Alexandrou, Anna Maria; Saarinen, Timo; Kujala, Jan; Salmelin, Riitta

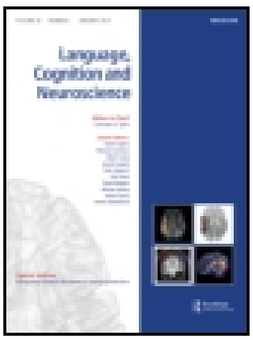
# Cortical entrainment

*Published in:*
Language, Cognition and Neuroscience

*DOI:*
10.1080/23273798.2018.1518534

Published: 02/07/2020

*Document Version*
Publisher's PDF, also known as Version of record



Routledge
Taylor & Francis Group

Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: http://www.tandfonline.com/loi/plcp21

# Cortical entrainment: what we can learn from studying naturalistic speech perception

Anna Maria Alexandrou, Timo Saarinen, Jan Kujala & Riitta Salmelin

**To cite this article:** Anna Maria Alexandrou, Timo Saarinen, Jan Kujala & Riitta Salmelin (2018): Cortical entrainment: what we can learn from studying naturalistic speech perception, Language, Cognition and Neuroscience, DOI: 10.1080/23273798.2018.1518534

**To link to this article:** https://doi.org/10.1080/23273798.2018.1518534



Published online: 13 Sep 2018.

LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2018.1518534

Routledge
Taylor & Francis Group

REGULAR ARTICLE

OPEN ACCESS Check for updates

# Cortical entrainment: what we can learn from studying naturalistic speech perception

Anna Maria Alexandrou [a,b], Timo Saarinen[a], Jan Kujala[a] and Riitta Salmelin[a,b]

[a]Department of Neuroscience and Biomedical Engineering, Aalto University, Espoo, Finland; [b]Aalto NeuroImaging, Aalto University, Espoo, Finland

**ABSTRACT**

The popular framework of cortical entrainment postulates that speech comprehension crucially depends on the continuous alignment of low-frequency cortical oscillatory activity with the amplitude envelope of perceived acoustic speech signals. The evidence for cortical entrainment mostly stems from tightly controlled experimental paradigms focusing on repeated perception of isolated sentences that feature a very constant speaking rate. However, these kinds of decontextualised and extremely regular stimuli do not reflect natural speech as we encounter it in real life. We thus advance the view that naturalistic experimental paradigms, utilising spontaneously produced speech as stimuli and suitable frequency-domain methodological tools, should be used to address an important question that remains open: whether cortical entrainment is observed during speech perception and comprehension in real-life communicative situations. In addition, we discuss how the phenomenon currently labelled as cortical entrainment might be confounded by a regular repetition of evoked responses.

**ARTICLE HISTORY**
Received 20 February 2018
Accepted 15 August 2018

**KEYWORDS**
Cortical rhythms; speech rhythm; speech perception; magnetoencephalography; cortical oscillations

## Introduction

Speech comprehension has been proposed to critically rely on cortical entrainment, that is, synchronisation between cortical signals and the envelope of the acoustic speech signal (acoustic amplitude envelope). This idea has been conceptualised through theoretical models (Giraud & Poeppel, 2012) and has found support through numerous experimental studies. Here, we offer an alternative view and suggest that presently, based on available evidence, it remains unresolved whether cortical entrainment actually takes place during speech perception. We propose that the tightly controlled experimental paradigms on which the body of empirical evidence for cortical entrainment builds on are not ideal for examining whether cortical entrainment is a neural mechanism supporting speech perception – and subsequent speech comprehension – during our every-day oral interactions. We then proceed to propose that studying perception of spontaneously produced speech would be more informative regarding the postulated role of cortical entrainment in speech perception. We first present the concept of cortical entrainment and provide a brief overview of the current findings on this topic. Then, we argue why these findings most likely do not represent actual evidence that cortical entrainment does indeed takes place. Finally, we propose that it is important to use naturalistic experimental paradigms in the quest to understand the mechanisms underlying speech perception and present our predictions regarding the cortical entrainment patterns that would likely be observed during perception of spontaneous speech. We also put forward that coherence and phase-locking values would be well-suited for examining the potential cortical entrainment patterns during naturalistic speech perception.

## Cortical entrainment in theory and in practice

The term "cortical entrainment" has become increasingly common in neuroscientific research in the recent years (for a review, see Calderone, Lakatos, Butler, & Castellanos, 2014). At its core, the framework of cortical entrainment is based on the assumption that cortical entrainment reflects the principle of attentional selection (Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008; Schroeder & Lakatos, 2009). Specifically, it has been suggested that cortical signals demonstrate oscillatory characteristics, and that the phase of these cortical oscillations is adjusted to ensure high sensitivity to relevant

**CONTACT** Anna Maria Alexandrou anna.alexandrou@aalto.fi

quasi-rhythmic or rhythmic sensory inputs. This phase adjustment has been thought to persist over time (Lakatos et al., 2005; Lakatos et al., 2008; Schroeder & Lakatos, 2009).

Speech is considered a rhythmic sensory input, even though it should be underlined that rhythm in speech is a multifaceted concept. Researchers from the field of acoustics and linguistics maintain that, instead of a regular reoccurrence of words and syllables, speech rhythm is rather a perceptual concept represented by metre and beat, concepts originating from the domain of music research. Specifically, some authors adopt the view that it is debatable whether speech rhythm exists (Nolan & Jeon, 2014), and others stress that the temporal regularities in the acoustic speech signal are not synonymous to speech rhythm (Cummins, 2012a, 2012b; Goswami & Leong, 2013).

Currently, however, a number of studies on cortical entrainment adopt a quantitative definition of speech rhythm that is based on physical characteristics of the acoustic speech signal: spectral analysis of the acoustic amplitude envelope indeed reveals that speech can demonstrate quasi-rhythmic properties (Tilsen & Arvaniti, 2013). This has sparked the notion that cortical entrainment would be a relevant concept also in perception of speech signals. During speech perception, the phase of ongoing cortical oscillations is thought to undergo an adjustment so that it matches the phase of the quasi-periodic acoustic amplitude envelope (for a review, see Peelle & Davis, 2012). Thus, cortical entrainment for speech perception is defined as a constant phase relationship (coupling) between the cortical signals and the acoustic amplitude envelope. An alternative definition of cortical entrainment is the observation of a constant phase of neural response to the same speech stimulus – this definition does not therefore examine the relationship between cortical signals and the speech stimulus *per se* (see e.g. Howard & Poeppel, 2010; Luo & Poeppel, 2007; Luo, Liu, & Poeppel, 2010).

Beyond its possible role as a mechanism for sensory selection, the coupling between cortical oscillations and the acoustic amplitude envelope has been further suggested to play an important functional role in speech comprehension. This proposed facilitatory role of cortical entrainment in linguistic processing has been advanced through several theoretical models. Inspired by the time-scales of habitual word (2–4 Hz) and syllable production frequencies (4–7 Hz), these models mainly focus on delta (< 4 Hz), theta (4–8 Hz) and gamma band (35–45 Hz) oscillatory activity (Ghitza, 2013; Peelle & Davis, 2012). The phenomenological *Tempo* model (Ghitza, 2011, 2012; Ghitza & Greenberg, 2009) and the computational model subsequently proposed by Giraud and Poeppel (2012) are based on nested oscillations, specifically on the observation that the phase of theta-band cortical oscillations modulates the power of gamma-band cortical oscillations (Canolty et al., 2006). In both models, alignment of cortical oscillatory activity in the auditory cortex with the speech input leads to the division of a temporally random, stimulus-induced spike train into manageable chunks that are suitable for further processing in higher-order cortical regions. Although these two models consider the speech signal as quasi-periodic (Cummins, 2012b) and cortical oscillations as perfectly periodic signals, which is an oversimplification (Arnal & Giraud, 2012; Wang, 2010), they provide a theoretical illustration of the nature of cortical entrainment during speech perception.

Numerous studies have claimed to provide empirical evidence for the existence of cortical entrainment. These studies have either considered the perception of

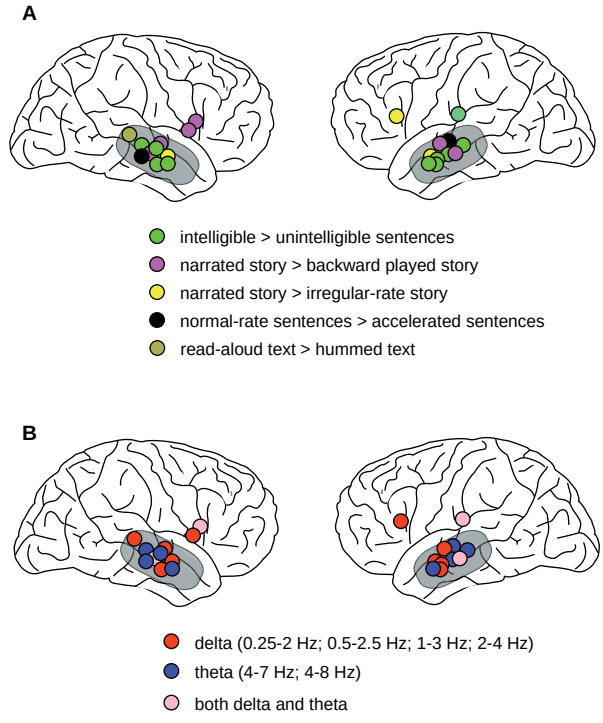



**Figure 1.** Overview of the spatiospectral topography of the findings labelled as "cortical entrainment" in the literature. The bilateral auditory cortices are highlighted in light grey colour. Each circle represents the findings of one study. ***A***, Findings categorised according to experimental paradigms: the different colours represent the experimental contrasts carried out in each study. ***B***, Findings categorised according to the frequency band in which the effects were observed. Findings from the following twelve studies are included: Ahissar et al. (2001); Bourguignon et al. (2013); Ding and Simon (2014); Doelling et al. (2014); Gross et al. (2013); Hertrich et al. (2013); Kayser et al. (2015); Keitel et al. (2017); Luo and Poeppel (2007); Mai, Minett, and Wang (2016); Park et al. (2015); Peelle et al. (2013).

isolated sentences or read-aloud texts (Figure 1A). As Figure 1A depicts, the findings of such studies mostly focus on the temporal regions bilaterally (with an emphasis on the left temporal region), with sporadic observations in higher-order regions. Spectrally, these findings considerably highlight the theta frequency band, and to a lesser degree the delta frequency band (Figure 1B).

## Studying cortical entrainment with non-naturalistic speech stimuli: is it informative?

Despite the evidence presented in Figure 1, we maintain that isolated sentences and non-naturalistic speech stimuli may not be optimal for examining the potential presence of cortical entrainment during speech perception, since they do not accurately represent instances of real-life language use. Specifically, the stimuli used in speech perception research in general, and in cortical entrainment research in particular, can be conceptualised as a continuum that ranges from completely artificial stimuli to completely natural stimuli. The isolated sentences used in many experimental paradigms devised to examine cortical entrainment (Ahissar et al., 2001; Ding, Melloni, Zhang, Tian, & Poeppel, 2016; Kösem, Basirat, Azizi, & Wassenhove, 2016; Luo & Poeppel, 2007; Meyer, Henry, Gaston, Schmuck, & Friederici, 2017; Millman, Johnson, & Prendergast, 2015; Zoefel, Archer-Boyd, & Davis, 2018) represent the non-natural end of this continuum. Even though, in some cases, isolated sentences appear in our every-day communication, they are encountered as a part of a continuous stream of utterances or through interactions with an interlocutor. In contrast, the sentences used in such experimental paradigms are semantically completely unrelated to each other and are repeated numerous times, a quite improbable occurrence in real-life communication. Often these sentences are degraded by using noise-vocoding, a technique that consists of parametrically modulating the number of frequency channels, and consequently the spectrotemporal detail in the speech signal: this renders a given speech stimulus partly or completely unintelligible (Peelle, Gross, & Davis, 2013; Scott, Rosen, Lang, & Wise, 2006; Smith, Delgutte, & Oxenham, 2002).

Continuous speech that is produced by trained speakers and is either read aloud (Di Liberto, O'Sullivan, & Lalor, 2015; Ding & Simon, 2012; Ding, Chatterjee, & Simon, 2014; Gross et al., 2013; Kayser, Ince, Gross, & Kayser, 2015; Keitel, Ince, Gross, & Kayser, 2017; Park, Ince, Schyns, Thut, & Gross, 2015) or rehearsed (Giordano et al., 2017; Zion-Golumbic et al., 2013) is further ahead in the artificial – natural continuum than isolated sentences. Although continuous stimuli do represent a step towards more naturalistic set-ups, read or rehearsed speech is quite infrequently encountered in real-life communicative situations: for instance, theatre performances, audio books and news readings represent quite small chunks of the speech we listen to in every-day life. Indeed, behavioural results have shown that read-aloud or rehearsed speech is more intelligible to listeners than speech used in every-day listening situations (Payton, Uchanski, & Braida, 1994; Uchanski, Choi, Braida, Reed, & Durlach, 1996), possibly because reading rate has been found to be invariably lower than speaking rate (Crystal & House, 1982; Hirose & Kawanami, 2002; Picheny, Durlach, & Braida, 1985). Furthermore, it is worth noting that listeners can identify spontaneous narratives from rehearsed ones on the basis of their linguistic content (Chawla & Krauss, 1994) and prosodic structure (Blaauw, 1994). Tree (1995) has examined the role of false starts and repetitions that characterise spontaneous speech in speech comprehension. As Tree (1995) underlines, the speech of professional speakers is fluent and clear, since it is especially designed for efficient information transmission to the listeners. However, when spontaneously producing speech, speakers speak and plan simultaneously (Ferreira & Swets, 2002). This is why spontaneous speech is characterised by significant disfluencies, in the forms of interruptions, repetitions, filler words and revisions, as well as a larger dynamic range of prosodic variations compared to read-aloud speech (Hirose & Kawanami, 2002). Behavioural work has further demonstrated that these characteristics of spontaneously produced speech shape subsequent speech comprehension (Brennan & Schober, 2001; Tree, 1995, 2001). Therefore, one may suggest that the rehearsed speech often used as "natural speech" represents only one, quite marginal and scarce case of speech perception, which is remarkably "simplified" compared to spontaneously produced speech.

Isolated sentences and read-aloud speech are thus quite detached from a real-life communicative context, and as such, they are not well suited as stimuli for examining if cortical entrainment comes into play during speech perception. However, one might claim that these non-naturalistic stimuli may nevertheless be informative about cortical entrainment patterns. Here, we wish to put forward the idea that cortical entrainment, as described by numerous theoretical models, has not yet been verified experimentally in the domain of speech research (for evidence of entrainment of endogenous cortical oscillations during other sensory tasks see Haegens & Zion Golumbic, 2018; Zoefel, ten Oever, & Sack, 2018). This kind of opinion may appear

surprising since, as Figure 1 depicts, there is a wealth of studies that refer to their findings as "cortical entrainment". Yet, in accordance with the original description of this neural phenomenon (e.g. Schroeder & Lakatos, 2009) we suggest that, to appropriately call one's findings as cortical entrainment, there should be a coupling between *ongoing* cortical oscillatory activity and the acoustic amplitude envelope. Moreover, crucially, this coupling should *persist over time*. The simplistic isolated linguistic stimuli employed in numerous experimental paradigms (see Figure 1A) that have been aimed at examining cortical entrainment, as well as the data analysis methods used in the context of these paradigms, do not, in our opinion, allow for such a conclusion to be made. These studies have not examined the existence of cortical oscillatory activity at the time of sensory stimulation, and they have mostly used sentences spoken at very constant rates, featuring clearer rhythmic patterning than spontaneous speech. Notably, the stimuli in these studies generally feature acoustic edges that are more prominent than in real-life speech, eliciting isolated evoked responses.

Indeed, isolated sentences and read-aloud texts are acoustically different from spontaneously produced speech. These acoustic differences are crucial when examining cortical entrainment. In particular, these stimuli differ in terms of speech rhythm. Speech rhythm can be quantified through frequency-domain analysis of speech acoustic signals; peaks in the resulting power spectrum illustrate the presence of rhythmic patterning in the signals (Alexandrou, Saarinen, Kujala, & Salmelin, 2016; Chandrasekaran, Trubanova, Stillittano, Caplier, & Ghazanfar, 2009). Spontaneously produced speech tends to feature quasi-rhythmic temporal patterning. The overall syllable frequency of speech produced at the normal/habitual speaking rate varies across individuals (3.5–6.2 Hz; Alexandrou et al., 2016). In addition to this, there are local speaking rate variations that occur within the same utterance of the same speaker (Miller, Grosjean, & Lomanto, 1984); there is indeed evidence that the speaking rate varies across time segments as short as 5 s (Alexandrou, Saarinen, Kujala, & Salmelin, 2018). The rhythmic patterning in speech is reflected through the power spectrum of the acoustic amplitude envelope (acoustic power spectrum). The power spectrum of spontaneously produced speech, computed according to the procedure described in Alexandrou et al. (2016) (Figure 2A, top), is rather flat and is characterised by a conspicuous *1/f* trend (Alexandrou et al., 2016; Chandrasekaran et al., 2009; Ruspantini et al., 2012). Thus, the present dataset may be even suggested to demonstrate a non-rhythmic pattern, compared with some other spontaneous speech datasets (Chandrasekaran et al., 2009). A similar pattern is observed for the modulation spectrum of the speech signal (Figure 2B): while this method of spectral estimation reveals the typical 4-Hz peak that reflects the frequency of firing of the auditory nerve or higher-level subcortical auditory nuclei in response to an incoming speech stimulus (Ding et al., 2017), it does not otherwise suggest the presence of any salient rhythmicity in the actual acoustic signal, as evidenced by the quite flat spectral pattern (Figure 2B).

In contrast, isolated sentences feature a salient rhythmic pattern since the words in the sentence are spoken at very constant intervals. In addition, these sentences consist of only a few, fairly short words, and usually the same number of words is used across all sentences in an experimental paradigm (Ahissar et al., 2001; Ding et al., 2016; Nourski et al., 2009). For this kind of short sentences, it is easier to maintain a steady speaking rate than for natural connected speech in which utterances are longer and sentences are syntactically more complex, with considerably more variation in word length (Clark & Wasow, 1998; Cummins & Port, 1998; Ferreira, 1991; Yaruss, 1999). The salient temporal patterning of sentence stimuli is manifested as prominent peaks in the acoustic power spectrum (~3–4 Hz; Figure 2A). The peaks observed in the acoustic power spectra are also readily distinguishable in the modulation spectra, with conspicuous peaks of normalised amplitude at lower frequencies (Figure 2B). The modulation spectra for sentences also demonstrate very sharp peaks centred at 1 Hz and 2 Hz, in agreement with Figure 2A (middle). Read-aloud texts also display fairly clear temporal patterning: there is evidence to suggest that, compared to speaking rate, reading rate remains quite stable (Uchanski et al., 1996). Thus, read-aloud speech, especially when produced by professional speakers or actors, has been proposed to feature quite a salient temporal structure (Uchanski et al., 1996). Indeed, even though read-aloud speech features quasi-rhythmic elements, power spectral analysis reveals a clearly distinguishable rhythmic structure, especially for some readers (Figure 2A, bottom). This is also evident in the modulation spectra (Figure 2B, bottom): there are observable peaks – albeit less salient than for isolated sentences – for frequencies between 2 and 4 Hz, in agreement with the spectra displayed in Figure 2A (bottom).

Moreover, read-aloud speech that is commonly used in studies examining continuous stimuli (see e.g. Gross et al., 2013; Kayser et al., 2015; Keitel et al., 2017) is characterised by prominent edges in its acoustic amplitude envelope. Acoustic edges represent large, salient increases in the acoustic envelope amplitude. Trained

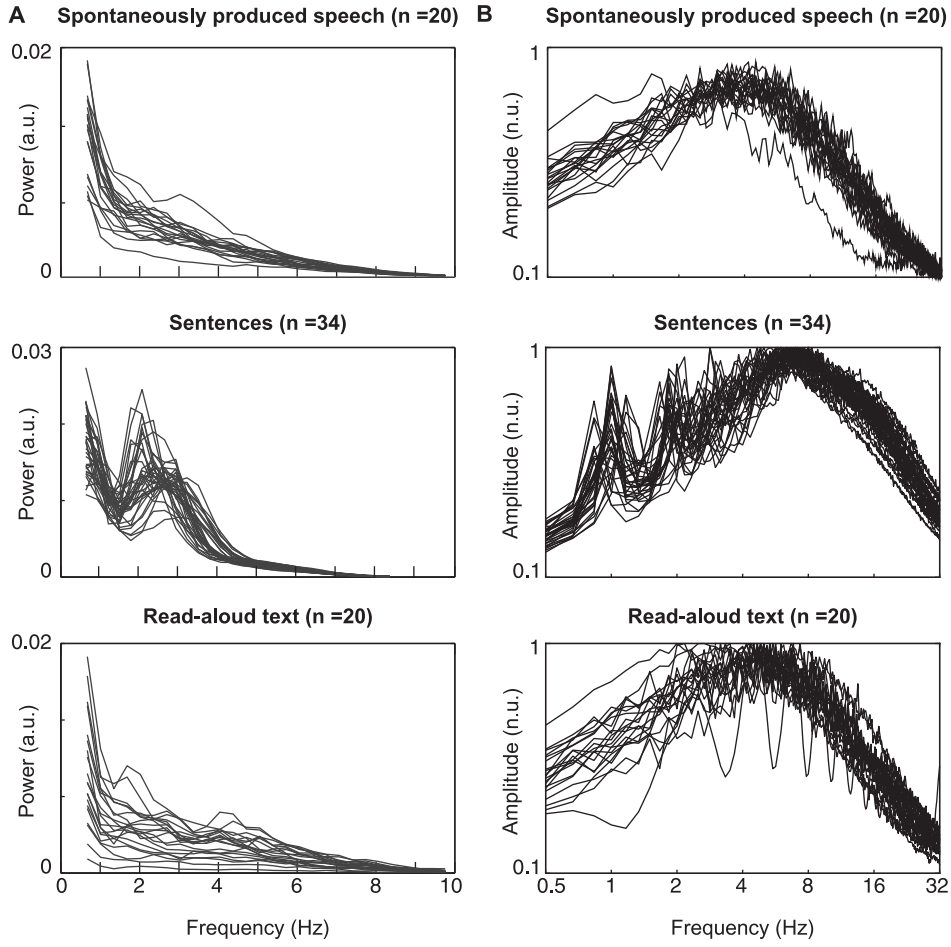



**Figure 2.** **A**, Power spectra of the acoustic amplitude envelope of different types of speech stimuli, estimated using the procedure described in Tilsen and Arvaniti (2013) and Alexandrou et al. (2016). Power (in arbitrary units; y-axis) is plotted against frequency (in Hz; x-axis). **B**, Modulation spectra of the acoustic amplitude envelope of different types of speech stimuli, computed using the procedure described in (Ding et al., 2017). Amplitude (in normalised units; y-axis) is plotted against frequency (in Hz, logarithmic scale; x-axis). **Top**: Spontaneously produced speech (4 min). Data are overlaid for the 20 study participants in Alexandrou et al. (2017). **Middle**: 1000 spoken 6-word sentences (mean duration $1.8 \pm 0.17$ s) taken from the GRID corpus (see Cooke, Barker, Cunningham, & Shao, 2006). Data are overlaid for all 34 speakers available in the corpus. **Bottom**: Read-aloud text (2 min). Data are overlaid for the 20 study participants in Alexandrou et al. (2017); same participants as for the spontaneously produced speech. Notice the qualitative difference in peaks in the power spectra between the three different types of speech stimuli.

speakers narrating stories based on a written text adopt a speaking style featuring an ample amount of prosodic gestures (including large variations in fundamental frequency and pause insertion) (Nakajima & Allen, 1993). In addition to notable prosodic patterning, read-aloud speech has a linguistically very salient structure: the linguistic content is segmented onto clearly identifiable sentences, phrases and paragraphs (Nakajima & Allen, 1993). This prosodic clarity is further amplified by the fact that read-aloud speech does not feature notable co-articulation effects, in contrast to spontaneously produced speech (Finke & Rogina, 1997). These inherent characteristics of read-aloud speech are acoustically manifested as prominent acoustic edges that are observed as a notable increase in the mean acoustic envelope amplitude as a function of time. In contrast, in spontaneously produced speech, prosodic features are less salient since complete sentences are quite rare (e.g. an utterance can be interrupted by other speaker (s)), co-articulation effects are quite prominent and the narrative style of speaking that accentuates prosodic structure is absent. To illustrate this point, the acoustic edges from acoustic signals recorded when a group of 20 individuals read aloud a text (Figure 3, black colour), and spontaneously produced speech (Figure 3, grey colour), were identified according to the procedure used by Gross et al. (2013). The edges represent points in the audio signal where a low-amplitude baseline window is followed by a sharp rise in the signal amplitude together with a sustained period where the audio signal remains at an elevated level. A qualitative comparison reveals that, indeed, the edges in spontaneously

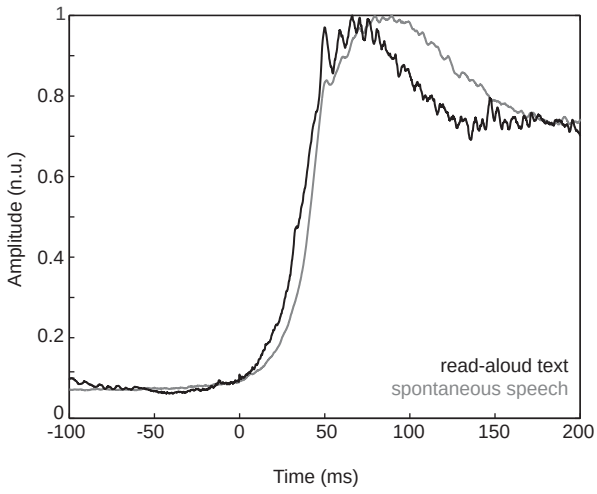



**Figure 3.** Acoustic edges' prominence in the amplitude envelope of speech signals. Mean acoustic envelope amplitude, normalised by the group-level mean amplitude (in normalised units; y-axis), is plotted against time (in ms; x-axis; data time-locked to edge onset). Data are displayed from 15 out of the 20 study participants in Alexandrou et al. (2017) for reading aloud a text (black colour) and spontaneously producing speech (grey colour) at the normal speaking rate. Notice the qualitative difference in the prominence of acoustic edges, illustrated by a shorter rise time of the amplitude of the acoustic envelope for read-aloud text than for spontaneously produced speech.

produced speech (Figure 3, grey colour) appear less sharp. Even though the difference is quite subtle, it can be observed that the amplitude of the acoustic envelope both increases (see the time-window between 0 and 50 ms) and decreases (see the time-window between 50 and 100 ms) over a longer period of time than in read-aloud texts (Figure 3, black colour) (see also Swerts, Strangert, & Heldner, 1996). Isolated sentences also feature quite prominent acoustic edges. This is because, firstly, they represent individual experimental trials, and thus emerge from complete silence; secondly, the manner in which they are spoken results in clear boundaries between consecutive words, especially for sentences that have been synthesised by concatenating individual words and inserting pauses between them (Ding et al., 2016). Consequently, these stimuli also feature large, abrupt increases in the amplitude of the acoustic envelope.

Hence, the speech stimuli typically used in the experimental paradigms that aim to examine cortical entrainment are characterised by, firstly, a quite regular rhythmic structure, and therefore a very narrow frequency content that notably emphasises a certain frequency range (especially in the case of isolated sentences), and secondly, prominent acoustic edges. We propose that, due to these acoustic characteristics, the perception of this kind of stimuli represents a form of direct rhythmic stimulation of the auditory cortices (perception of spontaneously produced speech most likely also results in this direct auditory stimulation but presumably to a lesser degree, since spontaneously produced speech is not as rhythmic). This consequently favours the observation of repeated, isolated neural responses. Hence, we maintain that, for the case of speech perception, the neural phenomenon that is widely labelled as "cortical entrainment" is, in all likelihood, a sequence of stimulus-driven neural responses. These neural responses may presumably be mostly regularly reoccurring auditory evoked responses, without, however, entirely excluding the possibility of induced responses. In support for our view, most studies (see legend of Figure 1 for a list of studies that have reported such emphasis; also see Ding et al., 2016; Hertrich, Dietrich, Trouvain, Moos, & Ackermann, 2012) invariably highlight the temporal regions, the main locus of evoked responses during speech perception (e.g. Mäkelä et al., 1993) (Figure 1A). Furthermore, although the origin of evoked responses remains uncertain, they have been proposed to be generated as a result of partial stimulus-induced phase resetting of multiple electroencephalographic processes (Makeig et al., 2002). This phase resetting has been suggested to occur as a consequence of acoustic edges in an incoming speech signal (Luo et al., 2010). The presence of acoustic edges in the experimental stimuli formed the basis for the analysis conducted by Gross et al. (2013), the results of which, although interpreted as cortical entrainment, could most likely reflect evoked responses.

We further suggest that the reason for which the findings labelled as "cortical entrainment" are mainly observed in the delta (2–4 Hz) and theta (4–8 Hz) frequency bands (Figure 1B) is a direct consequence and reflection of the frequency content of the stimulus. Especially for sentence stimuli, which feature very prominent periodicity in the form of a narrow frequency content, the neural responses occur at fixed intervals. These responses form an activity profile that, when analysed in the frequency domain, demonstrates prominent, low-frequency features that correspond to the frequency content of the stimulus (Zhou, Melloni, Poeppel, & Ding, 2016). These features are subsequently interpreted as "cortical entrainment" of cortical oscillations at a given frequency band, when they, in fact, represent evoked responses occurring at (semi-) constant rates. However, it is worth noting that this type of frequency-specific stimulation does somehow relate to cortical entrainment: for a given cortical system (in this case, the speech auditory system), there is a limited range of word and syllable frequencies (especially an upper limit) that neural signals can track in a perceived auditory

stimulus (Ahissar et al., 2001; Assaneo et al., 2016; Giraud et al., 2007; Keitel & Gross, 2016; Telkemeyer et al., 2009). This idea is also in line with previous work in which very specific stimulation frequencies have been used (Hertrich et al., 2012; Hertrich, Dietrich, & Ackermann, 2013; Howard & Poeppel, 2010; Luo et al., 2010; Luo, Boemio, Gordon, & Poeppel, 2007).

The interpretation that we present here is not new: the review articles by Ding and Simon (2014), Haegens and Zion Golumbic (2018) and Zoefel, ten Oever, et al. (2018) discuss the possibility that the evidence labelled as "cortical entrainment" is actually a superposition, or repetition of evoked potentials triggered by the edges in the speech signal (e.g. Howard & Poeppel, 2010). For instance, the effect strength is modulated as a function of the sharpness of acoustic edges (Doelling, Arnal, Ghitza, & Poeppel, 2014; Ghitza, 2011). Moreover, the effect strength is modulated in a task-specific manner via input from higher-order regions (e.g. Kayser et al., 2015; Keitel et al., 2017): this observation aligns with the reported top-down control of evoked responses (e.g. Debener, Herrmann, Kranczioch, Gembris, & Engel, 2003; Iversen, Repp, & Patel, 2009). Finally, (Zoefel & VanRullen, 2015) question, on a general level, whether the phenomenon referred to as entrainment reflects a low-level sensory response rather than a higher-level active mechanism that plays a role in linguistic processing. While Zoefel and VanRullen (2015) have chosen to label these sensory responses as "lower-level entrainment", we instead propose that, in the absence of solid evidence, the term "entrainment" should preferably be completely avoided and such findings should be rather referred to as time-locked, stimulus-driven activity.

However, we do acknowledge that some studies have sought to eliminate the possibility that the observed low-frequency oscillatory phenomena merely reflect evoked responses and have been able to identify actual phase modulations in endogenous oscillatory activity in response to sensory stimuli (Henry & Obleser, 2012; Meyer et al., 2017; Zion-Golumbic et al., 2013). In so doing, they have revealed that cortical entrainment might indeed take place (Haegens & Zion Golumbic, 2018; Zoefel, ten Oever, et al., 2018). Specifically, the potential presence of cortical entrainment (for a review, see Meyer, 2017) was evaluated outside the domain of speech research by delivering rhythmic stimulation and observing the persistence of oscillatory activity at that same frequency after stimulation had ceased (Dilley & Pitt, 2010; Hickok, Farahbod, & Saberi, 2015; Neuling, Rach, Wagner, Wolters, & Herrmann, 2012). Furthermore, it appears that cortical oscillations entrain to sensory stimuli even when rhythmic acoustic cues are less prominent (Calderone et al., 2014; Mathewson, Gratton, Fabiani, Beck, & Ro, 2009) or absent (Henry & Obleser, 2012). Recent evidence further suggests that this phenomenon could even be completely independent from exogenous rhythmic cues (Kösem et al., 2016; Meyer et al., 2017). Yet, it is worth noting that it is difficult to distinguish endogenous oscillatory entrainment from the regular reoccurrence of evoked responses (Haegens & Zion Golumbic, 2018; Zoefel, ten Oever, et al., 2018), and that such a practice has not been commonplace in cortical entrainment experimental paradigms in the domain of speech research.

At this point, we invite the reader to re-consider the definitions of cortical entrainment presented at the beginning of this article. We propose that the main definition of cortical entrainment advanced in Peelle and Davis (2012), that is, the existence of a systematic and consistent phase relationship between ongoing cortical oscillatory activity and natural connected speech that persists over time, and specifically over the entire duration of the stimulus, has not been solidly demonstrated with the existing experimental paradigms. As such, the proclaimed link between the empirical observations and theoretical models is not based on concrete and unequivocal evidence that cortical entrainment, as postulated in Ghitza and Greenberg (2009), Ghitza (2011, 2012) and Giraud and Poeppel (2012) actually takes place. Further research could benefit from a more thorough investigation of the origin of the experimentally observed effects.

We further propose that the alternative, less habitual definition of cortical entrainment mentioned by Peelle and Davis (2012) and presented earlier in this opinion piece (see section "Cortical entrainment in theory and in practice"), namely, the observation of a consistent phase of neural response to the same stimulus, time-locked to the input (e.g. Luo & Poeppel, 2007) is, instead, what is widely observed and reported in the literature (but see Meyer, 2017; Meyer et al., 2017). This second definition of cortical entrainment refers, in reality, to evoked responses and has been confounded with the first – in our opinion more fitting – definition of entrainment.

## What can naturalistic experimental paradigms teach us about cortical entrainment?

Given the current predominant trend in neuroscientific research to explain and elucidate the neural correlates of speech perception and subsequent speech comprehension through cortical entrainment models, we feel that it is crucial to make the shift towards ecologically valid experimental paradigms. For the reasons described

in the previous section, evidence from studies employing isolated sentences and continuous read-aloud texts cannot be readily extrapolated to perception of spontaneously produced speech during real-life communicative situations. We thus maintain that, unless naturalistic speech stimuli are used, it remains unclear whether cortical entrainment is a relevant neural mechanism in real-life speech processing. At this point, we would like to underline that we do not promote the use of naturalistic speech stimuli as a method of differentiating evoked responses from actual cortical entrainment, as evoked responses may also be recorded during perception of natural speech, but as an essential means of exploring whether cortical entrainment is indeed a pre-requisite for successful speech comprehension.

We propose that an optimal experimental paradigm for studying whether cortical entrainment takes place during naturalistic listening conditions (once it has been verified that the cortical activity preceding the stimulus indeed exhibits oscillatory characteristics) would consist of perception of spontaneously produced speech that demonstrates naturalistic experimental manipulations. By "naturalistic experimental manipulations", we refer to experimental conditions that capture different aspects and contexts of use of real-life speech. For instance, it has been suggested that cortical entrainment should be modulated by the amount of linguistic content in an utterance (Doelling et al., 2014; Ghitza, 2011), as well as its relevance for the listener (Lakatos et al., 2005). In order to examine this proposal, the effect of linguistic content can be probed by naturally modulating the linguistic content of speech (e.g. a university-level lecture consisting of difficult technical terms on a topic largely unfamiliar to the listener *vs.* a simple, spontaneously produced narrative of a routine real-life event). Real-life speech presents us with a variety and multitude of different linguistic contents, which we feel should be exploited and used to our advantage in experimental set-ups. However, we also wish to note that non-naturalistic experimental stimuli could potentially be used together with naturalistic speech stimuli in designing appropriate experimental contrasts in the aim to dissociate evoked responses from actual cortical entrainment. For instance, spontaneously produced speech could be contrasted to speech with the same linguistic content, which has been transcribed and then read aloud based on this written transcription. This would keep the linguistic content stable, but acoustically, the two stimuli would be markedly different (see Figures 2 and 3). However, irrespective of the naturalistic experimental paradigm employed, it is crucial to seek to disentangle evoked responses from actual cortical entrainment.

The aim of using natural connected speech as stimuli would be to elucidate the potential presence of cortical entrainment and whether it indeed plays a role in speech comprehension, instead of merely being a sensory mechanism. However, while using naturalistic stimuli might result in reduced occurrence and regularity of evoked responses compared to stimuli with more prominent acoustic structure and sharper acoustic edges, the overall weaker rhythmicity of the naturalistic speech signal would likely be accompanied by weaker cortical entrainment, as well. Indeed, we would like to suggest that there are at least two reasons why cortical entrainment might not be readily observable using natural speech stimuli. At the core of both lines of reasoning is the variable and fairly quasi-rhythmic pattern (or in some cases, non-rhythmic pattern; see Figure 2A and B, top) of the acoustic amplitude envelope of spontaneously produced speech. Indeed, the notion of cortical entrainment advanced by e.g. Lakatos et al. (2008) directly links cortical entrainment with the structure of the acoustic amplitude envelope, and specifically to the idea that cortical signals should entrain to rhythmic or quasi-rhythmic stimuli. The first alternative viewpoint is that cortical entrainment is a very weak, non-salient phenomenon in natural speech perception (also see Cummins, 2012a). The second alternative viewpoint is that cortical entrainment exists but is not readily detectable at the macroscopic level during perception of naturalistic speech.

In partial support of the first alternative viewpoint, the most notable demonstration of actual cortical entrainment has been observed in primates using simple, perfectly rhythmic stimuli in the auditory or in the audiovisual domain (Ghazanfar, Morrill, & Kayser, 2013; Lakatos et al., 2008; Steinschneider, Nourski, & Fishman, 2013). Further support comes from our own recent observations: using magnetoencephalography (MEG), we examined modulations in cortical signal power during perception of spontaneous speech that had been produced at the normal (or habitual) speaking rate, as well as fast and slow rates (Alexandrou, Saarinen, Mäkelä, Kujala, & Salmelin, 2017). According to Ghitza (2012), cortical entrainment would be reflected as a shift in the power of low-frequency oscillations. However, our examination of mean MEG power spectra for signals originating from the left auditory cortex, averaged across 20 subjects, suggests that there is no such shift in low-frequency oscillatory activity for perception of different-rate speech stimuli (Figure 4). Instead, modulations in gamma-band cortical activity were observed. Additionally, our use of cross-frequency coupling metrics (identical to those described in Gross et al., 2013) did not reveal any significant differences

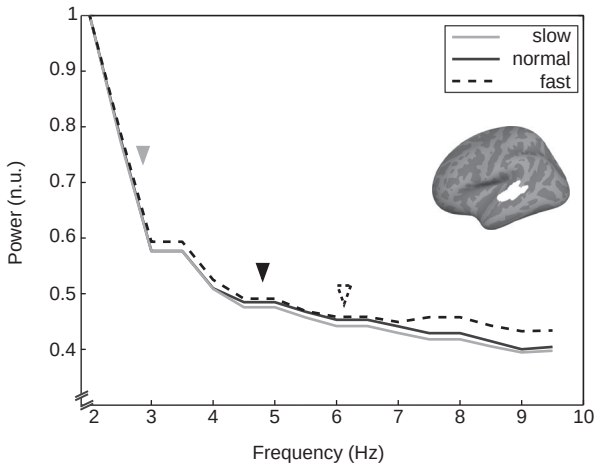



**Figure 4.** Mean power spectra, averaged across the 20 participants in Alexandrou et al. (2017) of cortical signals originating from one region of interest, the left temporal cortex (white colour). This region of interest was defined by combining the two clusters (encompassing the left superior temporal gyrus) that were found to demonstrate modulations in MEG signal power when contrasting perception of normal-rate speech to slow- and fast-rate speech; see Alexandrou et al. (2017). Signal power, normalised by its group-level maximum within the 0.5–10 Hz range (normalised units; y-axis), is plotted against frequency (in Hz; x-axis) for perception of slow-rate (grey line), normal-rate (solid black line) and fast-rate speech (dashed black line). The syllable production frequencies of the perceived speech are indicated by arrows for slow-rate (grey arrow; 2.1 Hz), normal-rate (black arrow; solid line; 4.8 Hz) and fast-rate speech (black arrow; dashed line; 6.3 Hz).

between different-rate speech stimuli. Subsequent work from our group (Alexandrou et al., 2018) has revealed that while there is coupling between cortical signals and speech stimuli, it is quite subtle and far from the very prominent activity patterns suggested by theoretical models. Thus, cortical entrainment might be a pattern of cortical activity that strongly emerges only in non-ecological circumstances and, albeit it could still play a role in subsequent stimulus processing, it is not the most predominant neurophysiological response during perception of naturalistic speech.

With regards to the second alternative viewpoint, cortical entrainment at a given frequency range should be observable for continuous speech that features a highly stable speaking rate. However, the acoustic amplitude envelope of spontaneously produced speech is inherently and unavoidably quasi-rhythmic due to naturally occurring intra-individual variations in speaking rate as a function of time. Thus, different frequencies of cortical oscillations would presumably entrain to the speech input at different moments in time. If this were the case, such frequency-variable coupling would not be easily detectable at the macroscopic level with the sensitivity afforded by the currently used methodological tools.

Nonetheless, if one were to observe cortical entrainment patterns through a naturalistic experimental approach, we predict that they would mostly involve cortical signals originating from higher-order cortical regions, instead of the currently emphasised lower-level cortical regions (that is, the auditory cortices) (also see Meyer, Sun, & Martin, 2018). This proposition is based on the observation that higher-order regions are associated with attentional selection and other attentional processes (e.g. Jensen, Kaiser, & Lachaux, 2007): this is in line with the idea of "active sensing", which is a core concept underlying cortical entrainment (Lakatos et al., 2008; Schroeder & Lakatos, 2009; Schroeder, Wilson, Radman, Scharfman, & Lakatos, 2010). Perception of natural, connected speech is thought to represent an active sensing mode that heavily relies on temporal predictions and expectations (Alexandrou et al., 2018; Engel, Fries, & Singer, 2001; Morillon & Schroeder, 2015). This view is also in line with the proposed existence of "high-level entrainment" that involves cortical activity originating from higher-order cortical regions and is modulated by predictions and other higher-level processes (Kösem et al., 2016; Zoefel & VanRullen, 2015). Indeed, studies that have used continuous stimuli have shown that also higher-level cortical regions track the incoming speech signal (see Figure 1A; also see Alexandrou et al., 2018; Borges, Giraud, Mansvelder, & Linkenkaer-Hansen, 2018; Puschmann et al., 2017). Further research is needed to determine whether the focus in cortical entrainment research should be shifted from the sensory representation areas to higher-level cortical regions.

As a consequence of the acoustic and linguistic complexities of spontaneously produced speech described above, the resulting signal-to-noise ratio is quite low. This fact poses a significant methodological challenge when addressing the question of whether cortical entrainment takes place during naturalistic speech perception experiments. In non-naturalistic experimental paradigms, signal-to-noise ratio can be improved by repeating a stimulus and by subsequently averaging the brain responses to that stimulus. However, we maintain that, in a purely naturalistic experimental paradigm, stimuli should be presented only once, to accurately simulate a real-life communicative context. This precludes averaging across repetitions as a means of enhancing the signal-to-noise ratio – although it should be noted that averaging across segments of natural speech that are similar in some respect (e.g. in speaking rate, or in the amount of linguistic content) may be possible in certain cases, depending on the nature of the stimuli and the hypothesis being tested. Thus, it is essential to develop specialised

methodological tools that are sensitive enough to the examined cortical effects. Other important methodological considerations are the number of participants and the amount of collected data.

Yet, we propose that the methodological challenges that stem from the lack of control in naturalistic experimental paradigms are counter-balanced by the ease of using real-life speech as stimuli, as it is readily available in our sensory environment. Regarding the measures of choice, one could consider a measure that quantifies the coupling between the phases of the audio envelope and the cortical signals (i.e. coherence). Specifically, coherence quantifies the relationship between two signals in the frequency domain and highlights any shared phase patterns between them, while suppressing noise and other types of random, uncorrelated activity in the signals (see Alexandrou et al., 2016). Furthermore, it tests also for non-zero-lag phase differences between the two signals. Indeed, despite its drawbacks (see e.g. Bastos & Schoffelen, 2016), coherence has proven to be a sensitive measure in examining the tracking of speech rhythm in natural connected speech (Alexandrou et al., 2018). While we here highlight coherence as a potential measure of choice, other alternatives such as phase-locking value or statistics (e.g. Lachaux, Rodriguez, Martinerie, & Varela, 1999) could also prove useful.

## Conclusion

We argue that caution is warranted when suggesting that cortical entrainment is a prerequisite for speech perception and comprehension, especially when the experimental paradigm examines perception of speech stimuli that have not been spontaneously produced, such as isolated sentences, or read-aloud texts. We propose that future studies should seek to utilise spontaneously produced speech as an optimal stimulus type to investigate the currently unclear functional role of cortical entrainment in real-life speech comprehension (see e.g. Peelle, 2018; Zoefel, Archer-Boyd, et al., 2018). We suggest that, in order for such an investigation to proceed in full force, it is also important to truly evaluate whether cortical entrainment as outlined, for instance, by Peelle and Davis (2012) actually exists or whether it is confounded by evoked responses.

## Acknowledgments

This work was financially supported by the Academy of Finland (grants #255349, #256459 and #283071 to RS, #257576 to JK), the Alfred Kordelin Foundation (grant #160143 to AA), the Emil Aaltonen Foundation (grant # 170011 N1 to AA), the Finnish Cultural Foundation (grant to TS), and the Sigrid Jusélius Foundation (grant to RS).

## Disclosure Statement

No potential conflict of interest was reported by the authors.

## Funding

This work was financially supported by the Academy of Finland [grant number #255349], [grant number #256459] and [grant number #283071] to RS, [grant number #257576] to JK, the Alfred Kordelin Foundation [grant number #160143] to AA, the Emil Aaltonen Foundation [grant number # 170011 N1] to AA, the Finnish Cultural Foundation (grant to TS), and the Sigrid Jusélius Foundation (grant to RS).

## ORCID

*Anna Maria Alexandrou* http://orcid.org/0000-0002-5342-3109

## References

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(23), 13367–13372.

Alexandrou, A. M., Saarinen, T., Kujala, J., & Salmelin, R. (2016). A multimodal spectral approach to characterize rhythm in natural speech. *The Journal of the Acoustical Society of America*, *139*(1), 215–226.

Alexandrou, A. M., Saarinen, T., Kujala, J., & Salmelin, R. (2018). Cortical tracking of global and local variations of speech rhythm during connected natural speech perception. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn_a_01295

Alexandrou, A. M., Saarinen, T., Mäkelä, S., Kujala, J., & Salmelin, R. (2017). The right hemisphere is highlighted in connected natural speech production and perception. *NeuroImage*, *152(C)*, 628–638.

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences*, *16*(7), 390–398.

Assaneo, M. F., Sitt, J., Varoquaux, G., Sigman, M., Cohen, L., & Trevisan, M. A. (2016). Exploring the anatomical encoding of voice with a mathematical model of the vocal system. *NeuroImage*, *141*, 31–39.

Bastos, A. M., & Schoffelen, J.-M. (2016). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in Systems Neuroscience*, *9*, 1–23.

Blaauw, E. (1994). The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication*, *14*(4), 359–375.

Borges, A. F. T., Giraud, A.-L., Mansvelder, H. D., & Linkenkaer-Hansen, K. (2018). Scale-free amplitude modulation of neuronal oscillations tracks comprehension of accelerated speech. *The Journal of Neuroscience*, *38*(3), 710–722.

Bourguignon, M., De Tiege, X., de Beeck, M. O., Ligot, N., Paquier, P., Van Bogaert, P., … Jousmäki, V. (2013). The pace of

prosodic phrasing couples the listener's cortex to the reader's voice. *Human Brain Mapping*, *34*(2), 314–326.
Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, *44*(2), 274–296.
Calderone, D. J., Lakatos, P., Butler, P. D., & Castellanos, F. X. (2014). Entrainment of neural oscillations as a modifiable substrate of attention. *Trends in Cognitive Sciences*, *18*(6), 300–309.
Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E., … Knight, R. T. (2006). High gamma power is phase-locked to theta oscillations in human neocortex. *Science*, *313*(5793), 1626–1628.
Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Computational Biology*, *5*(7), 1–18.
Chawla, P., & Krauss, R. M. (1994). Gesture and speech in spontaneous and rehearsed narratives. *Journal of Experimental Social Psychology*, *30*(6), 580–601.
Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology*, *37*(3), 201–242.
Cooke, M., Barker, J., Cunningham, S., & Shao, X. (2006). An audio-visual corpus for speech perception and automatic speech recognition. *The Journal of the Acoustical Society of America*, *120*(5), 2421–2424.
Crystal, T. H., & House, A. S. (1982). Segmental durations in connected speech signals: Preliminary results. *The Journal of the Acoustical Society of America*, *72*(3), 705–716.
Cummins, F. (2012a). Looking for rhythm in speech. *Empirical Musicology Review*, *7*(1-2), 28–35.
Cummins, F. (2012b). Oscillators and syllables: A cautionary note. *Frontiers in Psychology*, *3*, 1–2.
Cummins, F., & Port, R. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics*, *26*(2), 145–171.
Debener, S., Herrmann, C. S., Kranczioch, C., Gembris, D., & Engel, A. K. (2003). Top-down attentional processing enhances auditory evoked gamma band activity. *Neuroreport*, *14*(5), 683–686.
Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology*, *25*(19), 2457–2465.
Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, *21*(11), 1664–1670.
Ding, N., Chatterjee, M., & Simon, J. Z. (2014). Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *NeuroImage*, *88*, 41–46.
Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158–164.
Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews*, *81(B)*, 181–187.
Ding, N., & Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, *107*(1), 78–89.
Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, *8*, 311–317.
Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, *85*, 761–768.
Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience*, *2*(10), 704–716.
Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, *30*(2), 210–233.
Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*(1), 57–84.
Finke, M., & Rogina, I. (1997). Wide context acoustic modeling in read vs. spontaneous speech 1997 *IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 3, pp. 1743-1746). Los Alamitos: IEEE Computer Society Press.
Ghazanfar, A. A., Morrill, R. J., & Kayser, C. (2013). Monkeys are perceptually tuned to facial expressions that exhibit a theta-like speech rhythm. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(5), 1959–1963.
Ghitza, O. (2011). Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology*, *2*, 1–13.
Ghitza, O. (2012). On the role of theta-driven syllabic parsing in decoding speech: Intelligibility of speech with a manipulated modulation spectrum. *Frontiers in Psychology*, *3*, 1–12.
Ghitza, O. (2013). The theta-syllable: A unit of speech information defined by cortical function. *Frontiers in Psychology*, *4*, 1–5.
Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, *66*(1-2), 113–126.
Giordano, B. L., Ince, R. A. A., Gross, J., Schyns, P. G., Panzeri, S., & Kayser, C. (2017). Contributions of local speech encoding and functional connectivity to audio-visual speech perception. *eLife*, *6*, e24763.
Giraud, A.-L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S. J., & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron*, *56*(6), 1127–1134.
Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517.
Goswami, U., & Leong, V. (2013). Speech rhythm and temporal structure: Converging perspectives? *Laboratory Phonology*, *4*(1), 67–92.
Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, *11*(12), 1–14.
Haegens, S., & Zion Golumbic, E. (2018). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience & Biobehavioral Reviews*, *86*, 150–165.
Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(49), 20095–20100.
Hertrich, I., Dietrich, S., & Ackermann, H. (2013). Tracking the speech signal – time-locked MEG signals during perception of ultra-fast and moderately fast speech in blind and in sighted listeners. *Brain and Language*, *124*(1), 9–21.

Hertrich, I., Dietrich, S., Trouvain, J., Moos, A., & Ackermann, H. (2012). Magnetic brain activity phase-locked to the envelope, the syllable onsets, and the fundamental frequency of a perceived speech signal. *Psychophysiology*, *49*(3), 322–334.

Hickok, G., Farahbod, H., & Saberi, K. (2015). The rhythm of perception: Entrainment to acoustic rhythms induces subsequent perceptual oscillation. *Psychological Science*, *26*(7), 1006–1013.

Hirose, K., & Kawanami, H. (2002). Temporal rate change of dialogue speech in prosodic units as compared to read speech. *Speech Communication*, *36*(1), 97–111.

Howard, M. F., & Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *Journal of Neurophysiology*, *104*(5), 2500–2511.

Iversen, J. R., Repp, B. H., & Patel, A. D. (2009). Top-down control of rhythm perception modulates early auditory responses. *Annals of the New York Academy of Sciences*, *1169*(1), 58–73.

Jensen, O., Kaiser, J., & Lachaux, J.-P. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends in Neurosciences*, *30*(7), 317–324.

Kayser, S. J., Ince, R. A., Gross, J., & Kayser, C. (2015). Irregular speech rate dissociates auditory cortical entrainment, evoked responses, and frontal alpha. *The Journal of Neuroscience*, *35*(44), 14691–14701.

Keitel, A., & Gross, J. (2016). Individual human brain areas can be identified from their characteristic spectral activation fingerprints. *PLoS Biology*, *14*(6), e1002498.

Keitel, A., Ince, R. A., Gross, J., & Kayser, C. (2017). Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. *NeuroImage*, *147*, 32–42.

Kösem, A., Basirat, A., Azizi, L., & Wassenhove, V. v. (2016). High-frequency neural activity predicts word parsing in ambiguous speech streams. *Journal of Neurophysiology*, *116*(6), 2497–2512.

Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, *8*(4), 194–208.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, *320*(5872), 110–113. doi:10.1126/science.1154735

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., & Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of Neurophysiology*, *94*(3), 1904–1911.

Luo, H., Boemio, A., Gordon, M., & Poeppel, D. (2007). The perception of FM sweeps by Chinese and English listeners. *Hearing Research*, *224*(1-2), 75–83.

Luo, H., Liu, Z., & Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biology*, *8*(8), 1–13.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*(6), 1001–1010.

Mai, G., Minett, J. W., & Wang, W. S. Y. (2016). Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing. *NeuroImage*, *133*, 516–528.

Makeig, S., Westerfield, M., Jung, T.-P., Enghoff, S., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2002). Dynamic brain sources of visual evoked responses. *Science*, *295*(5555), 690–694.

Mäkelä, J. P., Ahonen, A., Hämäläinen, M., Hari, R., Llmoniemi, R., Kajola, M., … Salmelin, R. (1993). Functional differences between auditory cortices of the two hemispheres revealed by whole-head neuromagnetic recordings. *Human Brain Mapping*, *1*(1), 48–56.

Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., & Ro, T. (2009). To see or not to see: Prestimulus α phase predicts visual awareness. *The Journal of Neuroscience*, *29*(9), 2725–2732.

Meyer, L. (2017). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience*, 1–13. doi:10.1111/ejn.13748

Meyer, L., Henry, M. J., Gaston, P., Schmuck, N., & Friederici, A. D. (2017). Linguistic bias modulates interpretation of speech via neural delta-band oscillations. *Cerebral Cortex*, *27*(9), 4293–4302.

Meyer, L., Sun, Y., & Martin, A. E. (2018). Entrainment in disguise: The exogenous and endogenous cortical rhythms of speech and language processing. *PsyArXiv*.

Miller, J. L., Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica*, *41*(4), 215–225.

Millman, R. E., Johnson, S. R., & Prendergast, G. (2015). The role of phase-locking to the temporal envelope of speech in auditory perception and speech intelligibility. *Journal of Cognitive Neuroscience*, *27*(3), 533–545.

Morillon, B., & Schroeder, C. E. (2015). Neuronal oscillations as a mechanistic substrate of auditory temporal prediction. *Annals of the New York Academy of Sciences*, *1337*(1), 26–31.

Nakajima, S. y., & Allen, J. F. (1993). A study on prosody and discourse structure in cooperative dialogues. *Phonetica*, *50*(3), 197–210.

Neuling, T., Rach, S., Wagner, S., Wolters, C. H., & Herrmann, C. S. (2012). Good vibrations: Oscillatory phase shapes perception. *NeuroImage*, *63*(2), 771–778.

Nolan, F., & Jeon, H.-S. (2014). Speech rhythm: A metaphor? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1658), 20130396. http://dx.doi.org/10.1098/rstb.2013.0396

Nourski, K. V., Reale, R. A., Oya, H., Kawasaki, H., Kovach, C. K., Chen, H., … Brugge, J. F. (2009). Temporal envelope of time-compressed speech represented in the human auditory cortex. *Journal of Neuroscience*, *29*(49), 15564–15574.

Park, H., Ince, R. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology*, *25*(12), 1649–1653.

Payton, K. L., Uchanski, R. M., & Braida, L. D. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *The Journal of the Acoustical Society of America*, *95*(3), 1581–1592.

Peelle, J. E. (2018). Speech comprehension: Stimulating discussions at a cocktail party. *Current Biology*, *28*(2), R68–R70.

Peelle, J., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*, 1–17.

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, *23*(6), 1378–1387.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*, *28*(1), 96–103.

Puschmann, S., Steinkamp, S., Gillich, I., Mirkovic, B., Debener, S., & Thiel, C. M. (2017). The right temporoparietal junction supports speech tracking during selective listening: Evidence from concurrent EEG-fMRI. *The Journal of Neuroscience*, *37* (47), 11505–11516.

Ruspantini, I., Saarinen, T., Belardinelli, P., Jalava, A., Parviainen, T., Kujala, J., & Salmelin, R. (2012). Corticomuscular coherence is tuned to the spontaneous rhythmicity of speech at 2–3 Hz. *The Journal of Neuroscience*, *32*(11), 3786–3790.

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, *32*(1), 9–18.

Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Current Opinion in Neurobiology*, *20*(2), 172–176.

Scott, S. K., Rosen, S., Lang, H., & Wise, R. J. S. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech—A positron emission tomography study. *The Journal of the Acoustical Society of America*, *120*(2), 1075–1083.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, *416*(6876), 87–90.

Steinschneider, M., Nourski, K. V., & Fishman, Y. I. (2013). Representation of speech in human auditory cortex: Is it special. *Hearing Research*, *305*, 57–73.

Swerts, M., Strangert, E., & Heldner, M. (1996). F0 declination in read-aloud and spontaneous speech. In H. T. Bunnel & W. Idsardi (Eds.), *Proceedings of the fourth international conference on spoken language processing* (Vol. 3, pp. 1501–1504). Philadelphia, USA: IEEE.

Telkemeyer, S., Rossi, S., Koch, S. P., Nierhaus, T., Steinbrink, J., Poeppel, D., … Wartenburger, I. (2009). Sensitivity of newborn auditory cortex to the temporal structure of sounds. *The Journal of Neuroscience*, *29*(47), 14726–14733.

Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America*, *134*(1), 628–639.

Tree, J. E. F. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language*, *34*(6), 709–738.

Tree, J. E. F. (2001). Listeners' uses of *um* and *uh* in speech comprehension. *Memory & Cognition*, *29*(2), 320–326.

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., & Durlach, N. I. (1996). Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech, Language, and Hearing Research*, *39*(3), 494–509.

Wang, X.-J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological Reviews*, *90*(3), 1195–1268. doi:10.1152/physrev.00035.2008

Yaruss, J. S. (1999). Utterance length, syntactic complexity, and childhood stuttering. *Journal of Speech, Language, and Hearing Research*, *42*(2), 329–344.

Zhou, H., Melloni, L., Poeppel, D., & Ding, N. (2016). Interpretations of frequency domain analyses of neural entrainment: Periodicity, fundamental frequency, and harmonics. *Frontiers in Human Neuroscience*, *10*, 1–8.

Zion-Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., … Simon, J. Z. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron*, *77*(5), 980–991.

Zoefel, B., Archer-Boyd, A., & Davis, M. H. (2018). Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. *Current Biology*, *28*(3), 401–408. e405.

Zoefel, B., ten Oever, S., & Sack, A. T. (2018). The involvement of endogenous neural oscillations in the processing of rhythmic input: More than a regular repetition of evoked neural responses. *Frontiers in Neuroscience*, *12*, 1–13.

Zoefel, B., & VanRullen, R. (2015). The role of high-level processes for oscillatory phase entrainment to speech sound. *Frontiers in Human Neuroscience*, *9*, 1–12.