Neural tracking of continuous acoustics: properties, speech-specificity and open questions

Benedikt Zoefel, Anne Kösem

Received: 27 June 2023 | Revised: 17 November 2023 | Accepted: 22 November 2023

DOI: 10.1111/ejn.16221

REVIEW ARTICLE

EJN European Journal of Neuroscience FENS WILEY

Neural tracking of continuous acoustics: properties, speech-specificity and open questions

Benedikt Zoefel[1,2] | **Anne Kösem**[3]

[1]Centre de Recherche Cerveau et Cognition (CerCo), CNRS UMR 5549, Toulouse, France

[2]Université de Toulouse III Paul Sabatier, Toulouse, France

[3]Lyon Neuroscience Research Center (CRNL), INSERM U1028, Bron, France

Correspondence
Benedikt Zoefel, Centre de Recherche Cerveau et Cognition (CerCo), CNRS UMR 5549, Toulouse, France.
Email: benedikt.zoefel@cnrs.fr

Anne Kösem, Lyon Neuroscience Research Center (CRNL), INSERM U1028, Bron, France.
Email: anne.kosem@inserm.fr

Funding information
Agence Nationale de la Recherche, Grant/Award Numbers: ANR-21-CE37-0002, ANR-21-CE37-0003; Fondation Pour l'Audition, Grant/Award Number: FPA-RD-2021-10

Edited by: Dr. Edmund Lalor

Abstract

Human speech is a particularly relevant acoustic stimulus for our species, due to its role of information transmission during communication. Speech is inherently a dynamic signal, and a recent line of research focused on neural activity following the temporal structure of speech. We review findings that characterise neural dynamics in the processing of continuous acoustics and that allow us to compare these dynamics with temporal aspects in human speech. We highlight properties and constraints that both neural and speech dynamics have, suggesting that auditory neural systems are optimised to process human speech. We then discuss the speech-specificity of neural dynamics and their potential mechanistic origins and summarise open questions in the field.

KEYWORDS
dynamical systems, entrainment, neural oscillations, neural tracking, speech perception

1 | INTRODUCTION

Human speech is possibly the most relevant acoustic stimulus for our species, at least one we are continuously exposed to since birth. The fact that humans use speech to communicate assigns it a distinct role among the multitude of sounds we are confronted with. Naturally, the question how speech is processed in the brain has a long tradition in research (Carbonell & Lotto, 2014; Galantucci et al., 2006; Moore, 2000; Steinschneider et al., 2013) and produced important results. Studies on brain-function mapping have revealed a complex functional neuroanatomy of speech that comprises temporal, parietal and frontal regions of the cortex (Hickok & Poeppel, 2007, 2016; Rauschecker & Scott, 2009) as well as subcortical contributions (Kotz & Schwartze, 2010). Speech, however, is a dynamic signal and carries relevant acoustic and linguistic information in the temporal domain. Neural analysis of speech therefore requires continuous information processing at different time scales in parallel, from relatively brief phonemes to slower sentential information. Research in the neurobiology of speech has started to address this facet

Abbreviations: AM, amplitude-modulated; BOLD, blood-oxygen level dependent; EEG, electroencephalography; fMRI, functional magnetic resonance imaging; MEG, magnetoencephalography; STS, superior temporal sulcus; tACS, transcranial alternating current stimulation.

of speech processing by putting focus on the temporal aspect of neural activity. Research on temporarily resolved neural responses (i.e. *neural dynamics*) to speech is, of course, not new. However, this research has traditionally examined responses to isolated words, sometimes embedded in continuous speech, or their mismatch with expectations (Boddy, 1981; Kutas & Federmeier, 2011; Sohoglu et al., 2012). In this review, we focus on neural dynamics that are related to the temporal properties that speech itself possesses. As we describe, the role of neural dynamics for continuous speech processing and the challenges that go along with such a dynamic signal begin to be understood. In particular, neural dynamics have been shown to follow the temporal structure of spoken utterances at distinct time scales. This phenomenon is sometimes described as 'neural tracking' or 'neural entrainment in the broad sense' (Obleser & Kayser, 2019). Neural dynamics track a large variety of acoustic inputs, including simple tone sequences (Lakatos et al., 2008), beats (Nozaradan et al., 2012) and music (Doelling & Poeppel, 2015). However, entrained neural dynamics are also a fundamental part of prominent models of speech processing where they are thought to be necessary for successful speech comprehension and to contribute to the parsing of continuous speech into relevant linguistic units (Giraud & Poeppel, 2012).

A long-standing question revolves around the mechanistic origins of neural tracking and, in particular, whether it involves endogenous brain rhythms (Haegens & Zion Golumbic, 2018; Lakatos et al., 2019; Obleser & Kayser, 2019; Zoefel, ten Oever, & Sack, 2018). In this review, we first step away from this debate and focus, with no assumption on the underlying neural implementation, on dynamic properties of neural activity that are relevant for the processing of temporal patterns in speech. In a first part, we do not distinguish neural dynamics involved in the processing of sound from those that specifically process speech. Rather, we describe general properties and constraints of auditory neural dynamics and discuss in how far they might relate to challenges and demands that the dynamic complexity of speech imposes onto the neural system. In a second part, we summarise to what extent these dynamic constraints and properties are more pronounced or different during the processing of speech as compared to that of other auditory signals. For this purpose, we summarise results that allow us to contrast neural dynamics during speech with those observed during other sounds. This contrast already makes it clear that the term 'neural dynamics' entails neural processes that come from different areas and might achieve different tasks. Similarities (first part) and differences (second part) between these dynamics are the basis for what we describe here. In the last part of the review, we consider how far neural oscillatory models of speech processing can explain the described effects and propose testable hypotheses that result from this assumption. We conclude with open questions for research on neural dynamics and speech processing.

2 | PROPERTIES OF NEURAL DYNAMICS IN THE AUDITORY SYSTEM (AND BEYOND) AND HOW THEY RELATE TO THOSE OF HUMAN SPEECH

Continuous human speech has rapid and complex temporal dynamics and therefore requires fast and efficient temporal processing. In this chapter, we focus on properties of neural dynamics, particularly in the auditory system, that seem ideal to process temporal patterns in human speech. We discuss how these properties, along with their limits and constraints, might relate to the temporal characteristics of speech.

We here note that the term 'tracking' has become ubiquitous in research exploring neural responses to continuous speech by regressing these responses on various (acoustic and linguistic) features of speech. This research has revealed critical insights into the time course of neural dynamics in speech processing and demonstrated speech tracking on multiple levels (Brodbeck et al., 2018; Broderick et al., 2021; Weissbart et al., 2020). However, whereas this approach can assess the temporal evolution of neural responses evoked by selected features, it does not necessarily consider temporal properties of the speech itself. Referring to other recent reviews on regression-based speech tracking (Brodbeck & Simon, 2020; Sassenhagen, 2019), we therefore restrict our overview to measures that link neural dynamics and temporal aspects of speech more directly. One example is speech-brain coherence that quantifies the phase-locking between neural dynamics and speech at the *same frequencies*.

2.1 | Preferred dynamics, eigenfrequency (Figure 1)

Dynamic systems, including neural ones, often have an *eigenfrequency*, that is, a frequency they operate at in the absence of input, or a stimulus rate they most strongly respond to. Most studies point to two distinct *eigenfrequency* ranges for the auditory system: the delta–theta range (~2–8 Hz) and the gamma range (~30–40 Hz) (Boemio et al., 2005; Giraud et al., 2007). Human perceptual sensitivity to acoustic spectro-temporal modulations is highest between 2 and 5 Hz (Chi et al., 1999; Edwards & Chang, 2013). Brain imaging revealed that blood-oxygen level dependent (BOLD) responses to

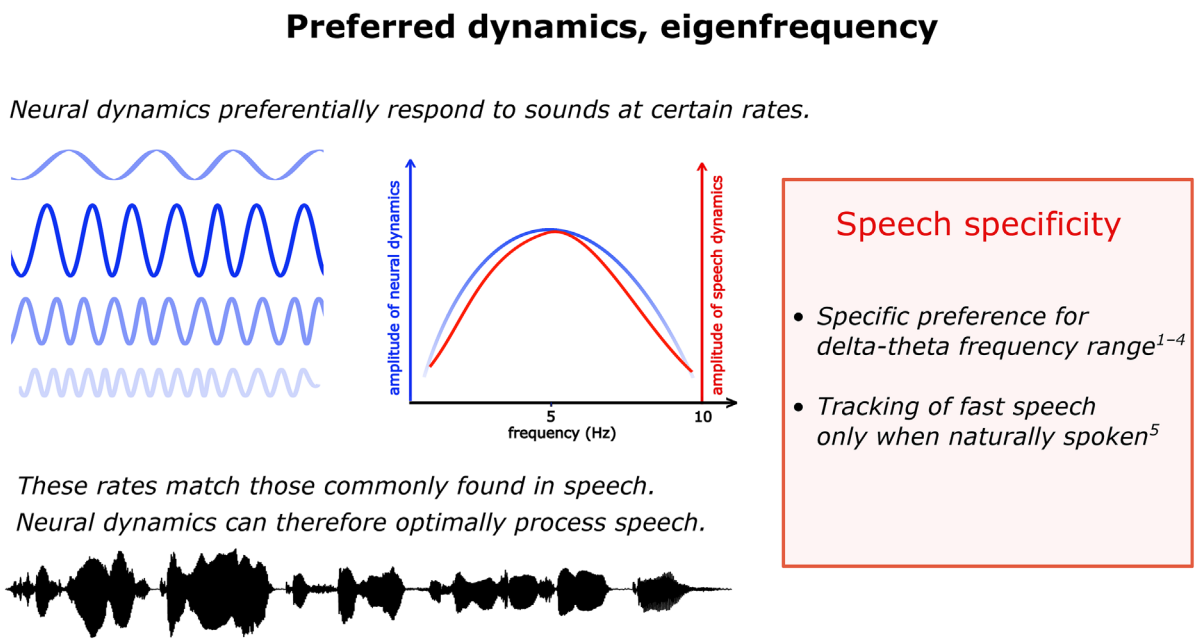

FIGURE 1 Preferred dynamics, eigenfrequency. References in speech-specificity box refer to: 1. Ding & Simon, 2014; 2. Etard & Reichenbach, 2019; 3. Keitel et al., 2018; 4. Molinaro & Lizarazu, 2018; 5. Zuk et al., 2021; 6. Hincapié Casas et al., 2021.

amplitude modulated (AM) sounds are strongest if these are presented at 4–5 Hz (Giraud et al., 2000; Tanaka et al., 2000). Rhythmic AM sounds also give rise to rhythmic fluctuations in auditory sensitivity that outlast the stimulus, but only at rates between ∼2 and 8 Hz (Farahbod et al., 2020; Hickok et al., 2015; L'Hermite & Zoefel, 2023). Non-rhythmic acoustic stimuli, such as the onset of broadband noise, produce similar fluctuations in neural dynamics and auditory sensitivity in the delta–theta range, although the exact frequency remains unclear (∼1–2 Hz in Kayser, 2019; ∼5 Hz in Teng et al., 2018; ∼6–8 Hz in Ho et al., 2017). Studies investigating neural tracking of higher stimulus rates reported an additional preference in the gamma range. Importantly, neural dynamics follow acoustic rhythms most reliably when these are presented at theta and gamma rates, while rates in between do not generate reliable tracking responses (Galambos et al., 1981; Giroud et al., 2020; Teng et al., 2017; Teng & Poeppel, 2020; Zaehle et al., 2010). This finding suggests the existence of two distinct eigenfrequency ranges. Nevertheless, gamma effects seem overall less clear than delta–theta ones, possibly due to an attenuation of higher frequencies in electroencephalography (EEG). Together, there is converging evidence that auditory dynamics 'prefer' certain stimulus rates and respond most readily to them.

Speech is a dynamic signal that has its own *eigenfrequencies*, that is, it conveys information over distinct time scales. This leads to linguistic 'building blocks' of speech, such as phonemes, syllables and words. Within each of these elements, the rate of information transmission is relatively stable. For example, phonemic features are typically of 20–50 ms duration, thus fluctuating at a rate of ∼20–50 Hz (Ghitza, 2011). Phonemes compose the syllables, which have a mean duration of 200–250 ms, corresponding to an average rate of 4–5 Hz (Greenberg, 1999; Strauß & Schwartz, 2017). In most languages, words are spoken at a rate of 100–200 words per minute, that is, at 1.5–3 Hz (Carver, 1973), although a systematic analysis of their regularity is still lacking. The acoustic speech signal also entails regular temporal structure at distinct time scales. Across languages, human speech contains broadband amplitude modulations that are strongest around 3–5 Hz (Ding et al., 2017; Varnet et al., 2017), roughly corresponding to the spoken syllabic rate (Greenberg, 1999). It also contains modulations in frequency (pitch) (Teoh et al., 2019) that the auditory system can convert to amplitude modulations (Ghitza et al., 2013). Stress patterns or intonational units, carrying prosodic information, can entail regular patterns that fluctuate around ∼1 Hz (Inbar et al., 2020). A point of caution, however, is that prosodic information shows considerable variability and its rhythmicity is not well investigated (Stehwien & Meyer, 2022; Tilsen & Arvaniti, 2013). Together, it is clear that the delta–theta range in human speech, driven by amplitude modulations and perceptualised as the syllable, is both the most pronounced and best explored rhythm in human speech.

This match between neural auditory eigenfrequencies and those of speech might explain some perceptual effects. Sounds that are amplitude-modulated at the delta–theta rate produce a distinct perceptual category (termed 'fluctuations') that disappears at faster or slower rates (Edwards & Chang, 2013). This observation suggests

that the tuning to delta-theta rates, common to both speech and auditory neural dynamics, also has a categorical impact on auditory perception. This link between speech and neural dynamics is also supported by studies reporting that blind listeners can understand speech at higher syllabic rates than a sighted population (Hertrich et al., 2013). This effect has been suggested to originate from a neural 'recycling' of visual areas for auditory processes (Van Ackeren et al., 2018). The *eigenfrequency* of primary visual regions (~10 Hz; Herrmann, 2001) is higher than the typical syllabic rate; if visual cortex is recruited during speech processing in the blind, then this might also lead to faster auditory eigenfrequencies and explain why blind people can understand faster speech.

2.2 | Constrained temporal flexibility (Figure 2)

Despite having 'preferred' frequencies, neural dynamics are not rigid and 'track' different acoustic rates in both non-speech (Doelling & Poeppel, 2015; Lakatos et al., 2008) and speech stimuli (Ahissar et al., 2001; Kösem et al., 2018), even when the stimulus is not perfectly isochronous (Doelling et al., 2022; Doelling & Assaneo, 2021; Kayser et al., 2015). Studies using transcranial brain stimulation to manipulate how neural dynamics adapt to acoustic rhythms showed that neural tracking causally modulates auditory and speech perception, an effect that has also been observed at various stimulation rates (Keshavarzi et al., 2020, 2021; Kösem et al., 2020; Riecke et al., 2015, 2018; van Bree et al., 2021; Wilsch et al., 2018; Zoefel et al., 2020; Zoefel, Archer-Boyd, & Davis, 2018). Importantly however, neural tracking has its limits: Neural dynamics fail to track the acoustic rhythm if it is too slow or too fast. These limits are defined by the system's *eigenfrequency* range: Most of the neural effects described in the previous section were observed for the delta–theta range but disappear if the stimulus is too fast or slow (Farahbod et al., 2020; Galambos et al., 1981; L'Hermite & Zoefel, 2023; Teng et al., 2017; Teng & Poeppel, 2020; Zaehle et al., 2010; but see Hertrich et al., 2012; Nourski et al., 2009, for neural responses that persist beyond the theta range). This suggests that neural dynamics are flexible but constrained by their *eigenfrequency*.

Human speech, despite having distinct temporal structure, also entails temporal variability in each of its constituents (Ramus et al., 1999). First, the rate of syllables and sentential phrases can vary as function of language (Coupé et al., 2019; Tilsen & Arvaniti, 2013; Varnet et al., 2017), speaker (Tilsen & Arvaniti, 2013), emotional state (Sobin & Alpert, 1999) and other factors. The mean syllabic rate of 4–5 Hz, common to most (if not all) languages (Ding et al., 2017), can result from averaging faster and slower syllables, especially in stress-timed languages (Strauß & Schwartz, 2017). However, variability in speech dynamics is structured and constrained by the time scales described above (Section 2.1). For example, although the syllabic rate is variable, it is rarely slower than 2 Hz or faster than 8 Hz. Thus, similar to neural dynamics, the temporal variability of each building block of speech (e.g. phrase, syllable, phoneme) is

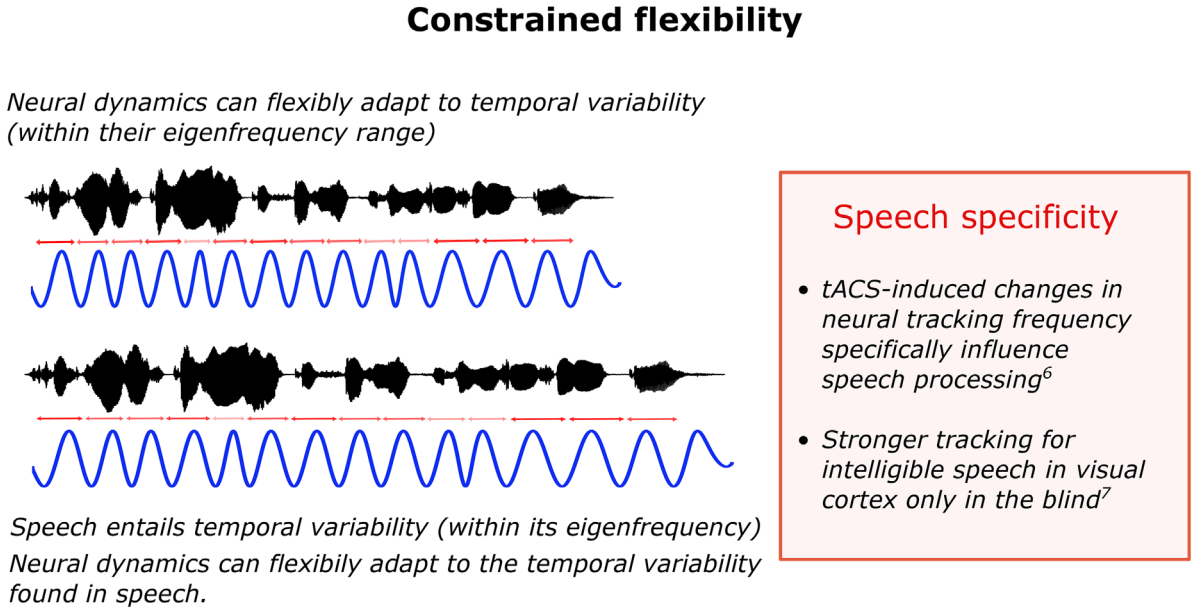

FIGURE 2 Constrained temporal flexibility. References in speech-specificity box refer to: 7. Zoefel, Archer-Boyd, & Davis, 2018; 8. Van Ackeren et al., 2018.

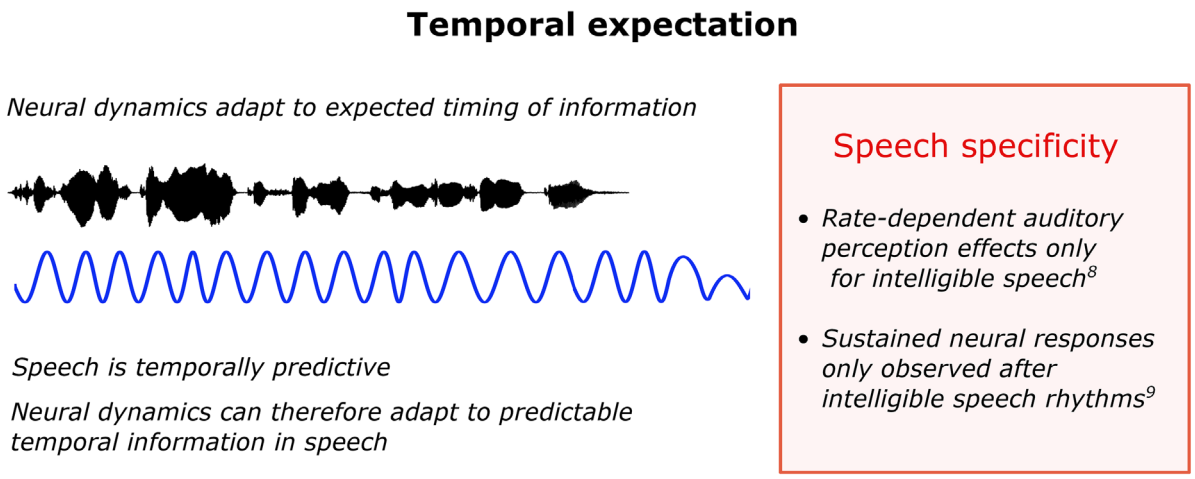

FIGURE 3 Temporal expectation. References in speech-specificity box refer to: 8. Pitt et al., 2016; 9. van Bree et al., 2021.

constrained to its typical (*eigenfrequency*) range. Indeed, speech understanding drops if word rate exceeds 4–5 Hz (Carver, 1973), or when the syllabic rate is above ~8–10 Hz (Ahissar et al., 2001; Ghitza & Greenberg, 2009; Hincapié Casas et al., 2021). Interestingly, comprehension of time-compressed, unintelligible speech is recovered if silent gaps are introduced between syllables (without slowing the time-compressed syllables themselves), suggesting that the restoration of a typical syllabic rate is key to successful speech perception (Ghitza & Greenberg, 2009).

Together, auditory neural dynamics show flexibility when it is most useful (within the *eigenfrequencies* of speech) but necessarily otherwise (outside of those ranges). This observation suggests that neural dynamics, particularly in the auditory system, are designed to cope with the temporal variability in the information they are exposed to.

2.3 | Temporal expectation (Figure 3)

To make sense of the world, the brain generates temporal predictions to anticipate future events (Friston, 2019). This function is of particular relevance for a modality confronted with a rapid stream of incoming information, such as the auditory one. Indeed, it has been shown that auditory perception is modulated by the temporal predictability of its target, in particular in the context of rhythmic scenarios. Sounds are more likely to be detected or more accurately perceived when they are presented at the beat of a preceding rhythm (Jones et al., 2002; Lawrance et al., 2014; ten Oever et al., 2014), a finding that is fundamental for the theory of 'auditory dynamic attending' (Bauer et al., 2015; Large & Jones, 1999). In line with these perceptual effects, an anticipatory adjustment of neural dynamics to expected information has been hypothesised and described in rhythmic (Kösem & van Wassenhove, 2017; Lakatos et al., 2013) and non-rhythmic scenarios (Bonnefond & Jensen, 2012; Breska & Deouell, 2017; Herbst & Obleser, 2017). This adjustment is often seen as a mechanism that aligns neural resources to expected upcoming events so that these are optimally processed (Schroeder & Lakatos, 2009). In line with this assumption, other studies have shown that the neural dynamics that track auditory rhythms are sustained, that is, neural dynamics keep fluctuating at the rhythm of the stimulus even when it stops (Bouwer et al., 2023; Lakatos et al., 2013; van Bree et al., 2021) or despite a change of temporal properties of the acoustic stimulus (Kösem et al., 2018; Lenc et al., 2020). These neural 'echoes' are also seen in corresponding perceptual data, changing rhythmically after a rhythmic acoustic stimulus (Saberi & Hickok, 2023). This effect is observed for acoustic rhythms between 2 and 8 Hz (Farahbod et al., 2020; L'Hermite & Zoefel, 2023), suggesting an involvement of neural dynamics with similar constrained temporal flexibility as described above. A recent study suggested similar echoes produced by speech prosody at lower frequencies, confirming their relevance for speech processing (Lamekina & Meyer, 2023). Together, neural echoes can be assumed to reflect anticipation that was induced by the rhythmicity of the stimulus and demonstrate temporal expectation in neural and perceptual dynamics.

Despite some variability, the timing of human speech is not random. The *average* timing of its constituents is predictable, as each of them possesses a typical rate (an *eigenfrequency*). Here, the syllabic rate (in the delta–theta range) is arguably the most stable in time (see Section 2.1). Beyond the average, the temporal variability itself can also be predictable. Across languages, a slowdown in rate is a robust predictor of a noun to be spoken

(Seifart et al., 2018). The duration of a syllable can also predict that of neighbouring ones (Greenberg, 1999; Greenberg et al., 2003; Strauß & Schwartz, 2017; but see Jadoul et al., 2016). Moreover, the variability of durational cues in speech can influence speech understanding throughout language development. Adults and babies are able to distinguish languages only based on speech rhythm properties (Nazzi et al., 1998; White et al., 2012). These variations in timing are therefore an acoustic feature that can be used for temporal predictions.

The fact that speech is predictable is nicely illustrated by various perceptual effects that link speech properties with neural ones. For example, speech perception is influenced by preceding speech rate so that some words are not perceived if the surrounding speech is pronounced at a fast or slow rate (Dilley & Pitt, 2010). Vowels can be perceived as short or long, depending on the rate of preceding speech, and this can alter the meaning of words in certain contexts (Bosker, 2017; Kösem et al., 2018). Interestingly, this effect is correlated with the neural echoes described in the previous paragraph: Kösem et al. (2018) showed that neural dynamics at a frequency that corresponds to the rate of a presented speech stimulus persists when the latter changes its rate and that this 'echo' biases the perception of an ambiguous syllable. Humans are also strikingly efficient in anticipating their turn in a conversation (Levinson, 2016). This anticipatory effect might involve a network of brain regions specialised for turn-taking in speech (Castellucci et al., 2022) and indicates that we continuously predict the end of the turn of our conversation partner.

Finally, the duration of spoken words is linked to their semantic predictability. There is an inverse correlation between the duration of spoken words and word frequency (Ten Oever et al., 2022). The onsets of words are also influenced by the sentential context, where they are slower (i.e. larger intervals between words), the less predictable a word is (Ten Oever et al., 2022). Perhaps as a consequence, speech can be processed better when spoken naturally. Adults understand speech in noise better when spoken at a natural rate, as compared to when it is made artificially rhythmic or spoken at an unnatural rhythm (Aubanel & Schwartz, 2020); neural tracking is stronger in response to naturally spoken fast speech, as compared to normal speech that has been accelerated (i.e. to a signal has a temporal structure that is unnatural for a fast speaking rate) (Hincapié Casas et al., 2021).

Human speech is not the only stimulus that is temporarily predictable. But given the rapid and complex temporal dynamics of the speech signal, it can only be processed efficiently with neural dynamics that can rapidly adapt to the expected timing of information. The fact that we seem to possess such adaptable neural dynamics again suggests that these meet the requirements imposed by dynamics of speech.

2.4 | Hierarchical structure (Figure 4)

Neural dynamics can follow abstract, structural features of an acoustic stimulus. Dynamics in primary auditory cortex of non-human primates delineate the perceived parsing of repetitive patterns in sounds (Barczak et al., 2018). In humans, neural activity aligns to higher-level structure in musical stimuli, such as when participants are asked to imagine a beat (Nozaradan et al., 2012), or when they detect changes in melodic sequences (Baltzell et al., 2019). In speech research, early studies reported speech-aligned neural activity on the syllable level (Ahissar et al., 2001; Luo & Poeppel, 2007; Peelle et al., 2013). Although this effect seems to include responses to sharp acoustic onsets (Doelling et al., 2014; Oganian et al., 2023), other research found that neural dynamics can track various structural or 'higher-level' features of speech (for reviews, see Ding & Simon, 2014; Zoefel & VanRullen, 2015a). Later, Ding et al. (2016) showed that, only when participants comprehend speech and are therefore able to parse it into various linguistic elements (e.g. phrases), brain responses follow the rate of these higher-level structures.

Structure is omnipresent in human speech which combines smaller units, such as syllables or words, into higher-level structures such as phrases or sentences. The ability of neural dynamics to follow the hierarchical structure of a stimulus might therefore have evolved from the necessity to do so in order to successfully comprehend speech. Several theoretical frameworks assume such a link, considering neural dynamics a 'tool' to parse speech into its various building blocks (e.g. Ghitza, 2013; Giraud & Poeppel, 2012).

We note that, despite the prevalence of hierarchical linguistic models in the corresponding literature (Ding et al., 2016; Giraud & Poeppel, 2012; Kazanina & Tavano, 2023), other models exist that do not assume such a hierarchy. Beyond the scope of the current review, we nevertheless refer to some of these (Frank et al., 2012; Lewis et al., 2006) for a more balanced overview of the current state of the art.

2.5 | Cross-modality and sensory-motor interactions

Most neural dynamics can be influenced by more than one (sensory-motor) modality (Ghazanfar & Schroeder, 2006). Activity in auditory regions does not only adapt to acoustic

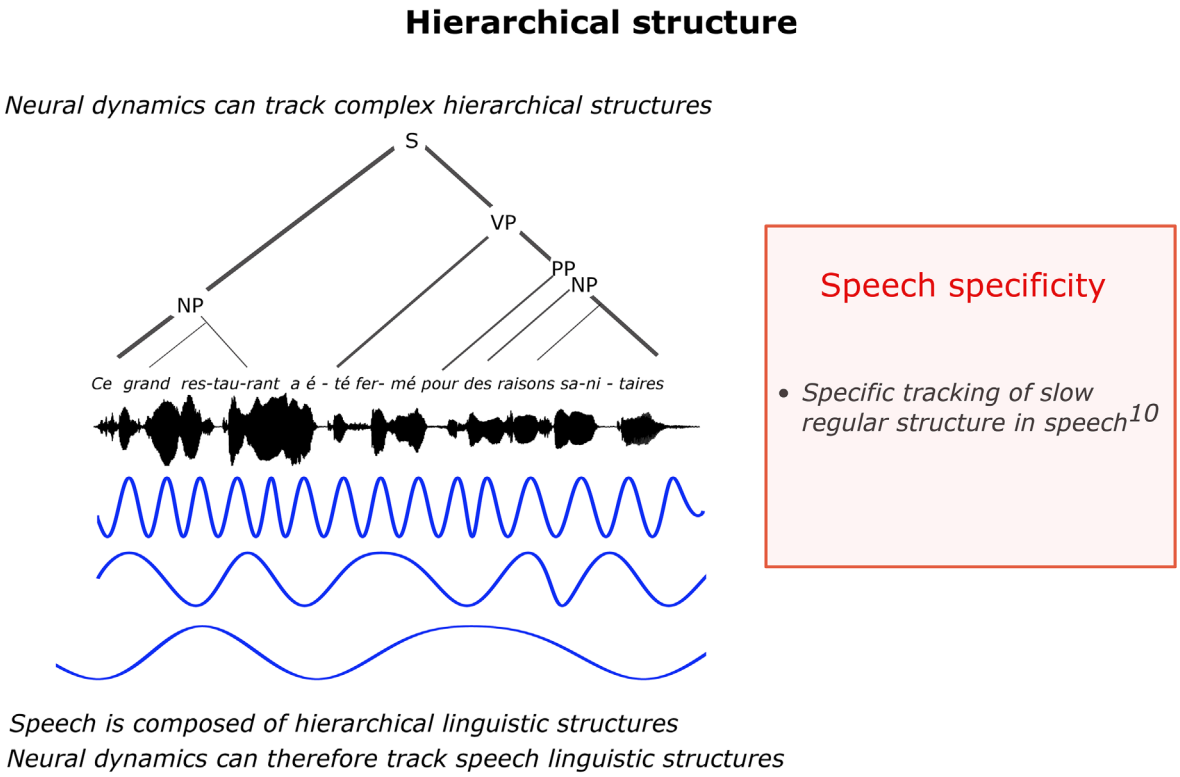

FIGURE 4 Hierarchical structure. References in speech-specificity box refer to: 10. Zuk et al., 2021.

stimuli but to input from other modalities as well, including simple visual rhythms (Besle et al., 2011; Kösem et al., 2014; Lakatos et al., 2008) or input from the motor system (Assaneo & Poeppel, 2018; Morillon et al., 2014).

Human speech is in most situations a cross-modal phenomenon. In face-to-face interactions, we see the other person move their lips when they talk. These visual speech cues usually precede acoustic information by tens to a hundred milliseconds, depending on the spoken utterance (Schwartz & Savario, 2014). Facial expressions, as well as beat gestures and semantic gestures, emphasise content and further contribute to speech understanding. For instance, congruent facial movements and gesture improve speech comprehension in noisy environments (Drijvers & Özyürek, 2017; Helfer, 1997).

The similar cross-modal organisation of neural dynamics and speech can lead to various perceptual phenomena. Most notably, the fact that visual speech cues precede acoustic information in speech makes the former a reliable cue to anticipate the latter, and visual information therefore influences the processing and interpretation of speech. In line with this assumption, neural dynamics in both visual and auditory regions track lip movements, even when presented without the accompanying sounds (Bourguignon et al., 2020; Giordano et al., 2017; Park et al., 2016, 2018). It has been proposed that visual cues reset auditory delta–theta dynamics to prepare them for upcoming acoustic information (Biau et al., 2021; Mégevand et al., 2020; Thorne & Debener, 2014). Audiovisual speech produces shorter latencies in neural responses than auditory-only speech (van Wassenhove et al., 2005), and the presentation of distinct acoustic and visual consonantal information can lead to the percept of a third consonant (Mcgurk & Macdonald, 1976). There is also evidence that we perceive acoustic and visual information to be synchronous when the latter *precedes* the former, an effect that might reflect the system's tuning to temporal statistics of human speech (Freeman et al., 2013; van Wassenhove et al., 2007). Finally, we highlight a recent study reporting that coupling between auditory and motor regions is most reliable when words are spoken at 4.5 Hz (Assaneo & Poeppel, 2018). This result does not only illustrate cross-modality of neural dynamics but reveals an *eigenfrequency* of auditory-motor synchronisation that suggests once more an optimisation to process human speech.

3 | SPEECH-SPECIFICITY OF NEURAL DYNAMICS

Some brain regions respond more readily to human speech than to other sounds, and increasingly so at higher levels of the auditory hierarchy (Landemard et al., 2021; Mesgarani et al., 2014; Scott et al., 2000). The existence of this speech-specific pathway (Saur et al., 2008;

Scott et al., 2000), together with the close match between temporal properties of speech and neural dynamics, might lead to the assumption that neural dynamics during speech differ from those observed during other sounds. In this section, we summarise findings that certain properties of neural dynamics only appear—or at least, they are particularly prominent—when confronted with intelligible speech, but not with other sounds.

3.1 | Intelligible speech does and does not produce stronger neural tracking responses than other sounds

As described above, neural dynamics follow the temporal evolution of auditory input irrespective of its complexity and have been shown to track speech, pure tones and various other non-speech stimuli (Barczak et al., 2018; Doelling & Poeppel, 2015; Lakatos et al., 2019; Nozaradan et al., 2012; Obleser & Kayser, 2019). Tracking of human speech is not restricted to human listeners and can also be observed in non-human primates (Zoefel et al., 2017). The magnitude of neural tracking (or entrainment) varies with certain properties of the auditory stimulus. Several studies have reported that a reduction in spectral detail of a speech stimulus—to a degree that makes it unintelligible—also reduces neural tracking, even if the broadband amplitude envelope remains unchanged (Meng et al., 2021; Molinaro & Lizarazu, 2018; Peelle et al., 2013). Temporal reversal does not only make speech unintelligible; it also attenuates neural dynamics aligned to it (Gross et al., 2013; Park et al., 2015). Reduced neural tracking of speech is also observed when acoustic edges in speech are removed (Doelling et al., 2014; Oganian & Chang, 2019) or when background noise or distracting speech signals are added (Ding & Simon, 2013; Rimmele et al., 2015; Zion Golumbic et al., 2013; Zoefel & VanRullen, 2015b).

Van Ackeren et al. (2018) contrasted magnetoencephalography (MEG) responses to intelligible and unintelligible noise-vocoded speech in blind and sighted participants. Unsurprisingly, both groups showed speech tracking in auditory regions. However, speech-aligned responses were also observed in primary visual cortex. Although this was the case for both participant groups, stronger tracking for intelligible speech in visual cortex was observed *only* in blind participants. Thus, the well-established reorganisation of visual cortex for auditory dynamics in the blind (Collignon et al., 2015; Voss & Zatorre, 2012) seems to entail specific processing of intelligible speech.

In all of these cases, the simultaneous reduction in neural tracking and speech comprehension was produced by changes in the acoustic signal, making it difficult to disentangle acoustic and linguistic effects on neural dynamics (Kösem & van Wassenhove, 2017). Some studies have failed to find a correlation between neural tracking and comprehension, including studies that manipulated intelligibility of speech independently of its acoustics (e.g. through training) (Dai et al., 2022; Howard & Poeppel, 2010; Kösem et al., 2023; Mai et al., 2016; Millman et al., 2015; Zoefel & VanRullen, 2016; Zou et al., 2019). Other studies found differences in neural tracking when contrasting participants presented with the identical physical stimulus, but who differ in their proficiency of a given language and its linguistic structure (Ding et al., 2016; Lizarazu et al., 2021), or their expectation about linguistic content (Di Liberto et al., 2018). In addition, brain stimulation studies showed that the manipulation of speech-aligned neural dynamics results in a change in speech perception, even if the speech stimulus itself remains unchanged (Keshavarzi et al., 2020, 2021; Riecke et al., 2018; van Bree et al., 2021; Wilsch et al., 2018; Zoefel et al., 2020; Zoefel, Archer-Boyd, & Davis, 2018).

Together, studies relating neural tracking and speech comprehension have produced mixed results. On the one hand, some findings support the notion that the intelligibility of speech per se can influence neural dynamics. On the other hand, this notion has not always been confirmed; in addition, caution is warranted when manipulation of speech intelligibility goes along with acoustic changes, considering that a small change in acoustic parameters can have strong effects on the neural tracking response (Dai et al., 2022; Kösem et al., 2023). In this case, stronger tracking for intelligible speech might reflect acoustic processing rather than language-related effects. Finally, paradigms described are prone to other biases that can affect outcomes. For instance, intelligible speech is a particularly relevant acoustic stimulus and thus prone to capture listeners' attention. Neural dynamics including tracking of auditory rhythmic stimuli are modulated by attention (Lakatos et al., 2013), and so is their alignment to both acoustic and symbolic information in speech (Dai et al., 2022; Ding et al., 2018). Stronger brain responses to intelligible speech might reflect stronger, attention-related neural activity that is not specific to speech (Reetzke et al., 2021). We therefore conclude that the magnitude of neural tracking, and its difference between intelligibility conditions, might not be optimal to reveal neural dynamics that are specific to speech or reflect its comprehension.

3.2 | Speech-specific neural dynamics

We use face-specific brain responses, observed in the human fusiform gyrus (McCarthy et al., 1997), as an

analogue to illustrate speech-specific neural dynamics. Neural activity in some parts of this brain region is stronger during the presentation of human faces as compared to non-face stimuli. To evoke face-specific activity, the face needs to be identified, and this is only possible based on certain visual patterns. This means that faces and non-faces will necessarily differ in visual properties and these differences can explain the observed neural results—just like speech and non-speech sounds will always have some acoustic differences, and these can produce differences in neural dynamics. It is interesting, however, that the identification of a face activates certain neural populations that are otherwise not active and might respond in a way that differs from other, more general populations. The same logic applies to speech-specific neural dynamics, which might need to be activated by certain acoustic patterns but, once activated, have distinct properties. As we explain in the following, these speech-specific circuits and their properties might produce neural responses to speech that are not only stronger but also different from those to other non-verbal sounds.

In support of speech-specific neural dynamics, there is evidence that the lower range of eigenfrequencies (Section 2.1; Figure 1) is special for neural populations tuned to human speech. Several studies reported that comprehension of natural speech is correlated with neural dynamics in the delta, but not theta frequency range (Ding & Simon, 2014; Etard & Reichenbach, 2019; Keitel et al., 2018; Molinaro & Lizarazu, 2018). A recent study showed that this 'preference' for low frequencies is specific to speech and not found for other stimuli like music (Zuk et al., 2021). Indeed, the topographical pattern of delta activity in response to speech seems distinct from more typical auditory processes (Bourguignon et al., 2018) and involves parietal sensors (Zuk et al., 2021). In contrast, theta dynamics more closely resembles typical auditory activity (Bourguignon et al., 2018; Zuk et al., 2021). A somewhat different result was obtained by Hincapié Casas et al. (2021), who used MEG to measure neural activity aligned to speech sentences spoken at a fast rate (nine syllables/s) and compared it with that to sentences spoken at a slower rate, but time-compressed to the fast rate. This time-compressed speech was not only significantly less intelligible than natural speech; it also did not entrain neural activity—in contrast to naturally fast speech that produced a reliable alignment between MEG signal and speech rhythm. More research is required to determine whether not only the lower but also the upper limit of the delta–theta range has a distinct role for the processing of human speech. Results are less clear for the speech-specificity of the gamma range (cf. Meyer, 2018). This is due to the large (and still under-investigated) variability in the rate of phonemes in speech that would require pronounced (and potentially implausible) flexibility in the corresponding neural dynamics. Consequently, it remains to be shown that the 40-Hz 'preference' reported, for example, by Galambos et al. (1981) is related to an optimisation to process phonemes.

Additional results suggest that the neural tracking response to intelligible speech (Section 2.2; Figure 2) differs from that to non-intelligible acoustic controls. Zoefel et al. (2018) manipulated speech tracking by varying the timing of transcranial alternating current stimulation (tACS) relative to rhythmic speech and measured consequences of this manipulation using functional magnetic resonance imaging (fMRI). They found that tACS-induced changes in tracking altered fMRI responses to speech, but this effect was only observed when the speech was intelligible (16-channel noise-vocoded speech) and not for an unintelligible, amplitude-matched control stimulus (one-channel noise-vocoded speech). Van Bree et al. (2021) presented rhythmic noise-vocoded speech that was either clearly intelligible or unintelligible and noise-like. They showed that intelligible speech produces rhythmic fluctuations in the MEG that outlast the rhythmic stimulus, the 'neural echo' described above. Importantly, this sustained rhythmic response was not present for unintelligible speech and measured at parietal MEG sensors rather than those typically capturing auditory responses. This finding implies that rhythmic echoes, possibly reflecting temporal expectation of upcoming events (Section 2.3; Figure 3), might be particularly pronounced in response to speech compared to other acoustic stimuli. It is of note that intelligible speech does produce not only stronger neural echoes but also stronger neural dynamics during its presence. An interesting follow-up study would include the design of speech and non-speech sounds that produce comparable neural dynamics during the sound and the test whether intelligible speech still produces stronger neural echoes in this case. Lastly, a recent study found that rhythmic irregularities in noise-vocoded speech are easiest to detect if it is intelligible (Zoefel et al., 2023). Moreover, rhythm perception was more accurate in an experimental group that perceived a (sine-wave) stimulus as speech, as compared to another group that did not. This finding is additional evidence that temporal prediction mechanisms, putatively carried by neural dynamic activity, are improved during speech processing as compared to non-verbal processing.

The extraction of linguistic and other symbolic features of speech (Section 2.4; Figure 4) requires speech-specific processing (by definition, linguistic features are specific to speech). However, their tracking (Ding et al., 2016, 2018; Har-shai Yahav & Zion Golumbic, 2021) could rely on an

unspecific circuit that aligns neural processing to stimulus properties (or structure) in the attentional focus. It has been proposed that neural dynamics characterise the nested recursive structure of various stimuli, such as language, but also music, spatial sequences or mathematical structures (Dehaene et al., 2022). For instance, neural oscillatory activity can reflect the complexity of geometrical sequences and parse geometrical primitives (Al Roumi et al., 2021) in the same way as it parses syntactic structures in language (Ding et al., 2016). Here, a demonstration of speech-specific tracking would require the comparison with an unintelligible control stimulus that does not entrain neural dynamics. One of the rare studies that used such a comparison is described above: Zuk et al. (2021) demonstrated low-frequency tracking that is specific for human speech. Nevertheless, we currently lack evidence whether a single higher-level circuit tracks rhythmic structure in a stimulus, independently of the stimulus' identity, or whether speech is parsed differently from other non-speech stimuli.

Although audiovisual neural dynamics and corresponding perceptual effects might reflect optimisation to process speech (Section 2.5), it remains unclear in how far these can be generalised to other sounds. For example, visual speech cues reset auditory dynamics in general, not only speech-specific ones (Biau et al., 2021). Some evidence for speech-specific effects has been reported for auditory-motor interactions. Delta activity that is associated with speech comprehension seems to be coupled specifically to beta oscillations originating from the motor system (Keitel et al., 2018). Unlike delta, the frequency of beta oscillations (13–30 Hz) does not match rates found in speech and might be due to characteristics found in the motor system. The speech-specific role of beta oscillations from sensory-motor interactions has been confirmed by Michaelis et al. (2021). They presented participants with speech and non-speech sounds and found that only the former produced an amplitude decrease of such oscillations in left sensorimotor clusters (indicating increased motor activity). Further evidence for the important role of beta oscillations from the motor system, and their entrainment to speech, comes from research demonstrating changes in these oscillations during developmental stuttering (Etchell et al., 2016; Mollaei et al., 2021).

4 | NEURAL OSCILLATIONS: A MECHANISTIC ORIGIN OF SPEECH-OPTIMISED NEURAL DYNAMICS?

In this review, we focus on neural dynamics, temporal patterns of neural activity that seem optimised to process human speech. Neural oscillations are a distinct class of neural dynamics (Buzsáki & Draguhn, 2004; van Bree et al., 2022; Wang, 2010) and possess certain properties that might underlie a specialisation to process speech.

4.1 | Properties of neural oscillations that suggest speech optimisation

Neural oscillations have been put forward as a mechanism that structures and gates information processing in time (Lisman & Jensen, 2013; Schroeder & Lakatos, 2009; VanRullen, 2016). Oscillations are regular fluctuations in the excitability of neural ensembles that lead to a rhythmic alternation between phases of stimulus amplification and suppression (Buzsáki & Draguhn, 2004). The alignment of neural dynamics to periodic or quasi-periodic stimulus features, described as 'tracking' and 'entrainment' above, is often assumed to involve such oscillations (Lakatos et al., 2008, 2019; Obleser & Kayser, 2019). According to initial theories (Large & Jones, 1999; Schroeder & Lakatos, 2009), by adapting to the rhythm of speech, endogenous oscillations can align their amplifying phases to important events in the speech stream and their suppressive phases to irrelevant ones (e.g. a distracting, competing speaker), thereby efficiently and elegantly allocating neural resources to when they are needed. Consequently, neural oscillations and their entrainment are often considered instrumental in speech processing (Giraud & Poeppel, 2012; Meyer, 2018; Peelle & Davis, 2012).

There is no doubt that neural dynamics can follow specific temporal features of speech and other sounds. Evidence for an actual involvement of endogenous oscillations is trickier to demonstrate and has been discussed in detail elsewhere (e.g. Zoefel, ten Oever, & Sack, 2018). We here ask whether the involvement of neural oscillations is a promising model to explain speech-constrained neural dynamics and focus on properties of oscillations that might support such a model:

- Endogenous neural oscillations have an *eigenfrequency* (Hutcheon & Yarom, 2000) and will respond more strongly to stimuli close to their preferred frequency (Fröhlich, 2015; Herrmann et al., 2016). This is in line with findings on auditory neural dynamics that also possess an *eigenfrequency* (Section 2.1).
- Neural oscillations flexibly adapt to the rate of rhythmic stimulation if (and only if) it falls into their *eigenfrequency* range (constrained flexibility) and can tolerate a certain amount of jitter in the stimulus rhythm (Doelling & Assaneo, 2021), as observed for neural dynamics processing speech (Section 2.2).

- Neural oscillations are apt to undergo inertia, a property that identifies oscillations in ambiguous situations (Thut et al., 2011). This leads to oscillatory activity outlasting rhythmic sensory and electric stimulation (e.g. Kösem et al., 2018; van Bree et al., 2021). In simple scenarios (e.g. phrases with relatively constant syllable rate), this neural echo is mechanistically relevant for temporal expectation (Section 2.3) as it aligns neural dynamics with the expected timing of upcoming events. Interestingly, this might include scenarios of turn-taking (Wilson & Wilson, 2005).
- Similar to human speech (Section 2.4), neural oscillations can have nested structures, where slower and faster rhythms are coupled. Oscillations might therefore be suitable to process the hierarchical structures that speech has (Ghitza, 2011, 2013; Giraud & Poeppel, 2012)
- Interactions between distinct oscillatory populations play an important role for many basic neural and cognitive functions (Akam & Kullmann, 2014). An interaction between distinct modalities (auditory, visual, motor; Section 2.5) is also necessary for successful speech perception. Oscillatory networks might therefore support cross-modal speech processing (Bauer et al., 2020). The observation of a visually-induced reset of auditory delta–theta oscillations (Biau et al., 2021) is in line with this assumption.

Together, those properties that reveal a close match between neural dynamics and human speech can also been found in neural oscillations (eigenfrequency, constrained flexibility, temporal expectation, cross-modality). Some of these properties are unique to, others characteristic for neural oscillations (van Bree et al., 2022). This supports the notion of neural oscillations being involved in the generation of the observed speech-constrained neural dynamics.

4.2 | Neural oscillations underlying speech processing: open questions

Whereas the neural oscillation framework provides clear strengths, several open questions remain that need to be answered in follow-up work. In the future, these answers might lead to a model that explains neural dynamics during speech perception by complementing neural oscillations with additional, not necessarily oscillatory, processes.

- While neural oscillators seem robust to a certain amount of external temporal variability (Doelling & Assaneo, 2021), it remains unclear how they adapt to the temporal variability in human speech. Unlike other relevant sounds like music, speech consists of frequent changes in rate and entails relatively irregular silent gaps between words or phrases. This leads to a temporal variability that is high in spoken speech (Ten Oever et al., 2022; Tilsen & Arvaniti, 2013; Varnet et al., 2017), and some researchers raised doubts about whether it is rhythmic at all (Jadoul et al., 2016; Nolan & Jeon, 2014). Whereas dominant rates in speech do imply *some* rhythmicity, it is clear that a perfectly sinusoidal oscillation would struggle to align to this rhythm. Indeed, if speech perception relied on such an oscillation, regularly spoken speech should be easier to understand, which is not the case (Aubanel & Schwartz, 2020). This does not necessarily rule out an involvement of oscillations as they possess means to change their instantaneous frequency and phase. Acoustic 'edges' might serve as a cue to 'reset' oscillations (Doelling et al., 2014), and visual cues might prepare oscillatory activity for upcoming acoustic information (Biau et al., 2021; Mégevand et al., 2020; Thorne & Debener, 2014). How exactly this is done remains to be investigated, as well as the question whether and why a rhythmic neural process (i.e. oscillation) that needs to be continuously adapted has an advantage over a non-rhythmic one (for a different perspective, see also Meyer et al., 2020).
- Neural oscillations are difficult to identify during human speech as they need to be disentangled from evoked activity that is repeated regularly due to the rhythmicity of the stimulus (Haegens & Zion Golumbic, 2018; Zoefel, ten Oever, & Sack, 2018). Although progress has been made recently (Doelling et al., 2019), most of the evidence for their involvement is relatively indirect (such as 'entrainment echoes'; Section 2.3), and we still lack methods to extract endogenous oscillations during rhythmic stimulation.
- Due to their relation to neural excitability (Buzsáki & Draguhn, 2004), incoming information is supposedly inhibited during the low-excitability part of the oscillation (Lakatos et al., 2013). While this might be beneficial for speech perception if this suppressive phase coincides with distracting information (e.g. a competing speaker; Zion Golumbic et al., 2013), this might not always be the case, given considerable temporal variability in speech. It is unclear how the system deals with potentially important information coinciding with the low-excitability phase of the oscillation (VanRullen et al., 2014). A related prediction is that the perception of speech segments, phonemes specifically, should depend on the phase of entrained neural oscillations. However, several studies have failed to find such effects (Bosker & Kösem, 2017; Kösem et al., 2020): In

these studies, only the rate but not the phase of a rhythmic stimulus (speech or tACS), assumed to entrain oscillations, modulated the perception of speech phonemes.

- Neural oscillations at frequencies that do not match those of speech also seem to play a role for speech processing (such as alpha oscillations; Strauß et al., 2014). The precise role of these oscillations, and whether they are speech-specific, remains unclear, although speech-induced changes in oscillatory power (Meyer, 2018; Prystauka & Lewis, 2019) as well as results from regression-based tracking (Brodbeck et al., 2018; Broderick et al., 2021; Di Liberto et al., 2021; Weissbart et al., 2020) might imply such a role. In any case, the mismatch between neural and stimulus frequencies makes this role less straightforward to interpret in light of temporal patterns found in speech.
- Preferred neural 'time scales' seem to increase along the cortical hierarchy (Giraud et al., 2000; Kiebel et al., 2008; Murray et al., 2014; Wolff et al., 2022; see also summary of corresponding effects in the auditory system in Edwards & Chang, 2013). Corresponding analyses for oscillatory activity are sparse, but first results revealed an opposite pattern of decreasing time scales, with prefrontal areas showing fastest dynamics (>20 Hz) (Capilla et al., 2022). This seems to contradict the notion that endogenous oscillations track different hierarchical levels of human speech and needs to be resolved in future work.

5 | OUTLOOK: QUESTIONS AND HYPOTHESES FOR RESEARCH ON SPEECH-SPECIFIC NEURAL DYNAMICS

We conclude this article with a list of open question and testable hypotheses for the exciting field of neural dynamics processing human speech.

- If face-specific neural activity requires the presence of certain features that are necessary to identify faces and activate face-specific brain areas, then similar speech-specific features might be necessary to activate speech-specific neural dynamics. It is likely that such features exist, given that we perceive speech as categorically different from most other sounds. It remains, however, an open question what these features are. In studies reporting speech-specific dynamics, any difference between intelligible and unintelligible (or non-)speech sounds might have produced them. For example, 16- and 1-channel noise-vocoded stimuli do not only differ in their intelligibility but also in their spectral complexity (Shannon et al., 1995). Time-compressed speech might have altered various acoustic features in addition to reduced intelligibility. We here propose that recognising human speech as such—based on (acoustic or linguistic) features that are distinct for speech and allow the listener to identify it—is crucial to activate speech-specific processing, a hypothesis that needs to be tested in the future and might reveal insights into the question of what makes human speech such a characteristic stimulus. A study by Overath et al. (2015) is important in this respect, demonstrating that parts of the superior temporal sulcus (STS) respond selectively to acoustic, temporal structure of speech (but not other sounds).
  In any case, if those characteristic features are not linguistic, then we should be able to reproduce them in non-speech stimuli that then activate the same speech-specific dynamics. As long as a non-speech stimulus mimics the critical properties of speech (e.g. its typical rate, association with visual cues and temporal predictability), it should produce neural dynamics that so far seem distinct for speech (Section 3.2).
- How speech-specific are neural dynamics reflecting temporal expectation? These dynamics should disappear when it has become clear that the temporal expectation has been violated. This can be tested and compared with similar effects observed for non-speech stimuli in which temporal expectation is manipulated. For expectations on hierarchically higher levels of linguistic information, the effect should only be observed for participants proficient in the language spoken. Moreover, whereas speech is easier to understand when it contains natural temporal variability (Aubanel & Schwartz, 2020), it remains unclear whether equivalent effects exist for non-speech sounds.
- Given the tight temporal correspondence between lip movements and speech, are temporal expectations given by visual cues more relevant for speech processing than for other audiovisual stimuli? Speech-specific dynamics might particularly rely on visual cues to anticipate auditory events that might otherwise difficult to predict, like the onset of a new phrase (Zoefel, 2021).
- Did neural dynamics and speech production co-evolve (Poeppel and Assaneo 2020), or was one shaped by the other? One the one hand, temporal constants of neural dynamics are conserved across species (Buzsáki et al., 2013). If basic neural architecture principles are indeed preserved throughout evolution, then dynamics of human speech might have adapted according to corresponding temporal constraints. On the other hand, neural dynamics are dependent on sensory experience (Kral, 2013). As one of the most prominent acoustic stimuli an individual is exposed to since birth,

exposure to speech might have constraint auditory cortices to adapt to its temporal dynamics.

- If neural dynamics are shaped by the exposure to speech, do they develop in parallel with language acquisition? Do listeners show differences in neural dynamics when presented with their native language as compared to other ones?
- Does this potential co-evolution have an impact on the processing of other auditory stimuli? Such a 'spillover effect' might explain why we are attracted to music—a stimulus that fluctuates at similar rates, is temporarily predictable but entails some variability and has therefore similar properties as human speech. Can the cross-modal wiring of neural dynamics explain why we like to dance to music (audio-motor interactions) or watch musicians during a concert (audiovisual interactions)? Related, a recent study suggested that audition's (delta/theta) *eigenfrequency* is indeed imposed onto eye movements during reading (Gagl et al., 2022), while another one confirmed that this effect indeed involves rhythmic brain activity (Henke et al., 2023).
- It seems to be a general property of the human brain that 'intrinsic time scales' become longer at higher levels of the cortical hierarchy (Edwards & Chang, 2013; Giraud et al., 2000; Kiebel et al., 2008; Murray et al., 2014; Wolff et al., 2022). Not much is known about how much of this phenomenon holds for speech processing. In particular, *eigenfrequencies* should decrease along the speech processing hierarchy, as relevant rates in speech also decrease.
- Do speech-specific neural dynamics localise to specific brain areas? It is of note that most speech-specific effects reported above are measured in regions (or at sensors) that do not show the strongest response to acoustic rhythms in general. Arguably, networks responding more readily to human speech than to other sounds (Saur et al., 2008; Scott et al., 2000) are likely to show such speech-specific neural dynamics, but this assumption requires confirmation. Moreover, these networks are large and contain sub-networks with distinct properties (e.g. *eigenfrequency*).

6 | CONCLUSION

In this review, we address the role of neural dynamics in the processing of speech and other sounds. We highlight that the brain can track various auditory signals and that this tracking has specific properties and constraints. These resemble characteristics of human speech and might therefore reflect the system's optimisation for speech processing. We also describe how neural dynamics during speech seem to be both quantitatively and qualitatively different from dynamics observed during other acoustic stimuli. More research is needed to understand the mechanistic origins of speech-specific dynamics and their impact on speech analysis.

AUTHOR CONTRIBUTIONS

Benedikt Zoefel: Conceptualization; funding acquisition; project administration; visualization; writing—original draft; writing—review and editing. **Anne Kösem:** Conceptualization; funding acquisition; project administration; visualization; writing—original draft; writing—review and editing.

ACKNOWLEDGEMENTS

This work was supported by the Agence Nationale de la Recherche (Grant Numbers ANR-21-CE37-0002 and ANR-21-CE37-0003) and the Fondation Pour l'Audition (Grant Number FPA-RD-2021-10).

CONFLICT OF INTEREST STATEMENT

The authors report no conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ejn.16221.

DATA AVAILABILITY STATEMENT

Not applicable (no data have been collected for this review).

ORCID

Benedikt Zoefel https://orcid.org/0000-0002-9800-2551
Anne Kösem https://orcid.org/0000-0002-2692-9999

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*, 13367–13372. https://doi.org/10.1073/pnas.201400998

Akam, T., & Kullmann, D. M. (2014). Oscillatory multiplexing of population codes for selective communication in the mammalian brain. *Nature Reviews. Neuroscience*, *15*, 111–122. https://doi.org/10.1038/nrn3668

Al Roumi, F., Marti, S., Wang, L., Amalric, M., & Dehaene, S. (2021). Mental compression of spatial sequences in human working memory using numerical and geometrical primitives. *Neuron*, *109*, 2627–2639.e4. https://doi.org/10.1016/j.neuron.2021.06.009

Assaneo, M. F., & Poeppel, D. (2018). The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. *Science Advances*, *4*, eaao3842. https://doi.org/10.1126/sciadv.aao3842

Aubanel, V., & Schwartz, J.-L. (2020). The role of isochrony in speech perception in noise. *Scientific Reports*, *10*, 19580. https://doi.org/10.1038/s41598-020-76594-1

Baltzell, L. S., Srinivasan, R., & Richards, V. (2019). Hierarchical organization of melodic sequences is encoded by cortical entrainment. *NeuroImage*, *200*, 490–500. https://doi.org/10.1016/j.neuroimage.2019.06.054

Barczak, A., O'Connell, M. N., McGinnis, T., Ross, D., Mowery, T., Falchier, A., & Lakatos, P. (2018). Top-down, contextual entrainment of neuronal oscillations in the auditory thalamocortical circuit. *Proceedings of the National Academy of Sciences*, *115*, E7605–E7614. https://doi.org/10.1073/pnas.1714684115

Bauer, A.-K. R., Debener, S., & Nobre, A. C. (2020). Synchronisation of neural oscillations and cross-modal influences. *Trends in Cognitive Sciences*, *24*, 481–495. https://doi.org/10.1016/j.tics.2020.03.003

Bauer, A.-K. R., Jaeger, M., Thorne, J. D., Bendixen, A., & Debener, S. (2015). The auditory dynamic attending theory revisited: A closer look at the pitch comparison task. *Brain Research, Predictive and Attentive Processing in Perception and Action*, *1626*, 198–210. https://doi.org/10.1016/j.brainres.2015.04.032

Besle, J., Schevon, C. A., Mehta, A. D., Lakatos, P., Goodman, R. R., McKhann, G. M., Emerson, R. G., & Schroeder, C. E. (2011). Tuning of the human neocortex to the temporal dynamics of attended events. *The Journal of Neuroscience*, *31*, 3176–3185. https://doi.org/10.1523/JNEUROSCI.4518-10.2011

Biau, E., Wang, D., Park, H., Jensen, O., & Hanslmayr, S. (2021). Auditory detection is modulated by theta phase of silent lip movements. *Current Research in Neurobiology*, *2*, 100014. https://doi.org/10.1016/j.crneur.2021.100014

Boddy, J. (1981). Evoked potentials and the dynamics of language processing. *Biological Psychology*, *13*, 125–140. https://doi.org/10.1016/0301-0511(81)90031-4

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, *8*, 389–395. https://doi.org/10.1038/nn1409

Bonnefond, M., & Jensen, O. (2012). Alpha oscillations serve to protect working memory maintenance against anticipated distracters. *Current Biology*, *22*, 1969–1974. https://doi.org/10.1016/j.cub.2012.08.029

Bosker, H. R. (2017). Accounting for rate-dependent category boundary shifts in speech perception. *Attention, Perception, & Psychophysics*, *79*, 333–343. https://doi.org/10.3758/s13414-016-1206-4

Bosker, H. R., & Kösem, A. (2017). An entrained rhythm's frequency, not phase, influences temporal sampling of speech. *Interspeech*, *2017*, 2416–2420.

Bourguignon, M., Baart, M., Kapnoula, E. C., & Molinaro, N. (2020). Lip-reading enables the brain to synthesize auditory features of unknown silent speech. *The Journal of Neuroscience*, *40*, 1053–1065. https://doi.org/10.1523/JNEUROSCI.1101-19.2019

Bourguignon, M., Molinaro, N., & Wens, V. (2018). Contrasting functional imaging parametric maps: The mislocation problem and alternative solutions. *NeuroImage*, *169*, 200–211. https://doi.org/10.1016/j.neuroimage.2017.12.033

Bouwer, F. L., Fahrenfort, J. J., Millard, S. K., Kloosterman, N. A., & Slagter, H. A. (2023). A silent disco: Persistent entrainment of low-frequency neural oscillations underlies beat-based, but not pattern-based temporal expectations. *Journal of Cognitive Neuroscience*, *35*, 990–1020. https://doi.org/10.1162/jocn_a_01985

Breska, A., & Deouell, L. Y. (2017). Neural mechanisms of rhythm-based temporal prediction: Delta phase-locking reflects temporal predictability but not rhythmic entrainment. *PLoS Biology*, *15*, e2001665. https://doi.org/10.1371/journal.pbio.2001665

Brodbeck, C., Hong, L. E., & Simon, J. Z. (2018). Rapid transformation from auditory to linguistic representations of continuous speech. *Current Biology*, *28*, 3976–3983.e5. https://doi.org/10.1016/j.cub.2018.10.042

Brodbeck, C., & Simon, J. Z. (2020). Continuous speech processing. *Current Opinion in Physiology*, *18*, 25–31. https://doi.org/10.1016/j.cophys.2020.07.014

Broderick, M. P., Di Liberto, G. M., Anderson, A. J., Rofes, A., & Lalor, E. C. (2021). Dissociable electrophysiological measures of natural language processing reveal differences in speech comprehension strategy in healthy ageing. *Scientific Reports*, *11*, 4963. https://doi.org/10.1038/s41598-021-84597-9

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, *304*, 1926–1929. https://doi.org/10.1126/science.1099745

Buzsáki, G., Logothetis, N., & Singer, W. (2013). Scaling brain size, keeping timing: Evolutionary preservation of brain rhythms. *Neuron*, *80*, 751–764. https://doi.org/10.1016/j.neuron.2013.10.002

Capilla, A., Arana, L., García-Huéscar, M., Melcón, M., Gross, J., & Campo, P. (2022). The natural frequencies of the resting human brain: An MEG-based atlas. *NeuroImage*, *258*, 119373. https://doi.org/10.1016/j.neuroimage.2022.119373

Carbonell, K. M., & Lotto, A. J. (2014). Speech is not special ... again. *Frontiers in Psychology*, *5*, 427. https://doi.org/10.3389/fpsyg.2014.00427

Carver, R. P. (1973). Effects of increasing the rate of speech presentation upon comprehension. *Journal of Educational Psychology*, *65*, 118–126. https://doi.org/10.1037/h0034783

Castellucci, G. A., Kovach, C. K., Howard, M. A., Greenlee, J. D. W., & Long, M. A. (2022). A speech planning network for interactive language use. *Nature*, *602*, 117–122. https://doi.org/10.1038/s41586-021-04270-z

Chi, T., Gao, Y., Guyton, M. C., Ru, P., & Shamma, S. (1999). Spectro-temporal modulation transfer functions and speech intelligibility. *The Journal of the Acoustical Society of America*, *106*, 2719–2732. https://doi.org/10.1121/1.428100

Collignon, O., Dormal, G., de Heering, A., Lepore, F., Lewis, T. L., & Maurer, D. (2015). Long-lasting crossmodal cortical reorganization triggered by brief postnatal visual deprivation. *Current Biology*, *25*, 2379–2383. https://doi.org/10.1016/j.cub.2015.07.036

Coupé, C., Oh, Y. M., Dediu, D., & Pellegrino, F. (2019). Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science Advances*, *5*, eaaw2594. https://doi.org/10.1126/sciadv.aaw2594

Dai, B., McQueen, J. M., Terporten, R., Hagoort, P., & Kösem, A. (2022). Distracting linguistic information impairs neural

tracking of attended speech. *Current Research in Neurobiology*, *3*, 100043. https://doi.org/10.1016/j.crneur.2022.100043

Dehaene, S., Roumi, F. A., Lakretz, Y., Planton, S., & Sablé-Meyer, M. (2022). Symbols and mental programs: A hypothesis about human singularity. *Trends in Cognitive Sciences*, *26*, 751–766. https://doi.org/10.1016/j.tics.2022.06.010

Di Liberto, G. M., Crosse, M. J., & Lalor, E. C. (2018). Cortical measures of phoneme-level speech encoding correlate with the perceived clarity of natural speech. *eNeuro*, *5*, ENEURO.0084-18.2018. https://doi.org/10.1523/ENEURO.0084-18.2018

Di Liberto, G. M., Nie, J., Yeaton, J., Khalighinejad, B., Shamma, S. A., & Mesgarani, N. (2021). Neural representation of linguistic feature hierarchy reflects second-language proficiency. *NeuroImage*, *227*, 117586. https://doi.org/10.1016/j.neuroimage.2020.117586

Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, *21*, 1664–1670. https://doi.org/10.1177/0956797610384743

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*, 158–164. https://doi.org/10.1038/nn.4186

Ding, N., Pan, X., Luo, C., Su, N., Zhang, W., & Zhang, J. (2018). Attention is required for knowledge-based sequential grouping: Insights from the integration of syllables into words. *The Journal of Neuroscience*, *38*, 1178–1188. https://doi.org/10.1523/JNEUROSCI.2606-17.2017

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience and Biobehavioral Reviews*, *81*, 181–187. https://doi.org/10.1016/j.neubiorev.2017.02.011

Ding, N., & Simon, J. Z. (2013). Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *The Journal of Neuroscience*, *33*, 5728–5735. https://doi.org/10.1523/JNEUROSCI.5297-12.2013

Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, *8*, 311. https://doi.org/10.3389/fnhum.2014.00311

Doelling, K.B., Arnal, L.H., & Assaneo, M.F. (2022). Adaptive oscillators provide a hard-coded Bayesian mechanism for rhythmic inference. BioRxiv. doi: https://doi.org/10.1101/2022.06.18.496664

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*(Pt 2), 761–768. https://doi.org/10.1016/j.neuroimage.2013.06.035

Doelling, K. B., & Assaneo, M. F. (2021). Neural oscillations are a start toward understanding brain activity rather than the end. *PLoS Biology*, *19*, e3001234. https://doi.org/10.1371/journal.pbio.3001234

Doelling, K. B., Assaneo, M. F., Bevilacqua, D., Pesaran, B., & Poeppel, D. (2019). An oscillator model better predicts cortical entrainment to music. *Proceedings of the National Academy of Sciences*, *116*, 10113–10121. https://doi.org/10.1073/pnas.1816414116

Doelling, K. B., & Poeppel, D. (2015). Cortical entrainment to music and its modulation by expertise. *PNAS*, *112*, E6233–E6242. https://doi.org/10.1073/pnas.1508431112

Drijvers, L., & Özyürek, A. (2017). Visual context enhanced: The joint contribution of iconic gestures and visible speech to degraded speech comprehension. *Journal of Speech, Language, and Hearing Research*, *60*, 212–222. https://doi.org/10.1044/2016_JSLHR-H-16-0101

Edwards, E., & Chang, E. F. (2013). Syllabic (~2–5 Hz) and fluctuation (~1–10 Hz) ranges in speech and auditory processing. *Hearing Research, Communication Sounds and the Brain: New Directions and Perspectives*, *305*, 113–134. https://doi.org/10.1016/j.heares.2013.08.017

Etard, O., & Reichenbach, T. (2019). Neural speech tracking in the theta and in the delta frequency band differentially encode clarity and comprehension of speech in noise. *The Journal of Neuroscience*, *39*, 5750–5759. https://doi.org/10.1523/JNEUROSCI.1828-18.2019

Etchell, A. C., Ryan, M., Martin, E., Johnson, B. W., & Sowman, P. F. (2016). Abnormal time course of low beta modulation in non-fluent preschool children: A magnetoencephalographic study of rhythm tracking. *NeuroImage*, *125*, 953–963. https://doi.org/10.1016/j.neuroimage.2015.10.086

Farahbod, H., Saberi, K., & Hickok, G. (2020). The rhythm of attention: Perceptual modulation via rhythmic entrainment is low-pass and attention mediated. *Attention, Perception, & Psychophysics*, *82*, 3558–3570. https://doi.org/10.3758/s13414-020-02095-y

Frank, S. L., Bod, R., & Christiansen, M. H. (2012). How hierarchical is language use? *Proceedings of the Royal Society B: Biological Sciences*, *279*, 4522–4531. https://doi.org/10.1098/rspb.2012.1741

Freeman, E. D., Ipser, A., Palmbaha, A., Paunoiu, D., Brown, P., Lambert, C., Leff, A., & Driver, J. (2013). Sight and sound out of synch: Fragmentation and renormalisation of audiovisual integration and subjective timing. *Cortex*, *49*, 2875–2887. https://doi.org/10.1016/j.cortex.2013.03.006

Friston, K. J. (2019). Waves of prediction. *PLoS Biology*, *17*, e3000426. https://doi.org/10.1371/journal.pbio.3000426

Fröhlich, F. (2015). Chapter 3 - Experiments and models of cortical oscillations as a target for noninvasive brain stimulation. In S. Bestmann (Ed.), *Progress in brain research* (pp. 41–73). Elsevier.

Gagl, B., Gregorova, K., Golch, J., Hawelka, S., Sassenhagen, J., Tavano, A., Poeppel, D., & Fiebach, C. J. (2022). Eye movements during text reading align with the rate of speech production. *Nature Human Behaviour*, *6*, 429–442. https://doi.org/10.1038/s41562-021-01215-4

Galambos, R., Makeig, S., & Talmachoff, P. J. (1981). A 40-Hz auditory potential recorded from the human scalp. *PNAS*, *78*, 2643–2647. https://doi.org/10.1073/pnas.78.4.2643

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*, 361–377. https://doi.org/10.3758/BF03193857

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*, 278–285. https://doi.org/10.1016/j.tics.2006.04.008

Ghitza, O. (2011). Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology*, *2*, 130. https://doi.org/10.3389/fpsyg.2011.00130

Ghitza, O. (2013). The theta-syllable: A unit of speech information defined by cortical function. *Frontiers in Psychology*, *4*, 138. https://doi.org/10.3389/fpsyg.2013.00138

Ghitza, O., Giraud, A.-L., & Poeppel, D. (2013). Neuronal oscillations and speech perception: Critical-band temporal envelopes are the essence. *Frontiers in Human Neuroscience*, *6*, 340. https://doi.org/10.3389/fnhum.2012.00340

Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, *66*, 113–126. https://doi.org/10.1159/000208934

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. https://doi.org/10.7554/eLife.24763

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*, 1127–1134. https://doi.org/10.1016/j.neuron.2007.09.038

Giraud, A.-L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., & Kleinschmidt, A. (2000). Representation of the temporal envelope of sounds in the human brain. *Journal of Neurophysiology*, *84*, 1588–1598. https://doi.org/10.1152/jn.2000.84.3.1588

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*, 511–517. https://doi.org/10.1038/nn.3063

Giroud, J., Trébuchon, A., Schön, D., Marquis, P., Liegeois-Chauvel, C., Poeppel, D., & Morillon, B. (2020). Asymmetric sampling in human auditory cortex reveals spectral processing hierarchy. *PLoS Biology*, *18*, e3000207. https://doi.org/10.1371/journal.pbio.3000207

Greenberg, S. (1999). Speaking in shorthand – A syllable-centric perspective for understanding pronunciation variation. *Speech Communication*, *29*, 159–176. https://doi.org/10.1016/S0167-6393(99)00050-3

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech—A syllable-centric perspective. *Journal of Phonetics, Temporal Integration in the Perception of Speech*, *31*, 465–485. https://doi.org/10.1016/j.wocn.2003.09.005

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*, e1001752. https://doi.org/10.1371/journal.pbio.1001752

Haegens, S., & Zion Golumbic, E. (2018). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience and Biobehavioral Reviews*, *86*, 150–165. https://doi.org/10.1016/j.neubiorev.2017.12.002

Har-shai Yahav, P., & Zion Golumbic, E. (2021). Linguistic processing of task-irrelevant speech at a cocktail party. *eLife*, *10*, e65096. https://doi.org/10.7554/eLife.65096

Helfer, K. S. (1997). Auditory and auditory-visual perception of clear and conversational speech. *Journal of Speech, Language, and Hearing Research*, *40*, 432–443. https://doi.org/10.1044/jslhr.4002.432

Henke, L., Lewis, A. G., & Meyer, L. (2023). Fast and slow rhythms of naturalistic reading revealed by combined eye-tracking and electroencephalography. *The Journal of Neuroscience*, *43*, 4461–4469. https://doi.org/10.1523/JNEUROSCI.1849-22.2023

Herbst, S. K., & Obleser, J. (2017). Implicit variations of temporal predictability: Shaping the neural oscillatory and behavioural response. *Neuropsychologia*, *101*, 141–152. https://doi.org/10.1016/j.neuropsychologia.2017.05.019

Herrmann, C. S. (2001). Human EEG responses to 1-100 Hz flicker: Resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Experimental Brain Research*, *137*, 346–353. https://doi.org/10.1007/s002210100682

Herrmann, C. S., Murray, M. M., Ionta, S., Hutt, A., & Lefebvre, J. (2016). Shaping intrinsic neural oscillations with periodic stimulation. *The Journal of Neuroscience*, *36*, 5328–5337. https://doi.org/10.1523/JNEUROSCI.0236-16.2016

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*, 9–21. https://doi.org/10.1016/j.bandl.2012.10.006

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*, 322–334. https://doi.org/10.1111/j.1469-8986.2011.01314.x

Hickok, G., Farahbod, H., & Saberi, K. (2015). The rhythm of perception: Entrainment to acoustic rhythms induces subsequent perceptual oscillation. *Psychological Science*, *26*, 1006–1013. https://doi.org/10.1177/0956797615576533

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews. Neuroscience*, *8*, 393–402. https://doi.org/10.1038/nrn2113

Hickok, G., & Poeppel, D. (2016). Neural basis of speech perception. In G. Hickok & S. L. Small (Eds.), *Neurobiology of language* (pp. 299–310). Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00025-0

Hincapié Casas, A. S., Lajnef, T., Pascarella, A., Guiraud-Vinatea, H., Laaksonen, H., Bayle, D., Jerbi, K., & Boulenger, V. (2021). Neural oscillations track natural but not artificial fast speech: Novel insights from speech-brain coupling using MEG. *NeuroImage*, *244*, 118577. https://doi.org/10.1016/j.neuroimage.2021.118577

Ho, H. T., Leung, J., Burr, D. C., Alais, D., & Morrone, M. C. (2017). Auditory sensitivity and decision criteria oscillate at different frequencies separately for the two ears. *Current Biology*, *27*, 3643–3649.e3. https://doi.org/10.1016/j.cub.2017.10.017

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*, 2500–2511. https://doi.org/10.1152/jn.00251.2010

Hutcheon, B., & Yarom, Y. (2000). Resonance, oscillation and the intrinsic frequency preferences of neurons. *Trends in Neurosciences*, *23*, 216–222. https://doi.org/10.1016/S0166-2236(00)01547-2

Inbar, M., Grossman, E., & Landau, A. N. (2020). Sequences of intonation units form a ~1 Hz rhythm. *Scientific Reports*, *10*, 15846. https://doi.org/10.1038/s41598-020-72739-4

Jadoul, Y., Ravignani, A., Thompson, B., Filippi, P., & de Boer, B. (2016). Seeking temporal predictability in speech: Comparing statistical approaches on 18 world languages. *Frontiers in*

Human Neuroscience, *10*, 586. https://doi.org/10.3389/fnhum.2016.00586

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*, *13*, 313–319. https://doi.org/10.1111/1467-9280.00458

Kayser, C. (2019). Evidence for the rhythmic perceptual sampling of auditory scenes. *Frontiers in Human Neuroscience*, *13*, 249. https://doi.org/10.3389/fnhum.2019.00249

Kayser, S. J., Ince, R. A. A., Gross, J., & Kayser, C. (2015). Irregular speech rate dissociates auditory cortical entrainment, evoked responses, and frontal alpha. *The Journal of Neuroscience*, *35*, 14691–14701. https://doi.org/10.1523/JNEUROSCI.2243-15.2015

Kazanina, N., & Tavano, A. (2023). What neural oscillations can and cannot do for syntactic structure building. *Nature Reviews. Neuroscience*, *24*, 113–128. https://doi.org/10.1038/s41583-022-00659-5

Keitel, A., Gross, J., & Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS Biology*, *16*, e2004473. https://doi.org/10.1371/journal.pbio.2004473

Keshavarzi, M., Kegler, M., Kadir, S., & Reichenbach, T. (2020). Transcranial alternating current stimulation in the theta band but not in the delta band modulates the comprehension of naturalistic speech in noise. *NeuroImage*, *210*, 116557. https://doi.org/10.1016/j.neuroimage.2020.116557

Keshavarzi, M., Varano, E., & Reichenbach, T. (2021). Cortical tracking of a background speaker modulates the comprehension of a foreground speech signal. *The Journal of Neuroscience*, *41*, 5093–5101. https://doi.org/10.1523/JNEUROSCI.3200-20.2021

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Computational Biology*, *4*, e1000209. https://doi.org/10.1371/journal.pcbi.1000209

Kösem, A., Bosker, H. R., Jensen, O., Hagoort, P., & Riecke, L. (2020). Biasing the perception of spoken words with transcranial alternating current stimulation. *Journal of Cognitive Neuroscience*, *32*, 1428–1437. https://doi.org/10.1162/jocn_a_01579

Kösem, A., Bosker, H. R., Takashima, A., Meyer, A., Jensen, O., & Hagoort, P. (2018). Neural entrainment determines the words we hear. *Current Biology*, *28*, 2867–2875.e3. https://doi.org/10.1016/j.cub.2018.07.023

Kösem, A., Dai, B., McQueen, J. M., & Hagoort, P. (2023). Neural tracking of speech envelope does not unequivocally reflect intelligibility. *NeuroImage*, *272*, 120040. https://doi.org/10.1016/j.neuroimage.2023.120040

Kösem, A., Gramfort, A., & van Wassenhove, V. (2014). Encoding of event timing in the phase of neural oscillations. *NeuroImage*, *92*, 274–284. https://doi.org/10.1016/j.neuroimage.2014.02.010

Kösem, A., & van Wassenhove, V. (2017). Distinct contributions of low- and high-frequency neural oscillations to speech comprehension. *Language, Cognition and Neuroscience*, *32*, 536–544. https://doi.org/10.1080/23273798.2016.1238495

Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: A timely subcortico-cortical framework. *Trends in Cognitive Sciences*, *14*, 392–399. https://doi.org/10.1016/j.tics.2010.06.005

Kral, A. (2013). Auditory critical periods: A review from system's perspective. *Neuroscience*, *247*, 117–133. https://doi.org/10.1016/j.neuroscience.2013.05.021

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Lakatos, P., Gross, J., & Thut, G. (2019). A new unifying account of the roles of neuronal entrainment. *Current Biology*, *29*, R890–R905. https://doi.org/10.1016/j.cub.2019.07.075

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*, 110–113. https://doi.org/10.1126/science.1154735

Lakatos, P., Musacchia, G., O'Connel, M. N., Falchier, A. Y., Javitt, D. C., & Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. *Neuron*, *77*, 750–761. https://doi.org/10.1016/j.neuron.2012.11.034

Lamekina, Y., & Meyer, L. (2023). Entrainment to speech prosody influences subsequent sentence comprehension. *Language, Cognition and Neuroscience*, *38*, 263–276. https://doi.org/10.1080/23273798.2022.2107689

Landemard, A., Bimbard, C., Demené, C., Shamma, S., Norman-Haignere, S., & Boubenec, Y. (2021). Distinct higher-order representations of natural sounds in human and ferret auditory cortex. *eLife*, *10*, e65566. https://doi.org/10.7554/eLife.65566

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*, 119–159. https://doi.org/10.1037/0033-295X.106.1.119

Lawrance, E. L. A., Harper, N. S., Cooke, J. E., & Schnupp, J. W. H. (2014). Temporal predictability enhances auditory detection. *The Journal of the Acoustical Society of America*, *135*, EL357–EL363. https://doi.org/10.1121/1.4879667

Lenc, T., Keller, P. E., Varlet, M., & Nozaradan, S. (2020). Neural and behavioral evidence for frequency-selective context effects in rhythm processing in humans. *Cereb Cortex Commun*, *1*, tgaa037. https://doi.org/10.1093/texcom/tgaa037

Levinson, S. C. (2016). Turn-taking in human communication – Origins and implications for language processing. *Trends in Cognitive Sciences*, *20*, 6–14. https://doi.org/10.1016/j.tics.2015.10.010

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*, 447–454. https://doi.org/10.1016/j.tics.2006.08.007

L'Hermite, S., & Zoefel, B. (2023). Rhythmic entrainment echoes in auditory perception. *The Journal of Neuroscience*, *43*, 6667–6678. https://doi.org/10.1523/JNEUROSCI.0051-23.2023

Lisman, J. E., & Jensen, O. (2013). The theta-gamma neural code. *Neuron*, *77*, 1002–1016. https://doi.org/10.1016/j.neuron.2013.03.007

Lizarazu, M., Carreiras, M., Bourguignon, M., Zarraga, A., & Molinaro, N. (2021). Language proficiency entails tuning cortical activity to second language speech. *Cerebral Cortex*, *31*, 3820–3831. https://doi.org/10.1093/cercor/bhab051

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*, 1001–1010. https://doi.org/10.1016/j.neuron.2007.06.004

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. https://doi.org/10.1016/j.neuroimage.2016.02.064

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, *9*, 605–610. https://doi.org/10.1162/jocn.1997.9.5.605

Mcgurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748. https://doi.org/10.1038/264746a0

Mégevand, P., Mercier, M. R., Groppe, D. M., Golumbic, E. Z., Mesgarani, N., Beauchamp, M. S., Schroeder, C. E., & Mehta, A. D. (2020). Crossmodal phase reset and evoked responses provide complementary mechanisms for the influence of visual speech in auditory cortex. *The Journal of Neuroscience*, *40*, 8530–8542. https://doi.org/10.1523/JNEUROSCI.0555-20.2020

Meng, Q., Hegner, Y. L., Giblin, I., McMahon, C., & Johnson, B. W. (2021). Lateralized cerebral processing of abstract linguistic structure in clear and degraded speech. *Cerebral Cortex*, *31*, 591–602. https://doi.org/10.1093/cercor/bhaa245

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*, 1006–1010. https://doi.org/10.1126/science.1245994

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience*, *48*, 2609–2621. https://doi.org/10.1111/ejn.13748

Meyer, L., Sun, Y., & Martin, A. E. (2020). Synchronous, but not entrained: Exogenous and endogenous cortical rhythms of speech and language processing. *Language, Cognition and Neuroscience*, *35*, 1089–1099. https://doi.org/10.1080/23273798.2019.1693050

Michaelis, K., Miyakoshi, M., Norato, G., Medvedev, A. V., & Turkeltaub, P. E. (2021). Motor engagement relates to accurate perception of phonemes and audiovisual words, but not auditory words. *Commun Biol*, *4*, 108. https://doi.org/10.1038/s42003-020-01634-5

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*, 533–545. https://doi.org/10.1162/jocn_a_00719

Molinaro, N., & Lizarazu, M. (2018). Delta (but not theta)-band cortical entrainment involves speech-specific processing. *The European Journal of Neuroscience*, *48*, 2642–2650. https://doi.org/10.1111/ejn.13811

Mollaei, F., Mersov, A., Woodbury, M., Jobst, C., Cheyne, D., & De Nil, L. (2021). White matter microstructural differences underlying beta oscillations during speech in adults who stutter. *Brain and Language*, *215*, 104921. https://doi.org/10.1016/j.bandl.2021.104921

Moore, D. R. (2000). Auditory neuroscience: Is speech special? *Current Biology*, *10*, R362–R364. https://doi.org/10.1016/S0960-9822(00)00479-6

Morillon, B., Schroeder, C. E., & Wyart, V. (2014). Motor contributions to the temporal precision of auditory attention. *Nature Communications*, *5*, 5255. https://doi.org/10.1038/ncomms6255

Murray, J. D., Bernacchia, A., Freedman, D. J., Romo, R., Wallis, J. D., Cai, X., Padoa-Schioppa, C., Pasternak, T., Seo, H., Lee, D., & Wang, X.-J. (2014). A hierarchy of intrinsic timescales across primate cortex. *Nature Neuroscience*, *17*, 1661–1663. https://doi.org/10.1038/nn.3862

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology. Human Perception and Performance*, *24*, 756–766. https://doi.org/10.1037/0096-1523.24.3.756

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

Nourski, K. V., Reale, R. A., Oya, H., Kawasaki, H., Kovach, C. K., Chen, H., Howard, M. A., & Brugge, J. F. (2009). Temporal envelope of time-compressed speech represented in the human auditory cortex. *The Journal of Neuroscience*, *29*, 15564–15574. https://doi.org/10.1523/JNEUROSCI.3065-09.2009

Nozaradan, S., Peretz, I., & Mouraux, A. (2012). Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *The Journal of Neuroscience*, *32*, 17572–17581. https://doi.org/10.1523/JNEUROSCI.3203-12.2012

Obleser, J., & Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends in Cognitive Sciences*, *23*, 913–926. https://doi.org/10.1016/j.tics.2019.08.004

Oganian, Y., & Chang, E. F. (2019). A speech envelope landmark for syllable encoding in human superior temporal gyrus. *Science Advances*, *5*, eaay6279. https://doi.org/10.1126/sciadv.aay6279

Oganian, Y., Kojima, K., Breska, A., Cai, C., Findlay, A., Chang, E., & Nagarajan, S. S. (2023). Phase alignment of low-frequency neural activity to the amplitude envelope of speech reflects evoked responses to acoustic edges, not oscillatory entrainment. *The Journal of Neuroscience*, *43*, 3909–3921. https://doi.org/10.1523/JNEUROSCI.1663-22.2023

Overath, T., McDermott, J. H., Zarate, J. M., & Poeppel, D. (2015). The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. *Nature Neuroscience*, *18*, 903–911. https://doi.org/10.1038/nn.4021

Park, H., Ince, R. A. 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*, 1649–1653. https://doi.org/10.1016/j.cub.2015.04.049

Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., & Gross, J. (2018). Representational interactions during audiovisual speech entrainment: Redundancy in left posterior superior temporal gyrus and synergy in left motor cortex. *PLoS Biology*, *16*, e2006558. https://doi.org/10.1371/journal.pbio.2006558

Park, H., Kayser, C., Thut, G., & Gross, J. (2016). Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. *eLife*, *5*, e14521. https://doi.org/10.7554/eLife.14521

Peelle, J., & Davis, M. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*, 320. https://doi.org/10.3389/fpsyg.2012.00320

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*, 1378–1387. https://doi.org/10.1093/cercor/bhs118

Pitt, M.A., Szostak, C. & Dilley, L.C. (2016). Rate dependent speech processing can be speech specific: Evidence from the perceptual disappearance of words under changes in context speech rate. *Atten Percept Psychophys 78*, 334–345. https://doi.org/10.3758/s13414-015-0981-7

Poeppel, D., Assaneo, M.F. Speech rhythms and their neural foundations.*Nat Rev Neurosci 21*, 322–334 (2020). https://doi.org/10.1038/s41583-020-0304-4

Prystauka, Y., & Lewis, A. G. (2019). The power of neural oscillations to inform sentence comprehension: A linguistic perspective. *Lang & Ling Compass*, *13*, e12347. https://doi.org/10.1111/lnc3.12347

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*, 265–292. https://doi.org/10.1016/S0010-0277(99)00058-X

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*, 718–724. https://doi.org/10.1038/nn.2331

Reetzke, R., Gnanateja, G. N., & Chandrasekaran, B. (2021). Neural tracking of the speech envelope is differentially modulated by attention and language experience. *Brain and Language*, *213*, 104891. https://doi.org/10.1016/j.bandl.2020.104891

Riecke, L., Formisano, E., Herrmann, C. S., & Sack, A. T. (2015). 4-Hz transcranial alternating current stimulation phase modulates hearing. *Brain Stimulation*, *8*, 777–783. https://doi.org/10.1016/j.brs.2015.04.004

Riecke, L., Formisano, E., Sorger, B., Başkent, D., & Gaudrain, E. (2018). Neural entrainment to speech modulates speech intelligibility. *Current Biology*, *28*, 161–169.e5. https://doi.org/10.1016/j.cub.2017.11.033

Rimmele, J. M., Zion Golumbic, E., Schröger, E., & Poeppel, D. (2015). The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. *Cortex*, *68*, 144–154. https://doi.org/10.1016/j.cortex.2014.12.014

Saberi, K., & Hickok, G. (2023). Forward entrainment: Evidence, controversies, constraints, and mechanisms. *Psychonomic Bulletin & Review*, *30*, 803–821. https://doi.org/10.3758/s13423-022-02220-y

Sassenhagen, J. (2019). How to analyse electrophysiological responses to naturalistic language with time-resolved multiple regression. *Language, Cognition and Neuroscience*, *34*, 474–490. https://doi.org/10.1080/23273798.2018.1502458

Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., & Weiller, C. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences*, *105*, 18035–18040. https://doi.org/10.1073/pnas.0805234105

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, *32*, 9–18. https://doi.org/10.1016/j.tins.2008.09.012

Schwartz, J. L., & Savariaux, C. (2014). No, there is no 150 ms lead of visual speech on auditory speech, but a range of audiovisual asynchronies varying from small audio lead to large audio lag. *PLoS Computational Biology*, *10* (7), e1003743.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*(Pt 12), 2400–2406. https://doi.org/10.1093/brain/123.12.2400

Seifart, F., Strunk, J., Danielsen, S., Hartmann, I., Pakendorf, B., Wichmann, S., Witzlack-Makarevich, A., de Jong, N. H., & Bickel, B. (2018). Nouns slow down speech across structurally and culturally diverse languages. *Proceedings of the National Academy of Sciences*, *115*, 5720–5725. https://doi.org/10.1073/pnas.1800708115

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304. https://doi.org/10.1126/science.270.5234.303

Sobin, C., & Alpert, M. (1999). Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. *Journal of Psycholinguistic Research*, *28*, 347–365. https://doi.org/10.1023/A:1023237014909

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *The Journal of Neuroscience*, *32*, 8443–8453. https://doi.org/10.1523/JNEUROSCI.5069-11.2012

Stehwien, S., & Meyer, L. (2022). Short-term periodicity of prosodic phrasing: Corpus-based evidence. *Proc. Speech Prosody*, 693–698.

Steinschneider, M., Nourski, K. V., & Fishman, Y. I. (2013). Representation of speech in human auditory cortex: Is it special? *Hearing Research, Communication Sounds and the Brain: New Directions and Perspectives*, *305*, 57–73. https://doi.org/10.1016/j.heares.2013.05.013

Strauß, A., & Schwartz, J.-L. (2017). The syllable in the light of motor skills and neural oscillations. *Language, Cognition and Neuroscience*, *32*, 562–569. https://doi.org/10.1080/23273798.2016.1253852

Strauß, A., Wöstmann, M., & Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00350

Tanaka, H., Fujita, N., Watanabe, Y., Hirabuki, N., Takanashi, M., Oshiro, Y., & Nakamura, H. (2000). Effects of stimulus rate on the auditory cortex using fMRI with 'sparse' temporal sampling. *Neuroreport*, *11*, 2045–2049. https://doi.org/10.1097/00001756-200006260-00047

ten Oever, S., Carta, S., Kaufeld, G., & Martin, A. E. (2022). Neural tracking of phrases in spoken language comprehension is automatic and task-dependent. *eLife*, *11*, e77468. https://doi.org/10.7554/eLife.77468

ten Oever, S., Schroeder, C. E., Poeppel, D., van Atteveldt, N., & Zion-Golumbic, E. (2014). Rhythmicity and cross-modal temporal cues facilitate detection. *Neuropsychologia*, *63*, 43–50. https://doi.org/10.1016/j.neuropsychologia.2014.08.008

Teng, X., & Poeppel, D. (2020). Theta and gamma bands encode acoustic dynamics over wide-ranging timescales. *Cerebral Cortex*, *30*, 2600–2614. https://doi.org/10.1093/cercor/bhz263

Teng, X., Tian, X., Doelling, K., & Poeppel, D. (2018). Theta band oscillations reflect more than entrainment: Behavioral and neural evidence demonstrates an active chunking process. *The European Journal of Neuroscience*, *48*, 2770–2782. https://doi.org/10.1111/ejn.13742

Teng, X., Tian, X., Rowland, J., & Poeppel, D. (2017). Concurrent temporal channels for auditory processing: Oscillatory neural entrainment reveals segregation of function at different scales. *PLoS Biology*, *15*, e2000812. https://doi.org/10.1371/journal.pbio.2000812

Teoh, E. S., Cappelloni, M. S., & Lalor, E. C. (2019). Prosodic pitch processing is represented in delta-band EEG and is dissociable from the cortical tracking of other acoustic and phonetic features. *European Journal of Neuroscience*, *50*, 3831–3842. https://doi.org/10.1111/ejn.14510

Thorne, J. D., & Debener, S. (2014). Look now and hear what's coming: On the functional role of cross-modal phase reset. *Hearing Research*, *307*, 144–152. https://doi.org/10.1016/j.heares.2013.07.002

Thut, G., Schyns, P., & Gross, J. (2011). Entrainment of perceptually relevant brain oscillations by non-invasive rhythmic stimulation of the human brain. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00170

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*, 628–639. https://doi.org/10.1121/1.4807565

Van Ackeren, M. J., Barbero, F. M., Mattioni, S., Bottini, R., & Collignon, O. (2018). Neuronal populations in the occipital cortex of the blind synchronize to the temporal dynamics of speech. *eLife*, *7*, e31640. https://doi.org/10.7554/eLife.31640

van Bree, S., Alamia, A., & Zoefel, B. (2022). Oscillation or not-why we can and need to know (commentary on Doelling and Assaneo, 2021). *The European Journal of Neuroscience*, *55*, 201–204. https://doi.org/10.1111/ejn.15542

van Bree, S., Sohoglu, E., Davis, M. H., & Zoefel, B. (2021). Sustained neural rhythms reveal endogenous oscillations supporting speech perception. *PLoS Biology*, *19*, e3001142. https://doi.org/10.1371/journal.pbio.3001142

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences*, *102*, 1181–1186. https://doi.org/10.1073/pnas.0408949102

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, *45*, 598–607. https://doi.org/10.1016/j.neuropsychologia.2006.01.001

VanRullen, R. (2016). Perceptual cycles. *Trends in Cognitive Sciences*, *20*, 723–735. https://doi.org/10.1016/j.tics.2016.07.006

VanRullen, R., Zoefel, B., & Ilhan, B. (2014). On the cyclic nature of perception in vision versus audition. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *369*, 20130214. https://doi.org/10.1098/rstb.2013.0214

Varnet, L., Ortiz-Barajas, M. C., Erra, R. G., Gervain, J., & Lorenzi, C. (2017). A cross-linguistic study of speech modulation spectra. *The Journal of the Acoustical Society of America*, *142*, 1976–1989. https://doi.org/10.1121/1.5006179

Voss, P., & Zatorre, R. J. (2012). Organization and reorganization of sensory-deprived cortex. *Current Biology*, *22*, R168–R173. https://doi.org/10.1016/j.cub.2012.01.030

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

Weissbart, H., Kandylaki, K. D., & Reichenbach, T. (2020). Cortical tracking of surprisal during continuous speech comprehension. *Journal of Cognitive Neuroscience*, *32*, 155–166. https://doi.org/10.1162/jocn_a_01467

White, L., Mattys, S. L., & Wiget, L. (2012). Language categorization by adults is based on sensitivity to durational cues, not rhythm class. *Journal of Memory and Language*, *66*, 665–679. https://doi.org/10.1016/j.jml.2011.12.010

Wilsch, A., Neuling, T., Obleser, J., & Herrmann, C. S. (2018). Transcranial alternating current stimulation with speech envelopes modulates speech comprehension. *NeuroImage*, *172*, 766–774. https://doi.org/10.1016/j.neuroimage.2018.01.038

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic Bulletin & Review*, *12*, 957–968. https://doi.org/10.3758/BF03206432

Wolff, A., Berberian, N., Golesorkhi, M., Gomez-Pilar, J., Zilio, F., & Northoff, G. (2022). Intrinsic neural timescales: Temporal integration and segregation. *Trends in Cognitive Sciences*, *26*, 159–173. https://doi.org/10.1016/j.tics.2021.11.007

Zaehle, T., Lenz, D., Ohl, F. W., & Herrmann, C. S. (2010). Resonance phenomena in the human auditory cortex: Individual resonance frequencies of the cerebral cortex determine electrophysiological responses. *Experimental Brain Research*, *203*, 629–635. https://doi.org/10.1007/s00221-010-2265-8

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., Goodman, R. R., Emerson, R., Mehta, A. D., Simon, J. Z., Poeppel, D., & Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron*, *77*, 980–991. https://doi.org/10.1016/j.neuron.2012.12.037

Zoefel, B. (2021). Visual speech cues recruit neural oscillations to optimise auditory perception: Ways forward for research on human communication. *Current Research in Neurobiology*, *2*, 100015. https://doi.org/10.1016/j.crneur.2021.100015

Zoefel, B., Allard, I., Anil, M., & Davis, M. H. (2020). Perception of rhythmic speech is modulated by focal bilateral transcranial alternating current stimulation. *Journal of Cognitive Neuroscience*, *32*, 226–240. https://doi.org/10.1162/jocn_a_01490

Zoefel, B., Archer-Boyd, A., & Davis, M. H. (2018). Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. *Current Biology*, *28*, 401–408.e5. https://doi.org/10.1016/j.cub.2017.11.071

Zoefel, B., Costa-Faidella, J., Lakatos, P., Schroeder, C. E., & VanRullen, R. (2017). Characterization of neural entrainment to speech with and without slow spectral energy fluctuations in laminar recordings in monkey A1. *NeuroImage*, *150*, 344–357. https://doi.org/10.1016/j.neuroimage.2017.02.014

Zoefel, B., Gilbert, R. A., & Davis, M. H. (2023). Intelligibility improves the perception of speech rhythm. *PLoS ONE*, *18*, e0279024.

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*. https://doi.org/10.3389/fnins.2018.00095

Zoefel, B., & VanRullen, R. (2015a). The role of high-level processes for oscillatory phase entrainment to speech sound. *Frontiers in Human Neuroscience*, *9*, 651. https://doi.org/10.3389/fnhum.2015.00651

Zoefel, B., & VanRullen, R. (2015b). Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. *The Journal of Neuroscience*, *35*, 1954–1964. https://doi.org/10.1523/JNEUROSCI.3484-14.2015

Zoefel, B., & VanRullen, R. (2016). EEG oscillations entrain their phase to high-level features of speech sound. *NeuroImage*, *124*, 16–23. https://doi.org/10.1016/j.neuroimage.2015.08.054

Zou, J., Feng, J., Xu, T., Jin, P., Luo, C., Zhang, J., Pan, X., Chen, F., Zheng, J., & Ding, N. (2019). Auditory and language contributions to neural encoding of speech features in noisy environments. *NeuroImage*, *192*, 66–75. https://doi.org/10.1016/j.neuroimage.2019.02.047

Zuk, N. J., Murphy, J. W., Reilly, R. B., & Lalor, E. C. (2021). Envelope reconstruction of speech and music highlights stronger tracking of speech at low frequencies. *PLoS Computational Biology*, *17*, e1009358. https://doi.org/10.1371/journal.pcbi.1009358

How to cite this article: Zoefel, B., & Kösem, A. (2024). Neural tracking of continuous acoustics: properties, speech-specificity and open questions. *European Journal of Neuroscience*, *59*(3), 394–414. https://doi.org/10.1111/ejn.16221