine oscillatory process [38,39]. For the hearing-relevant delta/theta-band activity, this has proved harder to establish empirically (but see [14,40,41]).

A second, and mechanistically equally important, question is whether the observed entrainment arises from feedforward transmission between the sensory periphery and higher brain structures. Alternatively, recurrent connections between the auditory cortex, the thalamus, and possibly other midbrain regions might instantiate entrainment. In the visual system, the collective evidence suggests that alpha-band activity can be induced by both thalamic and cortical mechanisms, which may partly operate in parallel, each relying on different neuromodulatory systems [35,39].

Thalamocortical interactions are central for gating the sensory transmission by rhythmic activity. For example, thalamic neurons can actively constrain the cycle of rhythmic activity during which sensory information is relayed to the cortex [42]. Concerning the auditory system, a recent landmark study [25] used extensive sampling and optogenetic dissection to show that corticofugal (i.e., thalamic projecting) neurons in layer 6 can mediate the phase reset of delta band-activity. Thereby, these cells change the selectivity of sound-encoding neurons, facilitating either stimulus detection or discrimination depending on the relative timing of the stimulus to the delta cycle.

These results point to a key role of thalamic structures in shaping entrainment and stimulus selection. Hence, thalamic regions may help to guide the exogenously or endogenously driven attentional entrainment [39], in line with proposals made by electrophysiologists for a long time [43]. Auditory entrainment induced by a rhythmic stimulus is not confined to the cortex but is seen in the auditory

Box 1. What Qualifies as Neural Entrainment and When Does It Matter?

One can provide different definitions of entrainment. We here consider a strict definition, motivated from physical principles, and a broader definition amenable to the technical constraints of most electrophysiological or imaging studies.

Entrainment in the Narrow Sense

For parsimony, the more common term ‘synchronization’ could be used instead. According to [80], we define synchronization as the ‘adjustment of rhythms of oscillators’ (p. 8). Neural entrainment thus requires one or more self-sustained oscillating processes to start with or, in a looser interpretation, an oscillatory process that is induced by an external source of energy. Whether such endogenous oscillatory processes are actually present in the brain often remains unclear in studies that call on ‘entrainment’.

It poses logical and technical challenges to confirm empirically the presence of the following requirements. (i) The neural oscillation is endogenous; that is, it can exist in the absence of an exogenous driving rhythm, either spontaneously or persistently after the offset of a stimulus. (ii) The oscillation is resilient to perturbation; that is, in the presence of exogenous rhythms the oscillator adapts its frequency within certain bounds. Importantly, this is necessary to be able to speak of entrainment to a signal with varying temporal regularities such as speech. It shows the tendency to return to its ‘eigen-frequency’ and its endogenous phase cycle on stimulus offset (Figure 1A). (iii) Related, and distinguishing entrainment from the weaker assumption of resonance, the phase of the endogenous oscillation will ‘linger’ for a number of cycles before returning to its endogenous phase.

In practice, few studies have tested these criteria. Notable exceptions have exploited the concept of the ‘Arnold tongue’ (e.g., [81,82]; Figure 1B). An Arnold tongue charts the observed amplitude of entrainment as a function of how close the entraining rhythm is to the oscillator’s eigenfrequency and of how strong the entraining signals is itself. Thus, an important manipulation that is only rarely employed is to vary the strength of an entraining physical feature (e.g., in the auditory domain the signal-to-noise ratio or the loudness of the entraining sound). Also, many studies on natural speech assume that the relevant acoustic signal is sufficiently regular to fit with such a definition of entrainment but fail to quantify the specific sensory regularities, while other studies either impose a strict regularity on the stimulus [83] or systematically manipulate these [61].

Entrainment in the Broad Sense

Often it is only demonstrated that the exogenous signal and the endogenous activity are temporally aligned (Figure 1C) rather than varying independently. Such temporal (e.g., phase) alignment may or may not be an instantiation of true entrainment in a strict sense. In many instances, the more likely and more parsimonious scenario is that a series of sensory inputs evokes a cascade of stereotyped, impulse-like evoked responses (left-bottom panel; but see [84]). This impulse response can be estimated with methods from system identification. Dedicated toolboxes are available that have greatly facilitated the study of entrainment and have popularized the estimation of (spectro)temporal response functions [85,86].

This alignment of brain and stimulus is often called ‘neural tracking’ irrespective of the generating mechanism. As Figure 1C shows, a spectral analysis of activity alone can be misleading in that evoked responses themselves can give rise to spectral peaks, here around 5 Hz, while no true oscillatory processes might be involved (as evidenced by the overlaid spectrum of the raw EEG signal). If authors insist on using the term ‘(neural) entrainment’ despite an unclear underlying process, we recommend using the term ‘entrainment in the broad sense’.

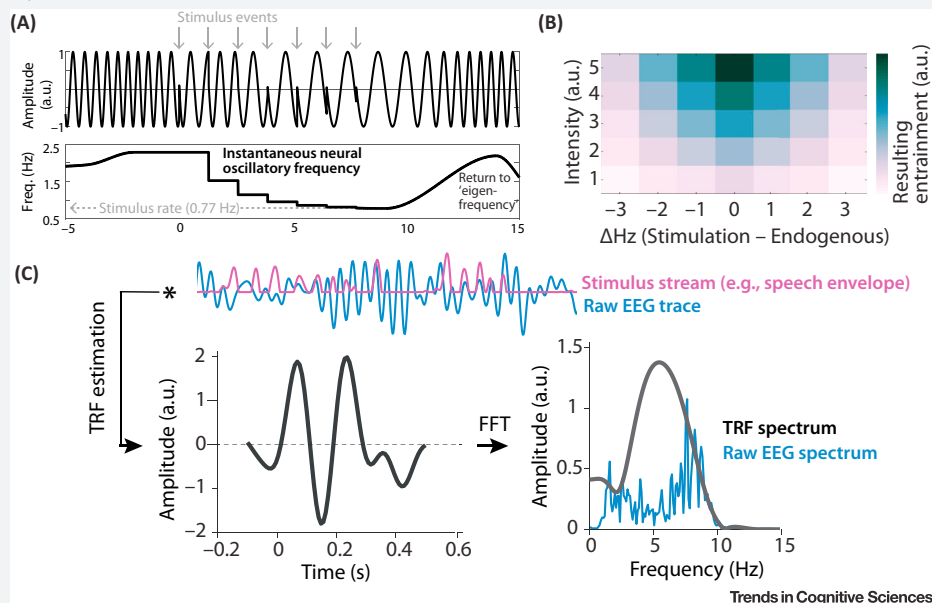


Figure 1. Entrainment in the Narrow Sense versus Entrainment in the Broad Sense.

(A) A neural oscillator exhibits an endogenous rhythm, which can be perturbed by regularly occurring stimuli (arrows). Note how the oscillator tends to quickly return to its eigenfrequency as the external drive ends. Adapted from [34]. (B) The Arnold tongue, expressing relative resilience to external perturbation with increasing distance between the endogenous rhythm (eigenfrequency) and the frequency of the external driver. The more similar an external driver is in frequency to an endogenous neural oscillator, the less energy is needed to obtain a certain level of entrainment. Adapted from [81]. (C) Neural tracking and the pitfalls of interpreting neural tracking falsely as neural entrainment in the narrow sense. As shown in this schematic, one can estimate a stimulus–response function by regressing a recorded signal against the stimulus stream (e.g., the amplitude envelope of speech). The resulting (impulse–)response function characterizes how the neural processes giving rise to the observed response ‘encode’ the stimulus. In the frequency domain (right-bottom panel), the merely evoked spectral energy prominent in the impulse response (here, 5 Hz) can be easily misinterpreted as a signature of entrainment in the narrow sense. However, a spectral profile in and by itself is not a test for an underlying oscillatory process. Here, the raw electroencephalography (EEG) frequency spectrum reveals no signature of an ongoing oscillation at 5 Hz.

thalamus and the pulvinar, with the pulvinar guiding, rather than following, auditory cortical entrainment [28]. High-level thalamic regions thus may actively modulate, or even control, entrainment and active listening. Future studies will need to reveal in more detail how subcortical regions actively participate in entrainment and active listening.

Speech Tracking as a Tool to Study Neural Representation

With these valuable mechanistic hypotheses in place we can now ask: how does entrainment pan out in a more realistic, conversational environment? Such real-life scenarios deviate considerably from the short-trial paradigms often employed in laboratory studies. They certainly require additional computations for segmenting and segregating individual sounds from a spatially, spectrally, and temporally complex mixture.

For such enriched listening scenarios, tools that quantify entrainment in a broader sense (speech tracking or, more generally, ‘neural tracking’; Box 1) have turned into a powerful approach to characterize the bidirectional relation between any type of sound stream (even irregular ones) and the brain, and allow us to study how acoustic or cognitive manipulations affect the encoding of speech. Neural tracking here refers to mathematical approaches to quantify how well the entrained activity is aligned to, or can be predicted from, specific features of the acoustic stimulus. Notably, for more translational research questions where, for example, the magnitude of a speech response (e.g., in infants) is of interest, it might be of secondary importance whether entrainment in the narrow sense is being observed; that is, whether both the stimulus and the neural response are genuinely rhythmic.

Popular tracking measures include the phase coherence between stimulus and brain activity and regression models that link brain and stimulus in either a forward (studying neural encoding) or a reverse or backward manner (reconstructing the stimulus). Inherently, these models also yield a quantification of impulse response functions. Such encoding/decoding models provide a good example for how approaches long established for single neurons (e.g., spectrotemporal receptive fields [44,45]) have recently been extended to macroscopic, multivariate neuroimaging signals.

Meaningful estimates of neural tracking can be obtained using various measures of mesoscopic brain activity, such as local field potentials derived intracranially in patients (ECoG) or noninvasive EEG or MEG activity, often band-pass filtered to contain fluctuations on a specific timescale only (e.g., [46–50]). Many studies not only focus on the phase of the neural signal but also consider the time course of the signal strength; that is, the envelope of the signal power. For example, the power envelope of gamma-band activity obtained from ECoG has proved a fruitful signal to characterize multiple aspects of speech encoding based on **power entrainment** [51] (Figure 2B).

One key insight derived using these approaches is that the neural tracking of a specific talker in higher-order auditory regions differs when the listener is attending (versus trying to ignore) this talker

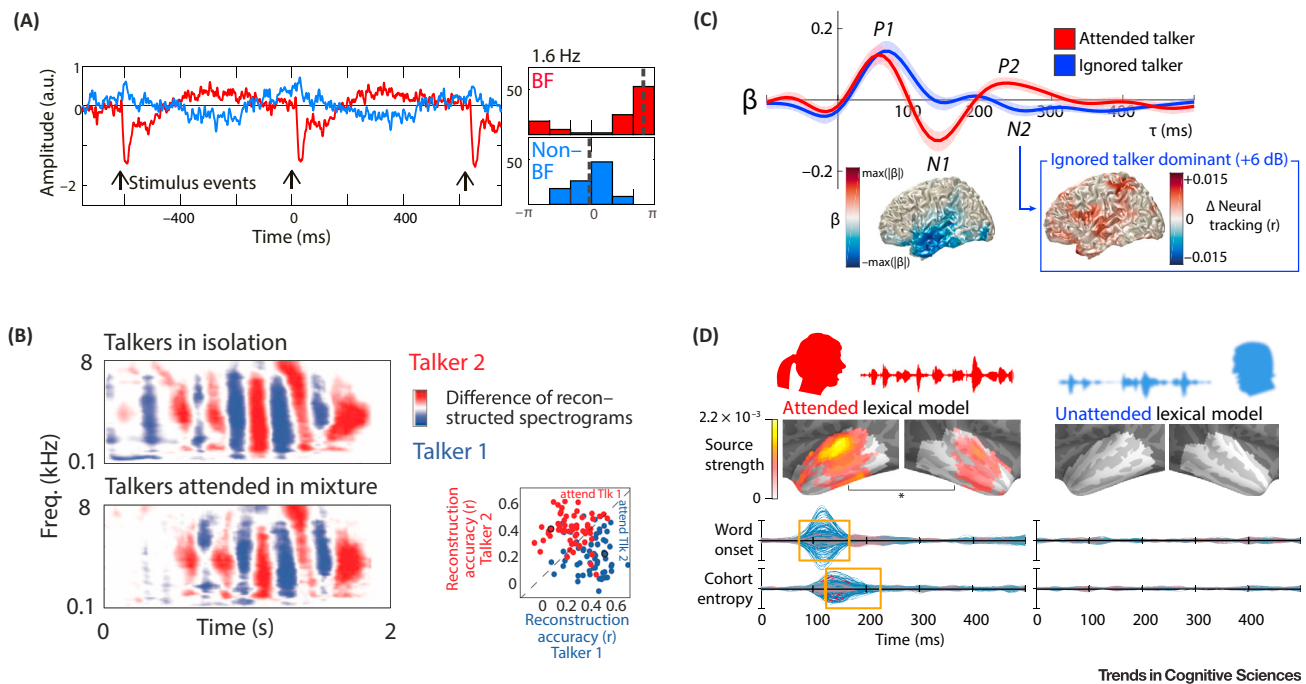


Figure 2. Attentional Filters for Auditory Sources Evident in the Neural Tracking Response.

(A) Neural entrainment of the low-frequency local field potentials in the macaque primary auditory cortex (A1) acts as a spectrotemporal filter. Attended (BF, best frequency; in red) and unattended (Non-BF; in blue) sites in A1 entrain at opposite phases to the stimulus sequence (black arrows; occurring at a rate of 1.6 Hz). This results in amplified, sharpened responses at the attended time points. Adapted, with permission, from [31]. (B) The neural tracking of an attended talker, here expressed as the reconstruction accuracy of this talker's acoustic signal (here reflected by the spectrogram) obtained from the high-frequency (gamma band) power envelopes in electrocorticography (ECoG) signals recorded over the posterior temporal lobe. When heard as a mixture of two talkers (Talker 1, blue; Talker 2, red), the attended-talker speech is represented very similarly in the entrained brain activity compared with when this talker is heard in isolation (i.e., when no attentional filtering is necessary). Inset: In correctly attended trials, the relative reconstruction accuracies for the two talkers (expressed as the correlation of actual and reconstructed speech) follow the attentional focus of the listener. Adapted, with permission, from [51]. (C) Neural tracking of an ignored talker differs from tracking of an attended talker also when quantified as a temporal response function estimated from electroencephalography (EEG) signals and source localized. The overall tracking response in the 100–200-ms time range emerges primarily from the temporal cortex. If the ignored talker is acoustically dominant (blue inset), however, ignored-talker tracking emerges in an additional response component (i.e., an enhanced N2 TRF component) with contributions from nonauditory, frontoparietal areas. This is also reflected in increased fidelity of the neural tracking of ignored speech (expressed as the correlation of predicted-from-ignored-speech versus actual EEG signal). Adapted, with permission, from [49]. (D) Selective attention to one listener not only affects the tracking of acoustic features [as in (C)] and simple 'linguistic boundaries' such as word onsets (top row), but also changes the tracking of relatively complex linguistic features such as cohort entropy (i.e., how uncertain-word identity is based on the previously heard phonemes) in neuroimaging signals, here from magnetoencephalography (MEG). The attentional modulation was graded for the acoustic features but largely suppressed the nonattended linguistic features in an all-or-none manner. Adapted, with permission, from [60].

(Figure 2B,C). The observed differences in the tracking of the speech envelope between attention conditions provide a window on how the auditory pathway differentiates two signals based on behavioral goals. In particular, this differential tracking suggests that the auditory pathway implements a 'spectrotemporal filter' (Figure 2A,C), emphasizing particular acoustic features with a specific time lag in the neural response, in line with the spatially specific entrainment of auditory cortical activity and the temporally selective gain modulation of individual neurons. This mechanism has been observed robustly across experimental approaches, including subcortical regions [52,53], in non-human primates' A1 [31], and in activity directly obtained from auditory regions in the human brain [46,51], and not least can be inferred from scalp-electrophysiological signals (e.g., [49]). It remains to be shown to what degree this apparent filtering mechanism reflects the enhancement of attended versus the suppression of irrelevant signals.

Box 2. Toward a Functionally and Brain-Region-Specific Understanding of Entrainment in Every-Day Listening

To understand how entrainment may facilitate the selection of one acoustic stream (e.g., one talker) and mediate the amplification of specific sounds (e.g., syllables), we require a better understanding of the specificity of entrainment and its modulation by attention. Studies reporting the local phase entrainment in A1 to pairs of regular tone sequences suggest that the nonattended stream does not induce entrainment in the respective tonotopic regions [31]. This apparent all-or-none effect is in stark contrast to the graded influence that attention has on neural firing rates or gamma-band activity [87] but may more directly relate to the perceptual benefit, where nonattended information can sometimes be retrieved if required but subjectively is largely eliminated. However, studies using EcoG [51,88] or neuroimaging [49] data demonstrate the feasibility of ‘reconstructing’ (i.e., classifying above chance) unattended stimuli from temporal lobe activity.

How can we reconcile such findings with the apparent all-or-none modulation of entrainment seen locally within the auditory cortex? One possibility is that neuroimaging data permit above-chance reconstruction of ignored talkers simply because of the pooling of multiple types of neural representations across larger parts of the brain. Most data were obtained from outside the brain, covering higher auditory regions in the STG (most EcoG data), or with less precise source location in the temporal lobe (MEG/EEG data). In particular, MEG/EEG studies are frequently used to study entrainment or speech tracking, but the underlying signals cannot be precisely localized to anatomically defined brain regions, less specific cell types. By contrast, direct recordings from the auditory cortex are mostly obtained from A1, while neuroanatomical and functional organizational studies have identified at least two subareas in the primate A1 and more than a dozen auditory regions between Heschl’s gyrus and the STG [89–91]. It thus remains to be seen how area specific the entrainment of rhythmic network activity is along the auditory pathways. Technical constraints on the study of individual auditory regions leave a large gap in our understanding, although recent evidence begins to paint a converging picture, in particular for high-level regions on the STG [92].

While most previous studies on auditory attention have focused on spectrotemporal cues that differentiate two competing talkers, in real life the spatial separation of two sound sources plays a central role. We would thus expect that entrainment can also be utilized to study neural implementations of spatial attentional filters. The spatial separation of two talkers improves the performance by which their acoustic streams (e.g., [48,54]) or even an exact spatial position can be recovered from a neural tracking response [55,56]. If, however, the temporal segregation of signals is compromised (e.g., by experimentally inducing a perceptually aligned sequence of two talkers), our ability to segregate two speakers based on their neural tracking response is much reduced [57]. Interestingly, in such circumstances the attentional locus is best reflected in dynamic changes of alpha power, a more heteromodal proxy of auditory attention [58,59]. Furthermore, the time course of alpha lateralization itself is then temporally modulated at the auditory stimulation rate, opening the window onto yet other forms of a temporal alignment of brain activity and the sensory environment [57].

Widening the Scope: Moving beyond Sensory Representations

Importantly, the same analytic approaches are now being used to study how higher-order acoustic, phonetic, or linguistic features and even multisensory speech signals such as lip movements are reflected in brain activity (Box 3). Addressing the critical question of attention, a recent study demonstrated that, when presented with a cocktail-party stimulus, the neural tracking of acoustic features reflects a mix of both talkers. However, when considering a lexical level of representation, the neural tracking reflected exclusively the attended talker, demonstrating that attentive listening can selectively affect specific types of acoustic or linguistic representations [60] (Figure 2D). To understand the role of entrainment as a mechanism in these studies, we will need to know whether the underlying neural circuits implement a genuine oscillatory process and to what degree the observed tracking depends on the precise regularity of the stimulus (e.g., [61]).

When combined with source localization, neural tracking holds the potential to delineate which precise speech features are represented when and where in the brain, paving the way toward a functionally specific understanding of how individual auditory regions contribute to selective listening and comprehension. In particular, this approach makes it feasible to characterize the neural encoding

of complex stimuli in notoriously difficult-to-study populations such as infants [62]. Still, many challenges remain, such as integrating the information obtained at different levels of granularity and understanding the role played by many of the anatomically defined auditory regions, which are difficult to access in most common experiments (Box 2).

What is also lacking is a clear understanding of whether and where in the brain entrainment signatures reflect the immediate perceptual outcome, rather than just some physical attribute of acoustic information. Some insights come from studies demonstrating the neural restoration of missing phonemic information in entrained **superior temporal gyrus (STG)** activity [63] or the change of syllable-induced percepts with specific phase configurations of neural activity [64]. A recent study showed that distinct brain regions entrain to different linguistic structures: in temporal regions entrainment at the time-scale of words was predictive of comprehension, while in motor regions entrainment to prosodic information did so [65]. Further work is required to better dissociate where brain activity apparently entrains to the sensory stimulus and where this entrainment is behaviorally relevant [66].

Neural Entrainment as a Tool to Manipulate the Brain

Ultimately, our goal must be to assess in compelling terms whether neural entrainment – in a narrow or a broad sense – constitutes a mechanism that causally drives speech comprehension, and perception more generally.

Toward that goal, studies have begun to enhance speech entrainment by concurrent electrical brain stimulation. These found significant, albeit small, benefits for comprehension [67–69], corroborating a mechanistically relevant role of entrainment. This line of work may also pave the way toward a translational use of entrainment via the augmentation of hearing using prosthetic devices. However, we have to remain cautious about drawing conclusions from these results. First, difficulties remain with inferring causal relations from what remain essentially observational data even when they involve brain stimulation [70–73]. Second, the considerable lack of spatioanatomical specificity of the stimulation approach remains, and the points of action in the brain remain unclear. Third, one should keep in mind the slippery slope of inferring the ‘representation’ of a given feature solely from our

Box 3. Does Neural Entrainment Delineate Linguistic Boundaries in Speech?

Conceptually, the classical view on phase entrainment predicts that entrained activity phase aligns to the regular word level or syllabic boundaries in speech [24,93]. Experimental studies only partly support this. A more parsimonious interpretation would be that neural activity entrains to discrete landmarks, such as sharp acoustic transients, which in speech often reflect consonant–vowel sequences [94,95]. However, results from clever stimulus manipulations show that entrainment can also be shaped by high-level linguistic processes and may not require such low-level acoustic regularities [30,96,97], a case in which the neural signal may be oscillatory but the stimulus not. Furthermore, by specifically crafting speech stimuli to carry syllabic, phrasal, and sentential structure at different precisely rhythmic timescales, studies were able to show that entrainment to acoustic features is automatic [83] and also emerges during sleep [98], while linguistic elements are tracked only when the stimulus is attended and comprehended [60]. Other studies are beginning to quantify the reflection of higher-order acoustic and phonetic categories [60,99–102] and semantic context [103] in the neural tracking response to natural speech. While it can pose a challenge to strictly dissociate acoustic-driven neural representations from those reflecting articulatory or perceptual categories in real-life stimuli [104,105], the converging evidence suggests that higher auditory regions along the STG encode acoustic–phonetic features, with more-posterior regions tracking these with higher temporal fidelity [92].

Bridging across sensory modalities, this picture is completed by studies demonstrating entrainment not only to acoustic but also to visual speech (lip movements) [106,107] and showing the integration of the two modalities in entrained activity [108]. While many studies have focused exclusively on either phase entrainment in low-frequency activity or power entrainment of high-gamma-band signals, both seem to carry, partly distinct, information about speech encoding [50].

Unfortunately, the analytical advances of recent years have not yet brought us much closer to a cause-and-effect-like (‘mechanistic’) understanding of how entrainment (in the narrow or broad sense) represents a necessary and/or sufficient neurobiological substrate for the human faculty of language.

Outstanding Questions

What are the microscopic mechanisms of entrainment in the auditory cortex and along the auditory pathways? Which cell types, connections, and neurotransmitters are involved? Cell-type-specific recordings and manipulations using optogenetics, possibly combined with detailed modeling studies, will be required.

How local is the entrainment along the auditory tonotopic map and how local (i.e., how feature specific) is the attentional modulation? How does entrainment along the auditory pathways look for a real-life cocktail-party scenario, where different voices may speak with different patterns of temporal regularities? Only invasive recordings with sufficient spatial resolution will be ultimately able to address this.

Does stronger entrainment always equate with better neural information representation and with better use of this information for comprehension? Studies with sufficiently large sample size and well-crafted behavioral paradigms probing distinct aspects of stimulus recognition and comprehension may provide insights.

Does the observed entrainment reflect genuine (i.e., induced) oscillations or a series of evoked responses? Studies using clever acoustic stimuli (e.g., sounds without clear temporal regularity) and the comparison of *in vivo* or *in vitro* data with computational modeling may be able to address this.

How can neuroimaging and electrophysiology converge on an understanding of entrainment? Most single-neuron data come from A1, while most EcoG results are obtained on the lateral STG, leaving a gap in the spatial coverage of the auditory cortex. Related to this, which MEG/EEG signals originate from auditory cortex and which from other temporal regions, and how can we separate these? Comparative studies bridging across measurement techniques and species are required (Box 2).

mathematical ability to decode it with above-chance accuracy from a given neural response (for reviews see, e.g., [74,75]).

Along a different route, recent work used variants of sound-driven entrainment to manipulate or improve cognitive functions such as sleep [76], time perception [77], or memory [78]. The rationale here is to use external rhythmic stimulation to shape the temporal structure of brain activity. For example, acoustic brain stimulation during slow-wave sleep was shown to entrain (i.e., boost subsequent cycles of) slow-wave activity and enhance episodic memory. Notably, in conceptual correspondence to the neural phase coding exemplified in Figure 1, this effect was specific to the acoustic perturbation occurring in phase with the ongoing activity [79].

These studies using ‘auditory entrainment’ as a technique to manipulate brain activity are a good reminder that stimulation via the natural auditory pathway, with its exquisite temporal resolution, should not be neglected amid rapidly developing techniques of electrical or optical stimulation [76]. Not least, unlike optogenetic or electrical stimulation, entrainment using the natural pathways is more likely to keep the entrained neural systems within the bounds of their normal operating dynamic range [70].

Concluding Remarks

Neural entrainment is a versatile concept for understanding how inherently fluctuating and often apparently rhythmic neural activity aligns to exogenous regularities in our environment. We have delineated the underlying processes from the vantage point of the attentive listener. When tracing entrainment back to auditory neurophysiology, it becomes clear how entrainment may facilitate the selection of task-relevant, attended information. The important delineation of true oscillatory entrainment versus entrainment in a broad sense notwithstanding, researchers can make use of the same analytical tools at different levels of description – from single neurons to whole-brain neuroimaging – to trace the reflection of increasingly sophisticated aspects of sensory and linguistic information in different parts of the brain.

With these tools in hand, we can now begin to ask the interesting questions (see Outstanding Questions) about how entrainment in general can causally shape perception, speech comprehension, and cognition in general. The scene is now set to home in on both the microscopic mechanisms and the more large-scale, brain-wide ramifications of neural entrainment for the listening brain.

Acknowledgments

Both authors are supported by the European Research Council (ERC) (ERC-CoG-2014; Grant 646657 to C.K. and Grant 646696 to J.O.). We thank three anonymous reviewers for their constructive feedback.

References

- Schroeder, C.E. and Lakatos, P. (2008) Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* 32, 9–18
- Luo, H. and Poeppel, D. (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010
- Einavoll, G.T. et al. (2013) Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nat. Rev. Neurosci.* 14, 770–785
- Buzsáki, G. et al. (2012) The origin of extracellular fields and currents – EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci.* 13, 407–420
- Panzeri, S. et al. (2015) Neural population coding: combining insights from microscopic and mass signals. *Trends Cogn. Sci.* 19, 162–172
- de Cheveigne, A. et al. (2018) Decoding the auditory brain with canonical component analysis. *Neuroimage* 172, 206–216
- Lalor, E.C. et al. (2009) Resolving precise temporal processing properties of the auditory system using continuous stimuli. *J. Neurophysiol.* 102, 349–359
- Rosen, S. (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 336, 367–373
- Nolan, F. and Jeon, H.-S. (2014) Speech rhythm: a metaphor? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369, 20130396
- Elliott, T.M. and Theunissen, F.E. (2009) The modulation transfer function for speech intelligibility. *PLoS Comput. Biol.* 5, e1000302
- Santoro, R. et al. (2014) Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLoS Comput. Biol.* 10, e1003412

12. Lakatos, P. et al. (2016) Global dynamics of selective attention and its lapses in primary auditory cortex. *Nat. Neurosci.* 19, 1707–1717
13. Henry, M.J. et al. (2014) Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proc. Natl. Acad. Sci. U. S. A.* 111, 14935–14940
14. Henry, M.J. et al. (2016) Neural microstates govern perception of auditory input without rhythmic structure. *J. Neurosci.* 36, 860–871
15. Lakatos, P. et al. (2005) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J. Neurophysiol.* 94, 1904–1911
16. Keitel, A. and Gross, J. (2016) Individual human brain areas can be identified from their characteristic spectral activation fingerprints. *PLoS Biol.* 14, e1002498
17. Haegens, S. and Zion-Golumbic, E. (2018) Rhythmic facilitation of sensory processing: a critical review. *Neurosci. Biobehav. Rev.* 86, 150–165
18. Zoefel, B. et al. (2018) The involvement of endogenous neural oscillations in the processing of rhythmic input: more than a regular repetition of evoked neural responses. *Front. Neurosci.* 12, 95
19. Kayser, C. et al. (2015) Rhythmic auditory cortex activity at multiple timescales shapes stimulus-response gain and background firing. *J. Neurosci.* 35, 7750–7762
20. Haegens, S. et al. (2011) α -Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. *Proc. Natl. Acad. Sci. U. S. A.* 108, 19377–19382
21. Chapeton, J.I. et al. (2019) Large-scale communication in the human brain is rhythmically modulated through alpha coherence. *Curr. Biol.* Published online July 25, 2019. <https://doi.org/10.1016/j.cub.2019.07.014>.
22. Kayser, C. (2009) Phase resetting as a mechanism for supramodal attentional control. *Neuron* 64, 300–302
23. Panzeri, S. et al. (2010) Sensory neural codes using multiplexed temporal scales. *Trends Neurosci.* 33, 111–120
24. Giraud, A.L. and Poeppel, D. (2012) Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517
25. Guo, W. et al. (2017) A corticothalamic circuit for dynamic switching between feature detection and discrimination. *Neuron* 95, 180–194.e5
26. Lakatos, P. et al. (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320, 110–113
27. O'Connell, M.N. et al. (2014) Layer specific sharpening of frequency tuning by selective attention in primary auditory cortex. *J. Neurosci.* 34, 16496–16508
28. Barczak, A. et al. (2018) Top-down, contextual entrainment of neuronal oscillations in the auditory thalamocortical circuit. *Proc. Natl. Acad. Sci. U. S. A.* 115, E7605–E7614
29. O'Connell, M.N. et al. (2015) Multi-scale entrainment of coupled neuronal oscillations in primary auditory cortex. *Front. Hum. Neurosci.* 9, 335
30. Zoefel, B. et al. (2017) Characterization of neural entrainment to speech with and without slow spectral energy fluctuations in laminar recordings in monkey A1. *Neuroimage* 150, 344–357
31. Lakatos, P. et al. (2013) The spectrotemporal filter mechanism of auditory selective attention. *Neuron* 77, 750–761
32. Iemi, L. and Busch, N.A. (2018) Moment-to-moment fluctuations in neuronal excitability bias subjective perception rather than strategic decision-making. *eNeuro* 5, ENEURO.0430–17.2018.
33. Schroeder, C.E. and Lakatos, P. (2009) Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* 32, 9–18
34. Obleser, J. et al. (2017) What do we talk about when we talk about rhythm? *PLoS Biol.* 15, e2002794
35. Henry, M.J. and Hermann, B. (2014) Low-frequency neural oscillations support dynamic attending in temporal context. *Timing Time Percep.* 2, 62–86
36. Zoefel, B. and VanRullen, R. (2015) Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. *J. Neurosci.* 35, 1954–1964
37. Kayser, S.J. et al. (2016) Prestimulus influences on auditory perception from sensory representations and decision processes. *Proc. Natl. Acad. Sci. U. S. A.* 113, 4842–4847
38. Vijayan, S. and Kopell, N.J. (2012) Thalamic model of awake alpha oscillations and implications for stimulus processing. *Proc. Natl. Acad. Sci. U. S. A.* 109, 18553–18558
39. Halassa, M.M. and Kastner, S. (2017) Thalamic functions in distributed cognitive control. *Nat. Neurosci.* 20, 1669–1679
40. Helfrich, R.F. et al. (2018) Neural mechanisms of sustained attention are rhythmic. *Neuron* 99, 854–865.e5
41. Herbst, S.K. and Obleser, J. (2019) Implicit temporal predictability biases slow oscillatory phase in auditory cortex and enhances pitch discrimination sensitivity. *bioRxiv*. Published online January 24, 2019. <https://doi.org/10.1101/410274>.
42. Lorincz, M.L. et al. (2009) Temporal framing of thalamic relay-mode firing by phasic inhibition during the alpha rhythm. *Neuron* 63, 683–696
43. Steriade, M. and Llinas, R.R. (1988) The functional states of the thalamus and the associated neuronal interplay. *Physiol. Rev.* 68, 649–742
44. de Boer, R. and Kuyper, P. (1968) Triggered correlation. *IEEE Trans. Biomed. Eng.* 15, 169–179
45. Theunissen, F.E. et al. (2001) Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network* 12, 289–316
46. Golumbic, E.M.Z. et al. (2013) Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. *Neuron* 77, 980–991
47. Ding, N. and Simon, J.Z. (2012) Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. U. S. A.* 109, 11854–11859
48. O'Sullivan, J.A. et al. (2015) Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cereb. Cortex* 25, 1697–1706
49. Fiedler, L. et al. (2019) Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. *Neuroimage* 186, 33–42
50. Gross, J. et al. (2013) Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol.* 11, e1001752
51. Mesgarani, N. and Chang, E.F. (2012) Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485, 233–236
52. Forte, A.E. et al. (2017) The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. *eLife* 6, 35
53. Maddox, R.K. and Lee, A.K.C. (2018) Auditory brainstem responses to continuous natural speech

- in human listeners. *eNeuro* 5, ENEURO.0441-17.2018.
54. Ding, N. and Simon, J.Z. (2012) Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J. Neurophysiol* 107, 78–89.
 55. Teoh, E.S. and Lalor, E.C. (2019) EEG decoding of the target speaker in a cocktail party scenario: considerations regarding dynamic switching of talker location. *J. Neural Eng.* 16, 036017
 56. Bednar, A. and Lalor, E.C. (2018) Neural tracking of auditory motion is reflected by delta phase and alpha power of EEG. *Neuroimage* 181, 683–691
 57. Wöstmann, M. et al. (2016) Spatiotemporal dynamics of auditory attention synchronize with speech. *Proc. Natl. Acad. Sci. U. S. A.* 113, 3873–3878
 58. Wöstmann, M. et al. (2017) The human neural alpha response to speech is a proxy of attentional control. *Cereb. Cortex* 27, 3307–3317
 59. Lim, S.J. et al. (2015) Selective attention to auditory memory neurally enhances perceptual precision. *J. Neurosci.* 35, 16094–16104
 60. Brodbeck, C. et al. (2018) Rapid transformation from auditory to linguistic representations of continuous speech. *Curr. Biol.* 28, 3976–3983.e5
 61. Kayser, S.J. et al. (2015) Irregular speech rate dissociates auditory cortical entrainment, evoked responses, and frontal alpha. *J. Neurosci.* 35, 14691–14701
 62. Jessen, S. et al. (2019) Quantifying the individual auditory and visual brain response in 7-month-old infants watching a brief cartoon movie. *Neuroimage* 202, 116060
 63. Leonard, M.K. et al. (2016) Perceptual restoration of masked speech in human cortex. *Nat. Commun.* 7, 13619
 64. ten Oever, S. and Sack, A.T. (2015) Oscillatory phase shapes syllable perception. *Proc. Natl. Acad. Sci. U. S. A.* 112, 15833–15837
 65. Keitel, A. et al. (2018) Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS Biol.* 16, e2004473
 66. Panzeri, S. et al. (2017) Cracking the neural code for sensory perception by combining statistics, intervention, and behavior. *Neuron* 93, 491–507
 67. Riecke, L. et al. (2018) Neural entrainment to speech modulates speech intelligibility. *Curr. Biol.* 28, 161–169.e5
 68. Wilsch, A. et al. (2018) Transcranial alternating current stimulation with speech envelopes modulates speech comprehension. *Neuroimage* 172, 766–774
 69. Zoefel, B. et al. (2018) Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. *Curr. Biol.* 28, 401–408.e5
 70. Jazayeri, M. and Afraz, A. (2017) Navigating the neural space in search of the neural code. *Neuron* 93, 1003–1014
 71. Krakauer, J.W. et al. (2017) Neuroscience needs behavior: correcting a reductionist bias. *Neuron* 93, 480–490
 72. Jonas, E. and Kording, K. (2017) Could a neuroscientist understand a microprocessor? *PLoS Comput. Biol.* 13, e1005268
 73. Marinescu, I.E. et al. (2018) Quasi-experimental causality in neuroscience and behavioural research. *Nat. Hum. Behav.* 2, 891–898
 74. de-Wit, L. et al. (2016) Is neuroimaging measuring information in the brain? *Psychon. Bull. Rev.* 23, 1415–1428
 75. Kriegeskorte, N. and Douglas, P.K. (2018) Cognitive computational neuroscience. *Nat. Neurosci.* 21, 1148–1160
 76. Lafon, B. et al. (2017) Low frequency transcranial electrical stimulation does not entrain sleep rhythms measured by human intracranial recordings. *Nat. Commun.* 8, 1199
 77. Morillon, B. et al. (2016) Temporal prediction in lieu of periodic stimulation. *J. Neurosci.* 36, 2342–2347
 78. Hanslmayr, S. et al. (2019) Modulating human memory via entrainment of brain oscillations. *Trends Neurosci.* 42, 485–499
 79. Ngo, H.-V.V. et al. (2013) Auditory closed-loop stimulation of the sleep slow oscillation enhances memory. *Neuron* 78, 545–553
 80. Pikovsky, A. et al. (2005) *Synchronization: A Universal Concept in Nonlinear Sciences* (Cambridge University Press)
 81. Notbohm, A. et al. (2016) Modification of brain oscillations via rhythmic light stimulation provides evidence for entrainment but not for superposition of event-related responses. *Front. Hum. Neurosci.* 10, 1868–1812.
 82. Fröhlich, F. and McCormick, D.A. (2010) Endogenous electric fields may guide neocortical network activity. *Neuron* 67, 129–143
 83. Ding, N. et al. (2016) Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* 19, 158–164
 84. Doelling, K.B. et al. (2019) An oscillator model better predicts cortical entrainment to music. *Proc. Natl. Acad. Sci. U. S. A.* 116, 10113–10121
 85. David, S.V. et al. (2015) Estimating sparse spectro-temporal receptive fields with natural stimuli. *Network* 18, 191–212
 86. Crosse, M.J. et al. (2016) The Multivariate Temporal Response Function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. *Front. Hum. Neurosci.* 10, 604
 87. Fritz, J.B. et al. (2007) Auditory attention – focusing the searchlight on sound. *Curr. Opin. Neurobiol.* 17, 437–455
 88. Zion Golumbic, E. et al. (2013) Visual input enhances selective speech envelope tracking in auditory cortex at a “cocktail party”. *J. Neurosci.* 33, 1417–1426
 89. Rauschecker, J.P. and Scott, S.K. (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724
 90. Jasmin, K. et al. (2019) Understanding rostral-caudal auditory cortex contributions to auditory perception. *Nat. Rev. Neurosci.* 20, 425–434
 91. Fullerton, B.C. and Pandya, D.N. (2007) Architectonic analysis of the auditory-related areas of the superior temporal region in human brain. *J. Comp. Neurol.* 504, 470–498
 92. Yi, H.G. et al. (2019) The encoding of speech sounds in the superior temporal gyrus. *Neuron* 102, 1096–1110
 93. Peelle, J.E. and Davis, M.H. (2012) Neural oscillations carry speech rhythm through to comprehension. *Front. Psychol.* 3, 320
 94. Doelling, K.B. et al. (2014) Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* 85, 761–768
 95. Oganian, Y. and Chang, E.F. (2018) A speech envelope landmark for syllable encoding in human superior temporal gyrus. *bioRxiv*. Published online August 9, 2018 <https://doi.org/10.1101/388280>.
 96. Zoefel, B. and VanRullen, R. (2016) EEG oscillations entrain their phase to high-level features of speech sound. *Neuroimage* 124, 16–23
 97. Meyer, L. et al. (2017) Linguistic bias modulates interpretation of speech via neural delta-band oscillations. *Cereb. Cortex* 27, 4293–4302

98. Makov, S. et al. (2017) Sleep disrupts high-level speech parsing despite significant basic auditory processing. *J. Neurosci.* 37, 7772–7781
99. Mesgarani, N. et al. (2014) Phonetic feature encoding in human superior temporal gyrus. *Science* 343, 1006–1010
100. Di Liberto, G.M. et al. (2015) Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Curr. Biol.* 25, 2457–2465
101. Steinschneider, M. et al. (2011) Intracranial study of speech-elicited activity on the human posterolateral superior temporal gyrus. *Cereb. Cortex* 21, 2332–2347
102. Khalighinejad, B. et al. (2017) Dynamic encoding of acoustic features in neural responses to continuous speech. *J. Neurosci.* 37, 2176–2185
103. Broderick, M.P. et al. (2018) Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. *Curr. Biol.* 28, 803–809.e3
104. Daube, C. et al. (2019) Simple acoustic features can explain phoneme-based predictions of cortical responses to speech. *Curr. Biol.* 29, 1924–1937.e9
105. Obleser, J. and Eisner, F. (2009) Pre-lexical abstraction of speech in the auditory cortex. *Trends Cogn. Sci.* 13, 14–19
106. Park, H. et al. (2015) Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Curr. Biol.* 25, 1649–1653
107. Hauswald, A. et al. (2018) A visual cortical network for deriving phonological information from intelligible lip movements. *Curr. Biol.* 28, 1453–1459.e3
108. Giordano, B.L. et al. (2017) Contributions of local speech encoding and functional connectivity to audio-visual speech perception. *eLife* 6, 1703