



# The Role of High-Level Processes for Oscillatory Phase Entrainment to Speech Sound

Benedikt Zoefel<sup>1,2\*</sup> and Rufin VanRullen<sup>1,2</sup>

<sup>1</sup> Université Paul Sabatier, Toulouse, France, <sup>2</sup> Centre de Recherche Cerveau et Cognition (CerCo), CNRS, UMR5549, Pavillon Baudot CHU Purpan, Toulouse, France

Constantly bombarded with input, the brain has the need to filter out relevant information while ignoring the irrelevant rest. A powerful tool may be represented by neural oscillations which entrain their high-excitability phase to important input while their low-excitability phase attenuates irrelevant information. Indeed, the alignment between brain oscillations and speech improves intelligibility and helps dissociating speakers during a “cocktail party”. Although well-investigated, the contribution of low- and high-level processes to phase entrainment to speech sound has only recently begun to be understood. Here, we review those findings, and concentrate on three main results: (1) Phase entrainment to speech sound is modulated by attention or predictions, likely supported by top-down signals and indicating higher-level processes involved in the brain’s adjustment to speech. (2) As phase entrainment to speech can be observed without systematic fluctuations in sound amplitude or spectral content, it does not only reflect a passive steady-state “ringing” of the cochlea, but entails a higher-level process. (3) The role of intelligibility for phase entrainment is debated. Recent results suggest that intelligibility modulates the behavioral consequences of entrainment, rather than directly affecting the strength of entrainment in auditory regions. We conclude that phase entrainment to speech reflects a sophisticated mechanism: several high-level processes interact to optimally align neural oscillations with predicted events of high relevance, even when they are hidden in a continuous stream of background noise.

## OPEN ACCESS

### Edited by:

Johanna Maria Rimmele,  
University Medical Center  
Hamburg-Eppendorf, Germany

### Reviewed by:

Christian Keitel,  
University of Glasgow, UK  
Elana Zion Golumbic,  
Bar Ilan University, Israel

### \*Correspondence:

Benedikt Zoefel  
zoefel@cerco.ups-tlse.fr

**Received:** 30 July 2015

**Accepted:** 16 November 2015

**Published:** 02 December 2015

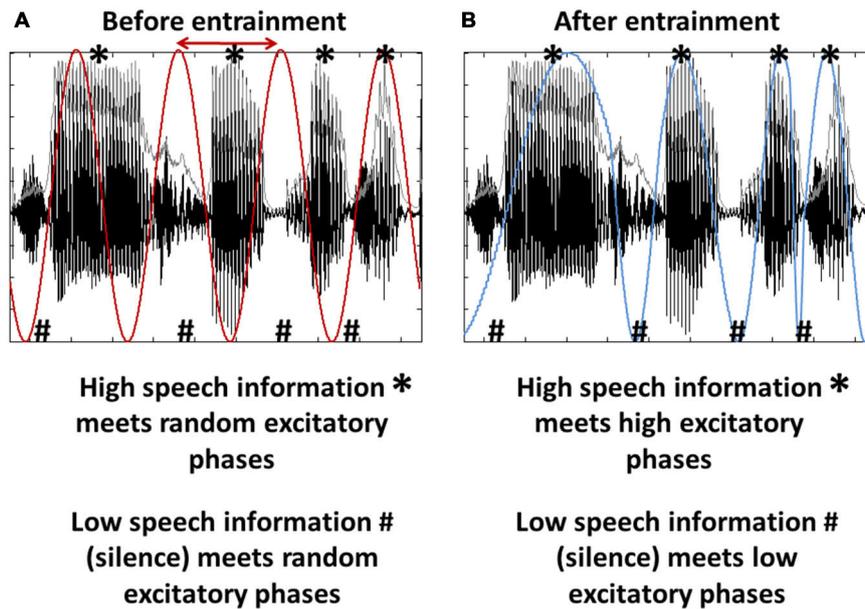
### Citation:

Zoefel B and VanRullen R (2015) The Role of High-Level Processes for Oscillatory Phase Entrainment to Speech Sound. *Front. Hum. Neurosci.* 9:651. doi: 10.3389/fnhum.2015.00651

**Keywords:** EEG, oscillation, phase, entrainment, high-level, speech, auditory, intelligibility

## PHASE ENTRAINMENT AS A TOOL FOR INPUT GATING

In virtually every situation of our life, the brain has to cope with an enormous amount of incoming information, only a fraction of which is essential for the scene’s interpretation or resulting behavior. Clearly, the brain must have evolved strategies to deal with this vast influx, and both amplification of relevant input and suppression of irrelevant information will be critical for survival. Based on recent research, one prominent tool for the described purpose are neural oscillations, assumed to reflect cyclic changes in the excitability of groups of neurons (Buzsáki and Draguhn, 2004; Rajkai et al., 2008; Mazzone et al., 2010). These *endogenous* fluctuations in neural excitability *per se* might seem without function at first glance, as long as they are *passive* and unrelated to the environment (**Figure 1A**). However, as previous studies showed, both on a theoretical (Schroeder et al., 2008, 2010;



**FIGURE 1 | Entrainment as a tool for input gating. (A)** Brain oscillations (red) are unrelated to the stimulus input, here a segment of speech sound. Note that both oscillation and speech sound are rhythmic ( $\sim 4$  Hz) and that the speech input consists of phases of high (\*) and low (#) informational content. Both phase and frequency (the latter to a certain extent; Ghitza, 2013, 2014) of the oscillations can be adjusted to match the input rhythm (red arrow), a phenomenon called phase entrainment. **(B)** Phase entrainment results in an alignment of the oscillation's high and low excitability phases (blue) with the input's high and low informational content. It can thus be used as a tool for input gating.

Schroeder and Lakatos, 2009; Ghitza, 2011; Morillon et al., 2015) and experimental level (Lakatos et al., 2005, 2008, 2013; Stefanics et al., 2010; Besle et al., 2011; Henry and Obleser, 2012; Henry et al., 2014; Morillon et al., 2014; Nozaradan, 2014; O'Connell et al., 2014; Arnal et al., 2015; Park et al., 2015), these oscillations might become an interesting tool when introducing the possibility that they can be *controlled* by the brain. By using the low and high excitability phases of those oscillations, the brain might *actively* “decide” what part of the incoming information should be amplified (the information coinciding with the oscillation's high excitability phase) and what part should be suppressed (the information coinciding with the oscillation's low excitability phase; **Figure 1B**). This phenomenon, the synchronization of an oscillatory system (here: brain oscillations) with external input has been termed *phase entrainment* (Schroeder and Lakatos, 2009). Of course, this kind of “input gating” can only be exploited functionally if the input is (1) rhythmic (i.e., predictable), (2) has a relatively stable frequency that the brain can entrain to, and (3) alternates between low and high informational content. Interestingly, one of the most salient stimuli in everyday life fulfills these criteria: speech sound. Although only considered “pseudo-rhythmic” (Cummins, 2012; but see Ghitza, 2013), the frequency of the speech envelope (roughly defined as the sum of energy across sound frequencies at a given point in time; shown as gray line in **Figure 1**) is relatively stable between 2 and 8 Hz and phases of low phonetic information (e.g., the silence between syllables) rhythmically alternate with phases of high phonetic information.

Indeed, the number of studies reporting an adaptation of neural oscillations to the envelope of speech sound is increasing continuously (Ding and Simon, 2012a,b, 2013, 2014; Peelle and Davis, 2012; Zion Golumbic et al., 2012, 2013b; Ding et al., 2013; Gross et al., 2013; Horton et al., 2013; Peelle et al., 2013; Power et al., 2013; Steinschneider et al., 2013; Doelling et al., 2014; Millman et al., 2015; Park et al., 2015). But not only speech sound is able to evoke an entrainment of neural oscillations, even simple stimuli, such as pure tones, have been found to produce phase entrainment (Stefanics et al., 2010; Besle et al., 2011; Gomez-Ramirez et al., 2011; Zoefel and Heil, 2013). Furthermore, rhythmic fluctuations in stimulus amplitude (which are present in both trains of pure tones and speech sound) introduce fluctuations at a level of auditory processing as low as the cochlea, a notion that is obviously not compatible with phase entrainment as an active or “high-level” process. Similar concerns have been raised by several authors in the last years (Obleser et al., 2012; Zion Golumbic et al., 2012; Ding et al., 2013; Peelle et al., 2013; Zoefel and Heil, 2013; Ding and Simon, 2014; VanRullen et al., 2014). Based on these concerns, it might be argued that a mere “following” of stimulus amplitude (leading to a series of evoked potentials) and the entrainment of endogenous neural oscillations might be completely different processes with different types of underlying mechanisms. Most studies investigating phase entrainment did not differentiate these components and might have measured a mix of evoked and entrained responses. For the sake of simplicity, and because it is not straightforward to disentangle the two, we will call both

processes “phase entrainment” throughout this manuscript, to describe an experimentally observable metric without assuming one or the other underlying process. However, we dedicated the last paragraph of Section “Phase Entrainment to High-Level Features of Speech Sound” to this issue, in which the controversy “evoked vs. entrained” is discussed in more detail.

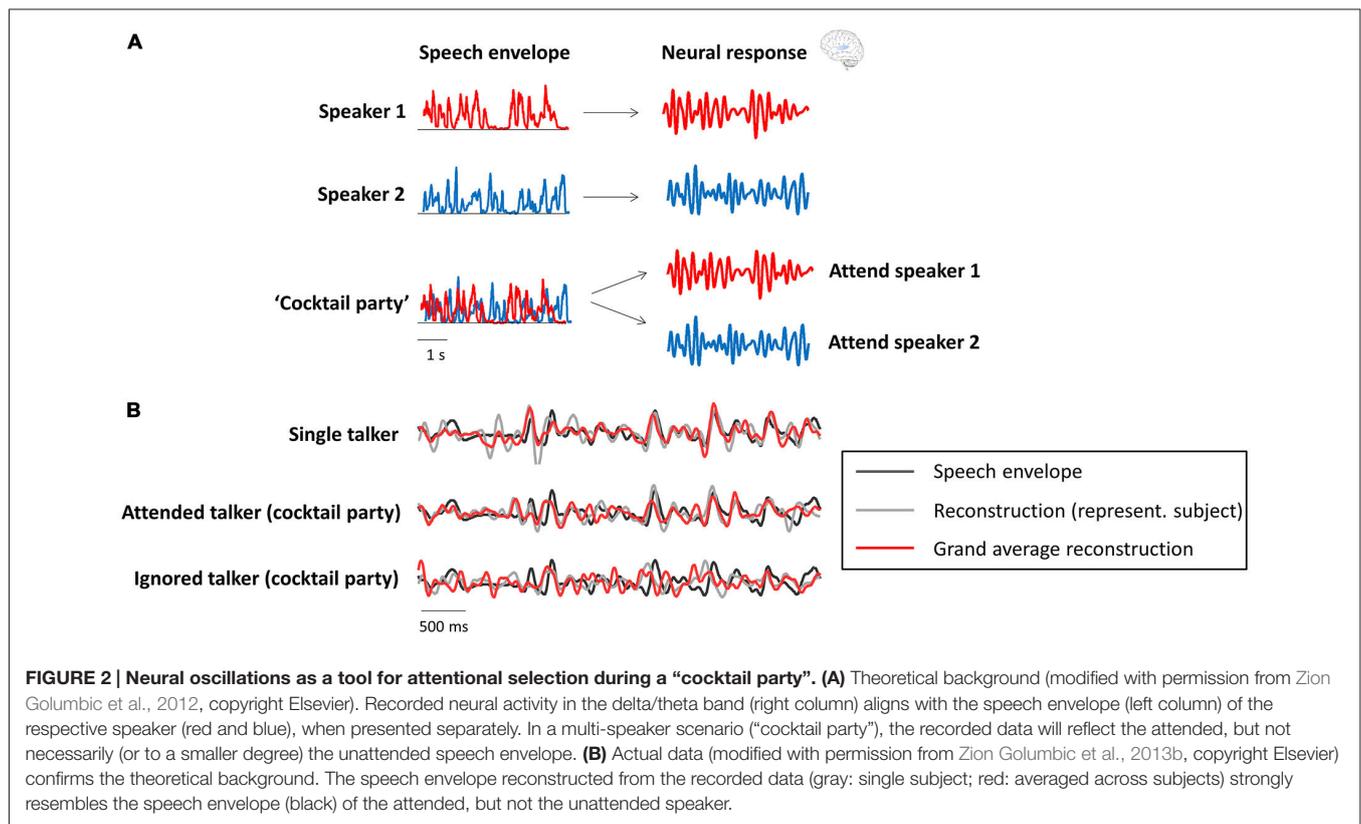
The issues outlined in the previous paragraph lead to the fact that the role of high-level processes for phase entrainment to speech sound is far from clear. Nevertheless, significant progress has been made within the last decade, and the aim of this review is to summarize the obtained results in a systematic way. The scope of this review is not a summary of existing literature showing an alignment between brain oscillations and speech sound, as comprehensive reviews have been published recently (Pelle and Davis, 2012; Zion Golumbic et al., 2012; Ding and Simon, 2014). Rather, we will focus on high-level processes that can modulate or even underlie this alignment. Critically, it is necessary to differentiate between (i) high-level *modulations* of phase entrainment and (ii) high-level *entrainment*: In (i), phase entrainment can be produced as a “following” response to a low-level rhythmic stimulus sequence (potentially in early brain areas, as early as the cochlea); however, the entrainment is modulated by high-level processes that include attention or predictions. In this review, low-level features of speech are defined as stimulus amplitude and spectral content, as those two properties can passively entrain the lowest level of auditory processing and evoke steady-state-potential-like (ASSR; Galambos et al., 1981) fluctuations in the cochlea. In contrast to (i), high-level *entrainment* (ii) represents phase entrainment that can be observed even in the absence of systematic fluctuations of low-level properties. In this case, a simple “following” of stimulus amplitude is not possible anymore. Thus, it is the process of phase entrainment *itself* that operates on a higher level, as a certain level of processing is required in order to adjust to the rhythm of high-level features. Convincing results have been obtained in the last years for both types of high-level processes, and we will address them in separate sections. We conclude this review with a section dedicated to the role of intelligibility for phase entrainment to speech sound, as the influence of semantic information on the brain’s adjustment to speech is currently a strongly debated topic.

## HIGH-LEVEL MODULATIONS OF PHASE ENTRAINMENT TO SPEECH SOUND

Certain cognitive processes, such as attention, expectation or interpretation, are often considered “high-level” functions of the human brain, as they require, for instance, evaluation, selection, and the comparison of the actual stimulation with experience (Lamme and Spekreijse, 2000; Gilbert and Li, 2013; Peelen and Kastner, 2014). A modulation of phase entrainment to speech sound by those cognitive processes would argue for phase entrainment being a process that is not restricted to a purely sensory mechanism, but rather the active gating mechanism (or “active sensing”; Schroeder et al., 2010) that was explained above.

Indeed, there is accumulating evidence for phase entrainment critically relying on attentional processes: one example is based on the so-called “cocktail party effect” (Cherry, 1953), describing a situation of several competing speakers, one of which has to be selected within the “noise” of the other, potentially distracting, speakers.

Several recent studies have shown a relation between the “cocktail party effect” and phase entrainment (the theoretical background is shown in **Figure 2A** and underlined by experimental results in **Figure 2B**). In Kerlin et al. (2010), two different speech streams were presented to the participants, one to each ear, and they were asked to selectively attend one of those two competing streams. They found that the representation of the attended speech stream in the delta/theta range (~2–8 Hz; the dominant frequency range of the speech envelope) of the electroencephalogram (EEG) signal was enhanced compared to that of the unattended stream. In other words, phase-locking between the EEG signal and the speech envelope of the attended stream was stronger than that between the EEG signal and the unattended stream. A similar paradigm was used in the studies by Ding and Simon (2012a), Horton et al. (2013) and Zion Golumbic et al. (2013b) in magnetoencephalographic (MEG), EEG and intracranial recordings in human subjects, respectively. All studies confirmed the finding that the phase of delta/theta brain oscillations “tracks” the envelope of speech sound, and that this “tracking” is enhanced when the speech is attended in a multi-speaker scenario. Interestingly, all studies reported that even the unattended speech signal is still represented (albeit weakly) in lower-level auditory cortices (i.e., regions closely related to sensory processing). However, as shown in the work by Zion Golumbic et al. (2013b), this unattended signal is “lost” in higher-level (e.g., frontal) regions. Ding and Simon (2012a) demonstrated that only the representation of the attended (and not the unattended) speech envelope varies as a function of stimulus intensity. This finding is important, because it suggests that attended and unattended inputs are processed separately in the brain, and that the alignment between neural phase and speech rhythm is used to form individual “auditory objects” (for a review on this notion, see Simon, 2015). In line with the notion of phase entrainment as an “amplifier-attenuator mechanism” (see “Phase Entrainment as a Tool for Input Gating”), Horton et al. (2013) reported cross-correlations between speech envelope and EEG signal for both attended and unattended streams, but with opposite signs, suggesting that phase entrainment is indeed used to amplify one stream while the other is attenuated. Finally, it has been shown in several studies that the speech envelope can be reconstructed (i.e., it can be identified which stimulus the listener is attending) in multi-speaker (Ding and Simon, 2012a; Zion Golumbic et al., 2013b; O’Sullivan et al., 2015) or noisy environments (Ding and Simon, 2013) by using the delta/theta phase of neural oscillations (but also their gamma power; Mesgarani and Chang, 2012; Zion Golumbic et al., 2013b). It is possible that in those kind of situations, where one speech stream has to be actively extracted from a noisy environment, attention is of particular importance for phase entrainment to speech sound, whereas clear speech can



be processed largely independently of attention (Wild et al., 2012).

Not only attention can be considered a high-level process: predictions reflect a comparison between present and previous experiences and its projection to the future and must therefore involve high-level functions of the brain (Friston, 2005; Arnal and Giraud, 2012). Indeed, it has been shown that predictions do influence phase entrainment to speech sound. For instance, in the “cocktail party” scenario described above, Zion Golumbic et al. (2013a) paired the auditory speech input with the speaker’s face and found that phase entrainment to the speech envelope was significantly enhanced by this visual input. Similar results were obtained by Arnal et al. (2011) using congruent and incongruent audiovisual stimuli (syllables) and by Luo et al. (2010) when subjects were watching audiovisual movies. A common interpretation of these findings is that, due to the slight delay between visual and auditory components of a conversation (the visual input preceding the auditory one), the former can be used to predict the timing of speech sound, thus enabling a better alignment between the oscillatory phase and speech envelope (Arnal et al., 2009, 2011; Zion Golumbic et al., 2013a; Perrodin et al., 2015; for a review, summarizing several existing theories, see Peelle and Sommers, 2015). A phase-reset of neural oscillations in primary auditory cortex by visual input seems to be an important underlying mechanism (Thorne and Debener, 2014; Mercier et al., 2015; Perrodin et al., 2015). Although this would indicate an involvement of low or intermediate hierarchical levels, we emphasize here that

a purely low-level mechanism is insufficient to explain many findings reported in the literature. For instance, introducing an additional delay between visual and auditory input disrupts the benefits of additional visual information for speech processing and incongruent visual information (which would result in a similar phase-reset as congruent information, assuming a purely low-level process) does not result in enhanced phase entrainment (e.g., Crosse et al., 2015; for a review, see Peelle and Sommers, 2015) but instead generates an increased neural response associated with error processing (Arnal et al., 2011). Finally, using a McGurk paradigm (McGurk and MacDonald, 1976; van Wassenhove et al., 2005) were able to show a correlation between the amount of prediction conveyed by the preceding visual input for the upcoming speech and the latency of speech processing. Together, these results speak for a mechanism that is tailored to speech-specific processing (Crosse et al., 2015) and against a purely low-level mechanism. The timing of the cross-modal phase-reset seems to have evolved in such a way that oscillations in the auditory system arrive at their high excitability phase exactly when the relevant auditory input is expected to be processed (Lakatos et al., 2009; Thorne and Debener, 2014). Finally, recent research suggests that not only the visual, but also the motor system plays a critical role for an efficient adjustment of excitability fluctuations in auditory cortex to expected upcoming events (Fujioka et al., 2012; Doelling et al., 2014; Morillon and Schroeder, 2015; Morillon et al., 2015). For instance, it has been suggested that the motor system possesses its own representation of expected auditory

events and can therefore prepare oscillations in auditory cortex for relevant upcoming stimuli (Arnal and Giraud, 2012; Arnal, 2012). This mechanism might underlie recent findings describing an enhanced segregation of relevant and irrelevant auditory events in the presence of rhythmic tapping (Morillon et al., 2014).

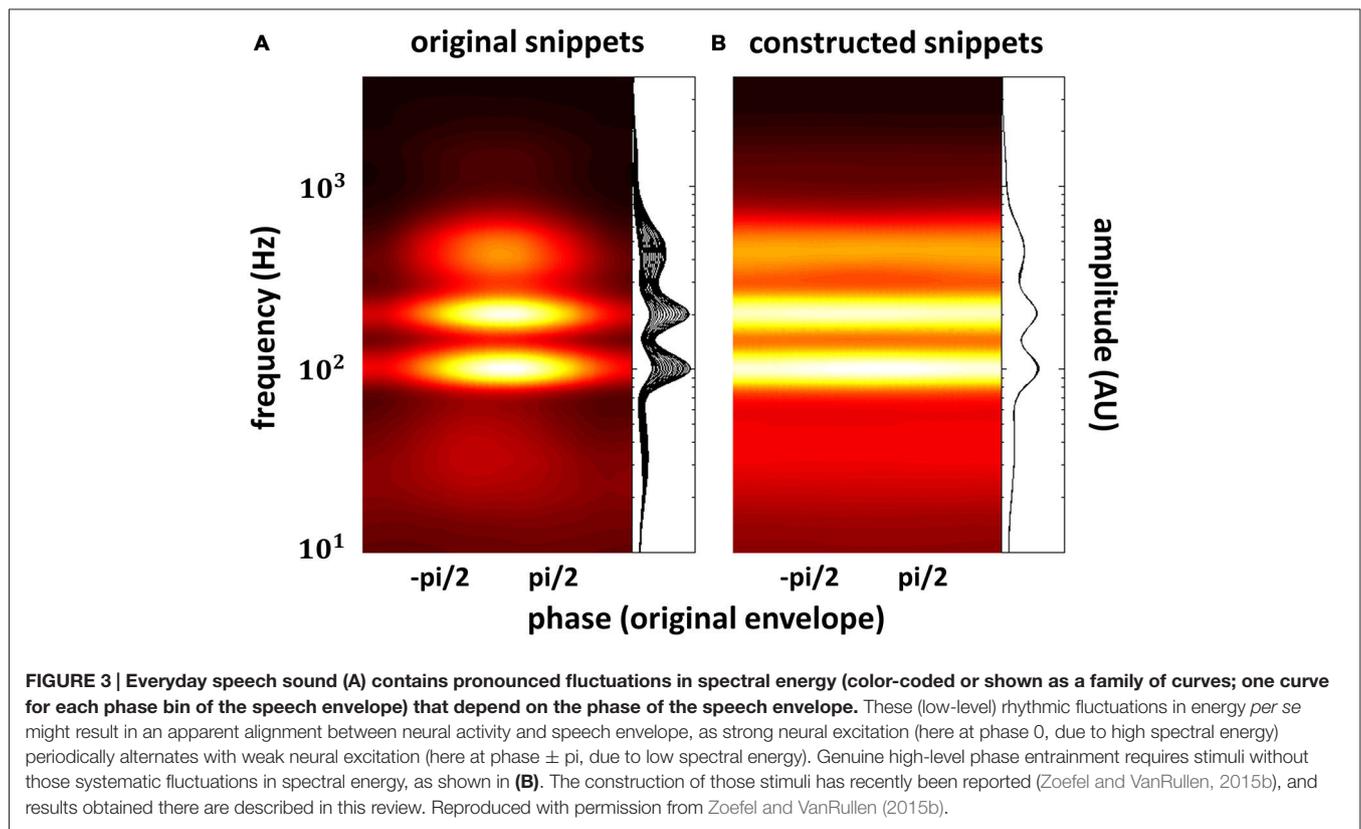
Not an experimental, but rather an analytical proof of high-level processes involved in phase entrainment was provided by two recent studies (Fontolan et al., 2014; Park et al., 2015). Fontolan et al. (2014) used Granger causality (Granger, 1969), applied on data recorded intracranially in human subjects, to demonstrate that information reflected in the phase of low-frequency oscillations in response to speech sound travels in top-down direction from higher-order auditory to primary auditory regions, where it modulates the power of (gamma) oscillations at higher frequencies. Park et al. (2015) analyzed their data, recorded with MEG, using transfer entropy measures (Schreiber, 2000). They were able to show that frontal and motor areas can modulate the phase of delta/theta oscillations in auditory cortex (note that the spatial resolution in this study was lower than for intracranial recordings. It is thus unclear whether these delta/theta oscillations correspond to those in higher-order auditory or primary auditory cortices described in Fontolan et al., 2014). Importantly, these top-down signals were correlated with an enhanced phase entrainment to speech sound when tracking of forward vs. backward speech was compared, indicating that higher-level processes can directly control the alignment between neural oscillations and speech sound.

The results described in this section strongly support the view that phase entrainment is a tool for attentional selection (Schroeder and Lakatos, 2009), filtering out irrelevant input and enhancing the representation of the attended stimulus in the brain. Predictions, potentially reflected by top-down mechanisms, help “designing” this filter by providing the timing for the alignment of “good” and “bad” phases of the oscillation to predicted relevant and irrelevant stimuli, respectively. This mechanism would not only help selecting relevant input in a noisy background, but also parse the speech signal at the same time: here, one cycle of the aligned oscillation would represent one segment of information (or “chunk”; Ghitza, 2011, 2013, 2014; Doelling et al., 2014) that is analyzed by means of faster oscillations (Giraud and Poeppel, 2012; Luo and Poeppel, 2012; for reviews, see Peelle and Davis, 2012; Ding and Simon, 2014). Thus, phase entrainment could function as a means of discretization (equivalent ideas are mentioned by Peelle and Davis, 2012; Zion Golumbic et al., 2012), similar to “perceptual cycles” commonly observed in vision (VanRullen et al., 2014).

## PHASE ENTRAINMENT TO HIGH-LEVEL FEATURES OF SPEECH SOUND

In the previous section, we have seen that high-level mechanisms of the brain, related to attention or prediction, clearly contribute to phase entrainment to speech sound. However, it should be noted that this contribution may just be *modulatory*: high-level mechanisms could merely *influence* a process, namely

phase entrainment, that *itself* might rely on purely low-level processes. Indeed, speech sound consists of large fluctuations in low-level properties (i.e., stimulus amplitude and spectral content) that might evoke systematic fluctuations in neural activity already at the earliest level of auditory processing: the cochlea. These fluctuations in neural activity accompanying changes in the speech envelope would be indistinguishable from an active entrainment response. It is therefore necessary to construct stimuli without systematic fluctuations in those low-level properties in order to prove genuine high-level entrainment. In a recent publication (Zoefel and VanRullen, 2015b), we were able to construct such stimuli and we review the most important findings in this section, together with supporting results from other studies. **Figure 3** shows the idea underlying stimulus construction in Zoefel and VanRullen (2015b). In everyday speech sound (**Figure 3A**), spectral energy (color-coded) clearly differs between different phases of the speech envelope. In the view of a single cochlear cell, this sound would periodically alternate between weak (e.g., at phase  $\pm \pi$ , which is the trough of the speech envelope) and strong excitation (e.g., at phase 0, which is the peak of the speech envelope). Consequently, at a larger scale, we would measure an oscillatory pattern of neural activity that strongly depends on envelope phase. This pattern, however, would only reflect the periodicity of the stimulation. Therefore, we constructed noise sound whose spectral energy was tailored to counterbalance spectral differences as a function of envelope phase of the original speech sound (for details of stimulus construction, see Zoefel and VanRullen, 2015b). This noise was mixed with the original speech and resulted in speech/noise sound that did, on average, not show those systematic differences in spectral content anymore (**Figure 3B**). Critically, as those stimuli remain intelligible, high-level features of speech (such as, but not restricted to, phonetic information) are still present and enable the listener to entrain to the speech sound that is now “hidden” inside the noise (note that the degree to which the speech is “hidden” in noise depends on the original envelope phase, with speech perceptually dominant at the original envelope peak, and noise perceptually dominant at the original envelope trough). We applied those stimuli in two studies: in the first (Zoefel and VanRullen, 2015b), a psychophysical study, we found that the detection of a short tone pip was significantly modulated ( $p$ -values shown in **Figure 4A**) by the remaining high-level features. Performance (**Figure 4B**) depended on the original envelope phase and thus differed between periods of dominant speech and noise. Note that speech and noise were spectrally matched; differences in performance could thus not be due to spectral differences between speech and noise, but rather due to the remaining high-level features that enable the listener to differentiate speech and noise. In the second study (Zoefel and VanRullen, 2015a), those stimuli were presented to listeners while their EEG was recorded. We found that EEG oscillations phase-lock to those high-level features of speech sound (**Figure 4C**), and the degree of entrainment (but not the phase relation between speech and EEG signal; see insets in **Figure 4C**) was similar to when the original everyday



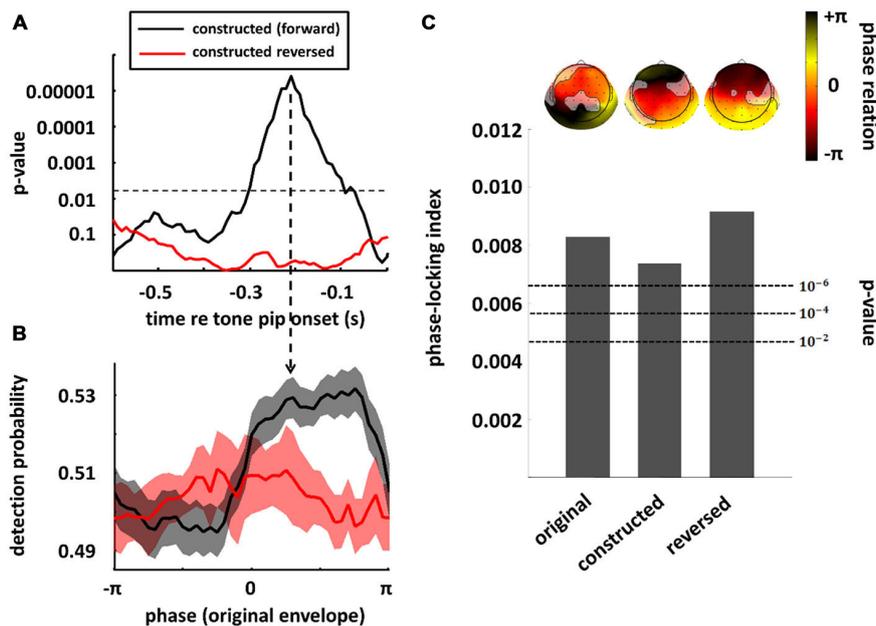
speech was presented. These results suggest an entrainment of neural oscillations as the mechanism underlying our perceptual findings.

It is not only interesting to investigate phase entrainment to speech stimuli *without* potentially entraining low-level features, but also to speech stimuli *only* containing the latter. This was done in a study by Ding et al. (2013) that might be seen as complementary to the other two described in this section. In their study, noise-vocoding (Green et al., 2002) was used in order to design stimuli where spectro-temporal fine structure (which can be considered as high-level features) was strongly reduced, but the speech envelope was essentially unchanged. Those stimuli were presented either in noise or in quiet, and MEG was recorded in parallel. Ding et al. (2013) showed that, indeed, the reduction of spectro-temporal fine structure in noise-vocoded speech results in a decline in phase entrainment as compared to that in response to natural speech sound. This result suggests that oscillations do not merely (and passively) follow the slow fluctuations in low-level features of speech (e.g., the speech envelope), as they are present in both natural and noise-vocoded speech. Instead, phase entrainment to speech sound involves an additional adjustment to rhythmic changes in spectro-temporal fine structure. It is important to mention that the effect was only observed in noise (and not in quiet), stressing the idea that separating speech and noise might be one of the main functions of phase entrainment to speech sound (see “Phase Entrainment as a Tool for Input Gating” and “High-Level Modulations of

Phase Entrainment to Speech Sound”). Using similar stimuli as in Ding et al. (2013), Rimmele et al. (2015) both extended their findings and built a bridge to our section “High-Level Modulations of Phase Entrainment to Speech Sound”. In contrast to Ding et al. (2013), they presented natural and noise-vocoded speech concurrently and asked their subjects to attend one of them while ignoring the other. Interestingly, they were able to show that the enhanced “envelope tracking” for natural compared to noise-vocoded speech (as in Ding et al., 2013) is only present when the speech is attended. They interpret their results as evidence for a high-level mechanism (“linguistic processing”) that is only possible when speech is in the focus of the listener’s attention, and only when speech contains spectro-temporal fine structure (i.e., high-level features). Finally, no attentional modulation of phase entrainment was found for noise-vocoded speech which might be taken as evidence for a tracking of low-level features that does not depend on top-down processes (e.g., attention; see “High-Level Modulations of Phase Entrainment to Speech Sound”).

Taken together, the results reported in this section suggest that phase entrainment to speech sound is not only a reflection of fluctuations in low-level features of speech sound, but entails an adaptation to phonetic information—and thus a genuine high-level process.

As briefly mentioned before, there is an ongoing debate which is directly related to the results presented in this section: as shown by Capilla et al. (2011), seemingly entrained



**FIGURE 4 | Perception and neural oscillations entrain to high-level features of speech sound.** Speech/noise stimuli without systematic fluctuations in amplitude or spectral content were constructed, whose high-level features are conserved and reflected by the original speech envelope (cf. **Figure 3**). In a psychophysical study (**A,B**), the detection of a tone pip was significantly modulated by those high-level features (black; in this plot, results obtained in the original experiment and a follow-up replication have been combined; both experiments are described in Zoefel and VanRullen, 2015b). The significance of this modulation is shown in (**A**) for different time lags relative to target onset, whereas the actual performance (at the time lag indicated by the vertical arrow in (**A**) is shown in panel (**B**). This effect was abolished when the speech/noise sound was reversed (red), indicating an important role of linguistic features (i.e., intelligibility) for behavioral consequences of the entrainment. In (**A**), the significance threshold is shown as a dashed line (corrected for multiple comparisons). In (**B**), standard error of mean (SEM) is shown by contours around the lines. When the same stimuli (and their original version of everyday speech) were presented in an EEG experiment (**C**) significant phase-locking between original speech envelope and EEG signal could be observed in all conditions (original, speech/noise sound and reversed speech/noise sound), suggesting that high-level features can entrain the phase of EEG oscillations, and do so even if the speech is unintelligible (note that acoustic high-level features remain present in the speech/noise sound, even when it is reversed, as the listener can still differentiate speech and noise). Bars show the average phase-locking across EEG channels, whereas the actual phase differences between EEG signal and original speech envelope, separately for each channel, are shown as insets above the bars (channels without significant entrainment are shaded out). *P*-values of phase entrainment, obtained by permutation tests, are shown as dashed lines. Note that, in contrast to the degree of entrainment which is comparable in all three conditions, the entrained phase does differ between everyday speech sound (original condition) and speech/noise sound in which systematic fluctuations in low-level features have been removed (constructed and constructed reversed conditions). Modified with permission from Zoefel and VanRullen (2015b) (**A,B**) and Zoefel and VanRullen (2015a), copyright Elsevier (**C**).

oscillations can be explained by a superposition of evoked responses (see also Keitel et al., 2014). Transferring this result to speech sound, it has been argued specifically that a phase-reset of neural oscillations by (e.g.) “acoustic edges” of speech might be an important mechanism underlying phase entrainment (Doelling et al., 2014; Howard and Poeppel, 2010)—assuming that these “edges” occur regularly in speech, a periodic sequence of phase-resets might thus be sufficient to explain the observed “phase entrainment”. This paragraph provides arguments against phase entrainment reflecting a purely passive mechanism, reflecting merely sequences of neural phase-resets or evoked potentials; however, we do emphasize here that most studies likely measure a mixture of evoked and entrained neural responses. As already outlined above, the two studies described in the first paragraph of this section (Zoefel and VanRullen, 2015a,b) support the notion that phase entrainment is more than a steady-state response to rhythmic stimulation: it can be observed even when the presented speech sound does not contain systematic fluctuations in amplitude or spectral content. Indeed, there are more

studies, using simpler, non-speech stimuli, that also support this conclusion. For instance, it has been found, for both vision (Spaak et al., 2014) and audition (Hickok et al., 2015), that behavioral performance fluctuates for several cycles after the offset of an entraining stimulus. A mere “following” of the stimulation would not produce these after-effects. Moreover, using entraining stimuli at threshold level, it has been shown that neural oscillations (as measured with EEG) entrain to the stimulation rate even when the stimulus is not perceived (e.g., in the case of several subsequent “misses”) and would therefore not evoke a strong neural response (Zoefel and Heil, 2013). Finally, in a clever experimental design, Herring et al. (2015) measured visual alpha oscillations (~8–12 Hz) after a single pulse of transcranial magnetic stimulation (TMS) that has previously been hypothesized to re-set (or entrain, in the case of multiple, rhythmic TMS pulses) endogenous oscillations (Thut et al., 2011). They then asked the question: how is the measured “alpha” modulated by attention? In the case of a simple evoked response (or “alpha- ringing”), the observed “alpha” would exhibit an increased amplitude when

attention is allocated to the visual domain; however, in the case of endogenous alpha, visual attention would decrease the alpha amplitude, as described already by Adrian (1944). Indeed, the latter is what Herring et al. (2015) observed. To conclude, although the issue remains open, there are promising first results suggesting that phase entrainment—to speech or other stimuli, including brain stimulation—is more than steady-state responses evoked by the rhythmic stimulation—it entails high-level processes and an adjustment of endogenous neural oscillations.

## THE ROLE OF INTELLIGIBILITY FOR PHASE ENTRAINMENT TO SPEECH SOUND

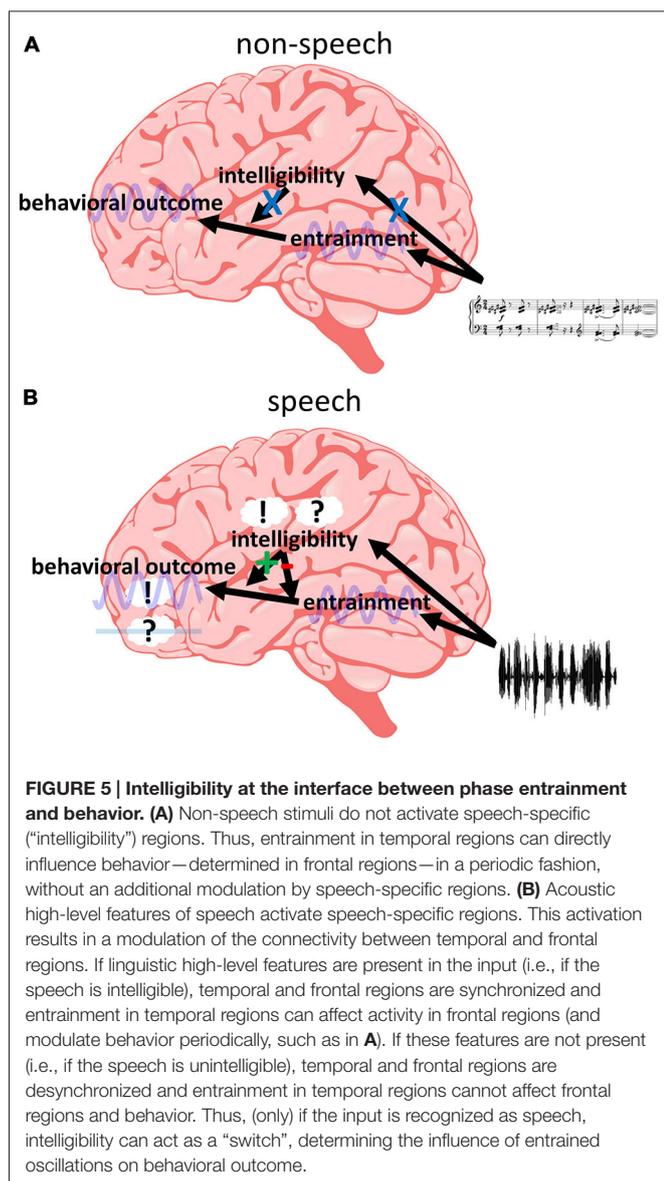
Of course, the ultimate goal of every conversation is to transmit information, and without intelligibility, this goal cannot be achieved. Thus, it is all the more surprising that the role of intelligibility for phase entrainment to speech is currently strongly debated. This controversy is due to seemingly contradictory results that have been published. On the one hand, both Ahissar et al. (2001) and Luo and Poeppel (2007) found a correlation between phase entrainment (i.e., alignment of delta/theta oscillations and speech envelope) and speech intelligibility, a finding that has been confirmed by recent studies (Ding et al., 2013; Doelling et al., 2014; Park et al., 2015). On the other hand, phase entrainment is not a phenomenon that is unique to speech sound and can also be found in response to much simpler stimuli, such as pure tones (Lakatos et al., 2005, 2008; Stefanics et al., 2010; Besle et al., 2011; Gomez-Ramirez et al., 2011; Zoefel and Heil, 2013). Also, the manipulation of speech intelligibility might destroy acoustic (i.e., non-semantic) properties of the sound that the brain actually entrains to (such as acoustic “edges”; Doelling et al., 2014), leading to a decline in phase entrainment and speech intelligibility at the same time, but without any relation between the two (Peelle and Davis, 2012; Millman et al., 2015). Moreover, several studies showed phase entrainment of neural oscillations to unintelligible speech sound (Howard and Poeppel, 2010; Peelle et al., 2013; Millman et al., 2015) suggesting that phase entrainment does not necessarily depend on intelligibility. The whole picture gets even more complicated, as, although phase entrainment to speech sound is possible even when the speech is unintelligible, it seems to be enhanced by intelligible speech in some (but not all) studies (Gross et al., 2013; Peelle et al., 2013; Park et al., 2015) and attention seems to be important for this enhancement (Rimmele et al., 2015). Further evidence that the role of intelligibility for phase entrainment is not trivial was reported in two of the studies described in the previous section. In Zoefel and VanRullen (2015b), it was found that perceptual entrainment to high-level features of speech sound is disrupted when the speech/noise sound is reversed (**Figures 4A,B**; red line) and this result was interpreted as a critical role of intelligibility for perceptual phase entrainment. On the other hand, in Zoefel and VanRullen (2015a), using the same reversed speech/noise stimuli, the observed EEG phase entrainment was similar to

that obtained in response to everyday speech and to (forward) speech/noise sound (**Figure 4C**), seemingly in contradiction to the behavioral results obtained in Zoefel and VanRullen (2015b).

How can we reconcile these studies, some of them clearly arguing against, and some for an important role of intelligibility for phase entrainment? Based on the current state of research, it is important to avoid overhasty conclusions and our interpretations have to remain speculative. Overall, phase entrainment seems to be a necessary, but not sufficient condition for speech comprehension. Speech intelligibility might not be possible without phase-locking, as we are not aware of any study reporting intelligible stimuli without oscillations (or perception) aligned to critical (low- and high-level) features of the speech sound. On the other hand, neural oscillations entrain to rhythmic structures (including reversed speech) even in the absence of intelligibility. It is clear that phase entrainment is a much more general phenomenon, and the brain might continuously scan its input for rhythmic patterns (indeed, popularity for auditory rhythms can be found in all cultures across the world and synchronization with rhythms—e.g., by clapping or dancing—is a general reaction to them). Once a rhythmic pattern has been detected, neural oscillations will align their phase to it (operating in the “rhythmic mode” described in Schroeder and Lakatos, 2009; see also Zoefel and Heil, 2013). Based on this notion, neural oscillations might always align to sound, as long as a rhythmic pattern can be detected (note that even the reversed speech/noise sound used in Zoefel and VanRullen, 2015a,b, contains a rhythmic pattern, as speech and noise can perceptually be differentiated). But what is the role of intelligibility? It is important to find a model that is at the same time parsimonious and can explain most results described in the literature. These findings are shortly summarized in the following:

1. Rhythmic non-speech stimuli, such as trains of pure tones, entrain neural oscillations (Lakatos et al., 2005; Besle et al., 2011; Gomez-Ramirez et al., 2011; Zoefel and Heil, 2013) and modulate behavior (Lakatos et al., 2008; Stefanics et al., 2010; Thut et al., 2012; Hickok et al., 2015).
2. Speech stimuli, both intelligible and unintelligible, entrain neural oscillations (Ahissar et al., 2001; Luo and Poeppel, 2007; Howard and Poeppel, 2010; Ding and Simon, 2012a; Ding et al., 2013; Peelle et al., 2013; Zion Golumbic et al., 2013b; Doelling et al., 2014; Millman et al., 2015; Park et al., 2015; Rimmele et al., 2015; Zoefel and VanRullen, 2015a).
3. The rhythm of speech only modulates behavior when speech is intelligible (Zoefel and VanRullen, 2015b).
4. Neural entrainment to intelligible speech might be increased when compared to unintelligible speech (Luo and Poeppel, 2007; Peelle et al., 2013; Doelling et al., 2014; Park et al., 2015; Rimmele et al., 2015). However, not all studies can confirm this result (Howard and Poeppel, 2010; Millman et al., 2015; Zoefel and VanRullen, 2015a).

One model that can potentially reconcile these findings is presented in **Figure 5**, and the different parts and implications of this model are discussed in the following. However, we



acknowledge that it is only one out of possibly several candidate models to explain the data available in the literature. Nevertheless, in our view, this model is currently the most parsimonious explanation for existing findings and we therefore focus our review on it. The first implication of our model is that different regions in the brain are “responsible” for different processes: Phase entrainment might be found throughout the whole auditory system, but most studies emphasize primary auditory cortex (A1; Lakatos et al., 2005, 2013; O’Connell et al., 2014) or early temporal regions (Gomez-Ramirez et al., 2011; Ding and Simon, 2012b; Zion Golumbic et al., 2013b). An influence of intelligibility is commonly related to regions specifically processing speech sound (Binder et al., 2000; Scott et al., 2000; Hickok and Poeppel, 2007; DeWitt and Rauschecker, 2012; Poeppel et al., 2012; Mesgarani et al., 2014). Finally, frontal regions are a likely candidate for behavioral outcome (Krawczyk, 2002; Coutlee and Huettel, 2012; Rushworth et al.,

2012; Romo and de Lafuente, 2013). In order to satisfy point (1), we assume that the entrainment in temporal regions can directly influence behavior as determined in frontal regions, as long as the entrainment is introduced by non-speech stimuli (Figure 5A). This results in a periodic modulation of performance as often described (Fiebelkorn et al., 2011; Vanrullen and Dubois, 2011; Landau and Fries, 2012; Thut et al., 2012; Song et al., 2014; Spaak et al., 2014; Zoefel and Sokoliuk, 2014; Hickok et al., 2015; note, however, that most studies report effects for the visual and not for the auditory system—it needs to be clarified whether this fact is biased by the number of studies investigating the visual system or whether there are genuine differences between the two systems). But not only non-speech stimuli can entrain temporal regions, the same is true for speech sound, irrespective of its intelligibility (point 2). However, speech intelligibility affects high-order auditory regions and they might directly influence the impact of temporal on frontal regions (Figure 5B). This notion is based on the increasing number of studies supporting the idea that the state of connectivity (or synchronization) between two (potentially distant) brain regions is crucial for perceptual outcome (Fries, 2005; Ruhnau et al., 2014; Weisz et al., 2014). Thus, speech intelligibility might modulate the state of connectivity between temporal and frontal regions. We hypothesize that speech-specific regions only become responsive if the input contains acoustic high-level (i.e., speech-specific) features of speech; otherwise these regions remain irrelevant and do not exhibit any modulatory effect on other regions or their connectivity. However, once the input is identified as speech (based on these acoustic features), linguistic features determine whether the modulatory effect is negative (desynchronizing temporal and frontal regions, resulting in no behavioral effect of the entrainment; in case of unintelligible speech) or positive (synchronizing temporal and frontal regions, resulting in a behavioral effect of the entrainment; in case of intelligible speech). This assumption satisfies point (3). In contrast to unintelligible speech, intelligible speech might result in an entrainment that also includes high-order (speech-specific) auditory regions. They might have to entrain to the speech sound in order to be able to synchronize temporal and frontal regions. That might be the reason that some studies show an increased entrainment for intelligible as compared to unintelligible speech whereas others do not (point 4). They might have captured the entrainment in those higher-level auditory regions—something which, due to the low spatial resolution in most EEG/MEG studies, is difficult to determine but could be resolved in future studies. More research is clearly needed: what are those behavioral variables that are differentially affected by intelligible and unintelligible speech? Where exactly are those brain regions hypothesized to be responsible for (or affected by) phase entrainment, for behavioral decisions and for the modulation of their relation by speech intelligibility? What are the mechanisms connecting these functional networks? Answering these questions has critical implications for our understanding of the brain’s processing of human speech and rhythmic input in general.

## CONCLUSION

Recently, phase entrainment has attracted researchers' attention as a potential reflection of the brain's mechanism to efficiently allocate attentional resources in time (for a recent review, see, e.g., Frey et al., 2015). Nevertheless, the periodicity of the stimulation itself complicates this interpretation, as the brain might simply follow the rhythm of its input. In this review, we presented an increasing amount of evidence that speaks against a merely passive role of neural oscillations for phase entrainment to speech sound. Instead, the brain might constantly predict the timing of relevant and irrelevant events of speech sound, including acoustic high-level features, and actively align neural oscillations so that they efficiently

boost the current locus of attention in a noisy background. Linguistic high-level features, reflecting intelligibility, might play a modulatory, and speech-specific, role by determining the behavioral consequences of phase entrainment to speech sound.

## ACKNOWLEDGMENTS

The authors thank Mitchell Steinschneider, Peter Lakatos, Daniel Pressnitzer, Alain de Cheveigné and Jesko Verhey for helpful comments and discussions. This study was supported by a Studienstiftung des deutschen Volkes (German National Academic Foundation) scholarship to BZ, and an ERC grant ("P-Cycles", number 614244) to RV.

## REFERENCES

- Adrian, E. D. (1944). Brain rhythms. *Nature* 153, 360–362. doi: 10.1038/153360a0
- Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., and Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl. Acad. Sci. U S A* 98, 13367–13372. doi: 10.1073/pnas.201400998
- Arnal, L. H. (2012). Predicting "when" using the motor system's beta-band oscillations. *Front. Hum. Neurosci.* 6:225. doi: 10.3389/fnhum.2012.00225
- Arnal, L. H., Doelling, K. B., and Poeppel, D. (2015). Delta-beta coupled oscillations underlie temporal prediction accuracy. *Cereb. Cortex* 25, 3077–3085. doi: 10.1093/cercor/bhu103
- Arnal, L. H., and Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends Cogn. Sci.* 16, 390p–398. doi: 10.1016/j.tics.2012.05.003
- Arnal, L. H., Morillon, B., Kell, C. A., and Giraud, A.-L. (2009). Dual neural routing of visual facilitation in speech processing. *J. Neurosci.* 29, 13445–13453. doi: 10.1523/JNEUROSCI.3194-09.2009
- Arnal, L. H., Wyart, V., and Giraud, A.-L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nat. Neurosci.* 14, 797–801. doi: 10.1038/nn.2810
- Besle, J., Schevon, C. A., Mehta, A. D., Lakatos, P., Goodman, R. R., McKhann, G. M., et al. (2011). Tuning of the human neocortex to the temporal dynamics of attended events. *J. Neurosci.* 31, 3176–3185. doi: 10.1523/JNEUROSCI.4518-10.2011
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb. Cortex* 10, 512–528. doi: 10.1093/cercor/10.5.512
- Buzsáki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science* 304, 1926–1929. doi: 10.1126/science.1099745
- Capilla, A., Pazo-Alvarez, P., Darriba, A., Campo, P., and Gross, J. (2011). Steady-state visual evoked potentials can be explained by temporal superposition of transient event-related responses. *PLoS One* 6:e14543. doi: 10.1371/journal.pone.0014543
- Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975–979. doi: 10.1121/1.1907229
- Coutlee, C. G., and Huettel, S. A. (2012). The functional neuroanatomy of decision making: prefrontal control of thought and action. *Brain Res.* 1428, 3–12. doi: 10.1016/j.brainres.2011.05.053
- Crosse, M. J., Butler, J. S., and Lalor, E. C. (2015). Congruent visual speech enhances cortical entrainment to continuous auditory speech in noise-free conditions. *J. Neurosci.* 35, 14195–14204. doi: 10.1523/JNEUROSCI.1829-15.2015
- Cummins, F. (2012). Oscillators and syllables: a cautionary note. *Front. Psychol.* 3:364. doi: 10.3389/fpsyg.2012.00364
- DeWitt, I., and Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proc. Natl. Acad. Sci. U S A* 109, E505–E514. doi: 10.1073/pnas.1113427109
- Ding, N., Chatterjee, M., and Simon, J. Z. (2013). Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *Neuroimage* 88C, 41–46. doi: 10.1016/j.neuroimage.2013.10.054
- Ding, N., and Simon, J. Z. (2012a). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. U S A* 109, 11854–11859. doi: 10.1073/pnas.1205381109
- Ding, N., and Simon, J. Z. (2012b). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J. Neurophysiol.* 107, 78–89. doi: 10.1152/jn.00297.2011
- Ding, N., and Simon, J. Z. (2013). Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *J. Neurosci.* 33, 5728–5735. doi: 10.1523/JNEUROSCI.5297-12.2013
- Ding, N., and Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Front. Hum. Neurosci.* 8:311. doi: 10.3389/fnhum.2014.00311
- Doelling, K. B., Arnal, L. H., Ghitza, O., and Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage* 85, 761–768. doi: 10.1016/j.neuroimage.2013.06.035
- Fiebelkorn, I. C., Foxe, J. J., Butler, J. S., Mercier, M. R., Snyder, A. C., and Molholm, S. (2011). Ready, set, reset: stimulus-locked periodicity in behavioral performance demonstrates the consequences of cross-sensory phase reset. *J. Neurosci.* 31, 9971–9981. doi: 10.1523/JNEUROSCI.1338-11.2011
- Fontolan, L., Morillon, B., Liegeois-Chauvel, C., and Giraud, A.-L. (2014). The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. *Nat. Commun.* 5:4694. doi: 10.1038/ncomms5694
- Frey, J. N., Ruhnau, P., and Weisz, N. (2015). Not so different after all: the same oscillatory processes support different types of attention. *Brain Res.* 1626, 183–197. doi: 10.1016/j.brainres.2015.02.017
- Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* 9, 474–480. doi: 10.1016/j.tics.2005.08.011
- Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622
- Fujioka, T., Trainor, L. J., Large, E. W., and Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic  $\beta$  oscillations. *J. Neurosci.* 32, 1791–1802. doi: 10.1523/JNEUROSCI.4107-11.2012
- Galambos, R., Makeig, S., and Talmachoff, P. J. (1981). A 40-Hz auditory potential recorded from the human scalp. *Proc. Natl. Acad. Sci. U S A* 78, 2643–2647. doi: 10.1073/pnas.78.4.2643
- Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Front. Psychol.* 2:130. doi: 10.3389/fpsyg.2011.00130
- Ghitza, O. (2013). The theta-syllable: a unit of speech information defined by cortical function. *Front. Psychol.* 4:138. doi: 10.3389/fpsyg.2013.00138
- Ghitza, O. (2014). Behavioral evidence for the role of cortical  $\theta$  oscillations in determining auditory channel capacity for speech. *Front. Psychol.* 5:652. doi: 10.3389/fpsyg.2014.00652

- Gilbert, C. D., and Li, W. (2013). Top-down influences on visual processing. *Nat. Rev. Neurosci.* 14, 350–363. doi: 10.1038/nrn3476
- Giraud, A.-L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517. doi: 10.1038/nn.3063
- Gomez-Ramirez, M., Kelly, S. P., Molholm, S., Sehatpour, P., Schwartz, T. H., and Foxe, J. J. (2011). Oscillatory sensory selection mechanisms during intersensory attention to rhythmic auditory and visual inputs: a human electrocorticographic investigation. *J. Neurosci.* 31, 18556–18567. doi: 10.1523/JNEUROSCI.2164-11.2011
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438. doi: 10.2307/1912791
- Green, T., Faulkner, A., and Rosen, S. (2002). Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants. *J. Acoust. Soc. Am.* 112, 2155–2164. doi: 10.1121/1.1506688
- Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., et al. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol.* 11:e1001752. doi: 10.1371/journal.pbio.1001752
- Henry, M. J., Herrmann, B., and Obleser, J. (2014). Entrained neural oscillations in multiple frequency bands comodulate behavior. *Proc. Natl. Acad. Sci. U S A* 111, 14935–14940. doi: 10.1073/pnas.1408741111
- Henry, M. J., and Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proc. Natl. Acad. Sci. U S A* 109, 20095–20100. doi: 10.1073/pnas.1213390109
- Herring, J. D., Thut, G., Jensen, O., and Bergmann, T. (2015). Attention modulates TMS-locked alpha oscillations in the visual cortex. *J. Neurosci.* 35, 14435–14447. doi: 10.1523/JNEUROSCI.1833-15.2015
- Hickok, G., Farahbod, H., and Saberi, K. (2015). The rhythm of perception: entrainment to acoustic rhythms induces subsequent perceptual oscillation. *Psychol. Sci.* 26, 1006–1013. doi: 10.1177/0956797615576533
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402. doi: 10.1038/nrn2113
- Horton, C., D'Zmura, M., and Srinivasan, R. (2013). Suppression of competing speech through entrainment of cortical oscillations. *J. Neurophysiol.* 109, 3082–3093. doi: 10.1152/jn.01026.2012
- Howard, M. F., and Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *J. Neurophysiol.* 104, 2500–2511. doi: 10.1152/jn.00251.2010
- Keitel, C., Quigley, C., and Ruhnau, P. (2014). Stimulus-driven brain oscillations in the alpha range: entrainment of intrinsic rhythms or frequency-following response? *J. Neurosci.* 34, 10137–10140. doi: 10.1523/JNEUROSCI.1904-14.2014
- Kerlin, J. R., Shahin, A. J., and Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a “cocktail party”. *J. Neurosci.* 30, 620–628. doi: 10.1523/JNEUROSCI.3631-09.2010
- Krawczyk, D. C. (2002). Contributions of the prefrontal cortex to the neural basis of human decision making. *Neurosci. Biobehav. Rev.* 26, 631–664. doi: 10.1016/s0149-7634(02)00021-0
- Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320, 110–113. doi: 10.1126/science.1154735
- Lakatos, P., Musacchia, G., O'Connell, M. N., Falchier, A. Y., Javitt, D. C., and Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron* 77, 750–761. doi: 10.1016/j.neuron.2012.11.034
- Lakatos, P., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., and Schroeder, C. E. (2009). The leading sense: supramodal control of neurophysiological context by attention. *Neuron* 64, 419–430. doi: 10.1016/j.neuron.2009.10.014
- Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., and Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J. Neurophysiol.* 94, 1904–1911. doi: 10.1152/jn.00263.2005
- Lamme, V. A., and Spekreijse, H. (2000). Modulations of primary visual cortex activity representing attentive and conscious scene perception. *Front. Biosci.* 5, D232–D243. doi: 10.2741/lamme
- Landau, A. N., and Fries, P. (2012). Attention samples stimuli rhythmically. *Curr. Biol.* 22, 1000–1004. doi: 10.1016/j.cub.2012.03.054
- Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54, 1001–1010. doi: 10.1016/j.neuron.2007.06.004
- Luo, H., Liu, Z., and Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol.* 8:e1000445. doi: 10.1371/journal.pbio.1000445
- Luo, H., and Poeppel, D. (2012). Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. *Front. Psychol.* 3:170. doi: 10.3389/fpsyg.2012.00170
- Mazzoni, A., Whittingstall, K., Brunel, N., Logothetis, N. K., and Panzeri, S. (2010). Understanding the relationships between spike rate and delta/gamma frequency bands of LFPs and EEGs using a local cortical network model. *Neuroimage* 52, 956–972. doi: 10.1016/j.neuroimage.2009.12.040
- McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0
- Mercier, M. R., Molholm, S., Fiebelkorn, I. C., Butler, J. S., Schwartz, T. H., and Foxe, J. J. (2015). Neuro-oscillatory phase alignment drives speeded multisensory response times: an electro-corticographic investigation. *J. Neurosci.* 35, 8546–8557. doi: 10.1523/JNEUROSCI.4527-14.2015
- Mesgarani, N., and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485, 233–236. doi: 10.1038/nature11020
- Mesgarani, N., Cheung, C., Johnson, K., and Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science* 343, 1006–1010. doi: 10.1126/science.1245994
- Millman, R. E., Johnson, S. R., and Prendergast, G. (2015). The role of phase-locking to the temporal envelope of speech in auditory perception and speech intelligibility. *J. Cogn. Neurosci.* 27, 533–545. doi: 10.1162/jocn\_a\_00719
- Morillon, B., Hackett, T. A., Kajikawa, Y., and Schroeder, C. E. (2015). Predictive motor control of sensory dynamics in auditory active sensing. *Curr. Opin. Neurobiol.* 31C, 230–238. doi: 10.1016/j.conb.2014.12.005
- Morillon, B., and Schroeder, C. E. (2015). Neuronal oscillations as a mechanistic substrate of auditory temporal prediction. *Ann. N Y Acad. Sci.* 1337, 26–31. doi: 10.1111/nyas.12629
- Morillon, B., Schroeder, C. E., and Wyart, V. (2014). Motor contributions to the temporal precision of auditory attention. *Nat. Commun.* 5:5255. doi: 10.1038/ncomms6255
- Nozaradan, S. (2014). Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130393. doi: 10.1098/rstb.2013.0393
- O'Connell, M. N., Barczak, A., Schroeder, C. E., and Lakatos, P. (2014). Layer specific sharpening of frequency tuning by selective attention in primary auditory cortex. *J. Neurosci.* 34, 16496–16508. doi: 10.1523/JNEUROSCI.2055-14.2014
- O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., et al. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cereb. Cortex* 25, 1697–1706. doi: 10.1093/cercor/bht355
- Obleser, J., Herrmann, B., and Henry, M. J. (2012). Neural oscillations in speech: don't be enslaved by the envelope. *Front. Hum. Neurosci.* 6:250. doi: 10.3389/fnhum.2012.00250
- Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., and Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Curr. Biol.* 25, 1649–1653. doi: 10.1016/j.cub.2015.04.049
- Peelen, M. V., and Kastner, S. (2014). Attention in the real world: toward understanding its neural basis. *Trends Cogn. Sci.* 18, 242–250. doi: 10.1016/j.tics.2014.02.004
- Peelle, J. E., and Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Front. Psychol.* 3:320. doi: 10.3389/fpsyg.2012.00320
- Peelle, J. E., Gross, J., and Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb. Cortex* 23, 1378–1387. doi: 10.1093/cercor/bhs118
- Peelle, J. E., and Sommers, M. S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex* 68, 169–181. doi: 10.1016/j.cortex.2015.03.006
- Perrodin, C., Kayser, C., Logothetis, N. K., and Petkov, C. I. (2015). Natural asynchronies in audiovisual communication signals regulate neuronal multisensory interactions in voice-sensitive cortex. *Proc. Natl. Acad. Sci. U S A* 112, 273–278. doi: 10.1073/pnas.1412817112

- Poeppl, D., Emmorey, K., Hickok, G., and Pyykkänen, L. (2012). Towards a new neurobiology of language. *J. Neurosci.* 32, 14125–14131. doi: 10.1523/JNEUROSCI.3244-12.2012
- Power, A. J., Mead, N., Barnes, L., and Goswami, U. (2013). Neural entrainment to rhythmic speech in children with developmental dyslexia. *Front. Hum. Neurosci.* 7:777. doi: 10.3389/fnhum.2013.00777
- Rajkai, C., Lakatos, P., Chen, C.-M., Pincze, Z., Karmos, G., and Schroeder, C. E. (2008). Transient cortical excitation at the onset of visual fixation. *Cereb. Cortex* 18, 200–209. doi: 10.1093/cercor/bhm046
- Rimmele, J. M., Zion Golumbic, E., Schröger, E., and Poeppl, D. (2015). The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. *Cortex* 68, 144–154. doi: 10.1016/j.cortex.2014.12.014
- Romo, R., and de Lafuente, V. (2013). Conversion of sensory signals into perceptual decisions. *Prog. Neurobiol.* 103, 41–75. doi: 10.1016/j.pneurobio.2012.03.007
- Ruhnau, P., Hauswald, A., and Weisz, N. (2014). Investigating ongoing brain oscillations and their influence on conscious perception - network states and the window to consciousness. *Front. Psychol.* 5:1230. doi: 10.3389/fpsyg.2014.01230
- Rushworth, M. F. S., Kolling, N., Sallet, J., and Mars, R. B. (2012). Valuation and decision-making in frontal cortex: one or many serial or parallel systems? *Curr. Opin. Neurobiol.* 22, 946–955. doi: 10.1016/j.conb.2012.04.011
- Schreiber, T. (2000). Measuring information transfer. *Phys. Rev. Lett.* 85, 461–464. doi: 10.1103/physrevlett.85.461
- Schroeder, C. E., and Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci.* 32, 9–18. doi: 10.1016/j.tins.2008.09.012
- Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends Cogn. Sci.* 12, 106–113. doi: 10.1016/j.tics.2008.01.002
- Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., and Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Curr. Opin. Neurobiol.* 20, 172–176. doi: 10.1016/j.conb.2010.02.010
- Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123(Pt. 12), 2400–2406. doi: 10.1093/brain/123.12.2400
- Simon, J. Z. (2015). The encoding of auditory objects in auditory cortex: insights from magnetoencephalography. *Int. J. Psychophysiol.* 95, 184–190. doi: 10.1016/j.ijpsycho.2014.05.005
- Song, K., Meng, M., Chen, L., Zhou, K., and Luo, H. (2014). Behavioral oscillations in attention: rhythmic  $\alpha$  pulses mediated through  $\theta$  band. *J. Neurosci.* 34, 4837–4844. doi: 10.1523/JNEUROSCI.4856-13.2014
- Spaak, E., de Lange, F. P., and Jensen, O. (2014). Local entrainment of  $\alpha$  oscillations by visual stimuli causes cyclic modulation of perception. *J. Neurosci.* 34, 3536–3544. doi: 10.1523/JNEUROSCI.4385-13.2014
- Stefanics, G., Hangya, B., Hernádi, I., Winkler, I., Lakatos, P., and Übert, I. (2010). Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *J. Neurosci.* 30, 13578–13585. doi: 10.1523/JNEUROSCI.0703-10.2010
- Steinschneider, M., Nourski, K. V., and Fishman, Y. I. (2013). Representation of speech in human auditory cortex: is it special? *Hear. Res.* 305, 57–73. doi: 10.1016/j.heares.2013.05.013
- Thorne, J. D., and Debener, S. (2014). Look now and hear what's coming: on the functional role of cross-modal phase reset. *Hear. Res.* 307, 144–152. doi: 10.1016/j.heares.2013.07.002
- Thut, G., Miniussi, C., and Gross, J. (2012). The functional importance of rhythmic activity in the brain. *Curr. Biol.* 22, R658–R663. doi: 10.1016/j.cub.2012.06.061
- Thut, G., Schyns, P. G., and Gross, J. (2011). Entrainment of perceptually relevant brain oscillations by non-invasive rhythmic stimulation of the human brain. *Front. Psychol.* 2:170. doi: 10.3389/fpsyg.2011.00170
- van Wassenhove, V., Grant, K. W., and Poeppl, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. U S A* 102, 1181–1186. doi: 10.1073/pnas.0408949102
- Vanrullen, R., and Dubois, J. (2011). The psychophysics of brain rhythms. *Front. Psychol.* 2:203. doi: 10.3389/fpsyg.2011.00203
- VanRullen, R., Zoefel, B., and Ilhan, B. (2014). On the cyclic nature of perception in vision versus audition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130214. doi: 10.1098/rstb.2013.0214
- Weisz, N., Wühle, A., Monittola, G., Demarchi, G., Frey, J., Popov, T., et al. (2014). Prestimulus oscillatory power and connectivity patterns predispose conscious somatosensory perception. *Proc. Natl. Acad. Sci. U S A* 111, E417–E425. doi: 10.1073/pnas.1317267111
- Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., and Johnsrude, I. S. (2012). Effortful listening: the processing of degraded speech depends critically on attention. *J. Neurosci.* 32, 14010–14021. doi: 10.1523/JNEUROSCI.1528-12.2012
- Zion Golumbic, E., Cogan, G. B., Schroeder, C. E., and Poeppl, D. (2013a). Visual input enhances selective speech envelope tracking in auditory cortex at a “cocktail party”. *J. Neurosci.* 33, 1417–1426. doi: 10.1523/JNEUROSCI.3675-12.2013
- Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., et al. (2013b). Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. *Neuron* 77, 980–991. doi: 10.1016/j.neuron.2012.12.037
- Zion Golumbic, E. M., Poeppl, D., and Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: a behavioral and neural perspective. *Brain Lang.* 122, 151–161. doi: 10.1016/j.bandl.2011.12.010
- Zoefel, B., and Heil, P. (2013). Detection of near-threshold sounds is independent of EEG phase in common frequency bands. *Front. Psychol.* 4:262. doi: 10.3389/fpsyg.2013.00262
- Zoefel, B., and Sokoliuk, R. (2014). Investigating the rhythm of attention on a fine-grained scale: evidence from reaction times. *J. Neurosci.* 34, 12619–12621. doi: 10.1523/JNEUROSCI.2134-14.2014
- Zoefel, B., and VanRullen, R. (2015a). EEG oscillations entrain their phase to high-level features of speech sound. *Neuroimage* 124, 16–23. doi: 10.1016/j.neuroimage.2015.08.054
- Zoefel, B., and VanRullen, R. (2015b). Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. *J. Neurosci.* 35, 1954–1964. doi: 10.1523/JNEUROSCI.3484-14.2015

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Zoefel and VanRullen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.