5 Evaluating Neural Tracking of Rhythmic Information in Speech: Some Caveats and Challenges

VÉRONIQUE BOULENGER

5.1 Introduction

As expert listeners, we do not always notice how complex speech decoding is. Understanding speech means handling a continuous stream with unclear acoustic boundaries between elements and in which approximately 200 words occur per minute. Yet we have to decode this information in a handful of seconds. In the highly sophisticated cascade of processes that enables the achievement of this, one of the first steps is to segment the incoming signal into discrete elements onto which linguistic information can then be mapped. To this aim, listeners exploit a fundamental property of speech, its rhythmicity. Rhythm also characterizes intrinsic brain activity: without the need for any external stimulus, neuronal populations naturally show cyclic variations of excitability at different frequencies (see also Chapter 3). As nature knows best, brain oscillations are able to synchronize to the different timescales in speech, a process thought to be central to discretize the signal into phoneme-, syllable-, and word-sized chunks. After sketching the importance of rhythm in speech and introducing current oscillatory models of speech perception, this chapter reviews neurophysiological evidence for the coupling of neural oscillations to speech modulations, syllabic rhythm in particular. Our aim is not to provide an exhaustive review of the literature but to draw attention to topics that may help the reader to better grasp the neural oscillatory dynamic underlying speech perception. We specifically consider speech rate variations as good candidates to examine brain tracking of speech rhythm and also discuss the functional relationship between neural oscillations and speech intelligibility. We finally review some evidence supporting a role of motor regions as temporal predictors for speech parsing.

5.2 The Processing Benefits of (Quasi-)Rhythmic Information in Speech

The (quasi-)rhythmicity of speech is central for listeners to parse the continuous acoustic stream into distinct linguistic elements. Despite not being strictly periodic, the temporal structure of

https://doi.org/10.1017/9781009295888.007 Published online by Cambridge University Press

connected speech shows acoustic cues that occur at fairly regular intervals (see also Chapters 9 and 14). Syllables are produced at a frequency of ~5 Hz across languages (Poeppel and Assaneo, 2020), and stressed syllables marking phrasal units occur approximately every 500 ms, at 2 Hz. These recurrent speech patterns over time serve as landmarks to predict the upcoming critical information, therefore facilitating speech segmentation and ultimately comprehension. Early on during language development, listeners exploit and take advantage of this rhythmical organization of prominences in speech to extract phonological, lexical, and morphosyntactic information (Christophe et al., 2003). English infants are, for instance, able to segment disyllabic words from unfolding speech based on stress cues, namely the strong–weak syllabic structure typical of words from their native language (Jusczyk et al., 1993). Sidiras et al. (2017) additionally revealed that 10-year-old Greek-native children better identified two-syllable words in a babble when these were presented in synchrony with a previous isochronous rhythmic sequence, that is, when a prominent element was expected to occur. The benefit of rhythm for speech processing is also manifest in adults. Detecting a target phoneme in disyllabic English words is easier if stressed syllables carrying this phoneme are presented at regular temporal intervals (Quené and Port, 2005; see also Roncaglia-Denissen et al., 2013, for evidence of syntactic processing). In the same vein, listening to nonverbal rhythmic sequences improves subsequent phonological processing in nonwords, provided that target phonemes appear on beat with the rhythmical prime (Cason and Schön, 2012). The study by Ghitza and Greenberg (2009) perfectly illustrates the use of temporal expectations in speech processing. They assessed the intelligibility of sentences time-compressed to one third of their original duration, with silent intervals of varying durations inserted periodically or aperiodically between consecutive speech fragments. Sentence identification was maximal when silences were inserted periodically so as to restore the initial temporal structure of speech. In other words, despite the fact speech fragments were compressed and had limited intelligibility, hearing them at the time they were expected was sufficient to increase intelligibility (see also Ghitza, 2014; Penn et al., 2018). What counts for successful speech decoding is therefore not only the content of the information but when this information is likely to happen. Along this line, the surrounding speech rate can shape phonemic perception, a process known as rate normalization. Hearing a sentence at a fast rate biases subsequent perception towards longer speech sounds compared to when the sentence is produced at a slower rate (Bosker, 2017; Bosker and Ghitza, 2018; Dilley and Pitt, 2010; Maslowski et al., 2019; Reinisch, 2016; Reinisch and Sjerps, 2013; see also Lamekina and Meyer, 2022). As an example, Dutch words embedding an ambiguous vowel between short /ɑ/ and long /a:/ are identified as /ta:k/ ("task") following a fast-rate carrier sentence, but as /tɑk/ ("branch") after a sentence at a slower rate (Kösem et al., 2018). The rationale here is that listeners make some expectations about the occurrence of the future syllables based on the syllabic rate in the previous carrier sentence. For a fast rate, short syllables are expected, hence leading to an overestimation of the target vowel duration: vowels are perceived as longer than they actually are.

Altogether, these lines of research emphasize the predictive function of (quasi-)rhythmic information in speech. One computational mechanism that has been suggested to underpin the establishment of these predictions is the synchronization of brain rhythmic activity to speech temporal structure.

https://doi.org/10.1017/9781009295888.007 Published online by Cambridge University Press

5.3 Neuronal Oscillations as a Key Mechanism for Speech Parsing

Endogenous oscillatory brain activity (Buzsáki and Draguhn, 2004; Fries, 2005, 2015; Varela et al., 2001) plays a crucial role in speech processing (see Chapter 3). By synchronizing, or entraining, to the quasi-rhythmic modulations in the acoustic signal, neuronal oscillations discretize the incoming stream into "packets" of different sizes that can then be mapped onto linguistic units. Contemporary neurocognitive models of speech processing advocate a close correspondence between the timescales in speech (Rosen, 1992) and auditory cortical activity in the delta (~1–4 Hz), theta (~4–8 Hz), and low-gamma (~25–45 Hz) frequency bands (Ghitza, 2011; Giraud and Poeppel, 2012; Peelle and Davis, 2012; Poeppel, 2003; see also Chapter 9). The phase of delta band oscillations parses the signal into large units such as words and phrases, thus encoding prosodic information, whereas theta activity phase-locks to the low-frequency modulations in the amplitude envelope, peaking at 4–7 Hz and conveying syllabic information. On the other hand, the low-gamma amplitude encodes faster modulations pertaining to phonetic features such as voicing and formant transitions. The parsing into speech units of varying sizes furthermore operates simultaneously but hierarchically through theta/gamma nesting, where the phase of theta oscillations drives gamma amplitude. Substantial evidence from electro- and magnetoencephalography (E/MEG) supports this brain dynamic during speech processing, highlighting how neuronal oscillations at different frequencies orchestrate to deal with the multi-time resolution of speech. In the following sections, we review the main existing findings for brain-to-speech coupling, focusing on theta band tracking of syllabic information in the case of speech rate variations. We then discuss whether and how oscillatory coupling is modulated by and contributes to speech intelligibility, a topic that is still debated in the literature. Lastly, we review evidence that beyond the auditory cortex, brain motor regions functionally contribute to speech decoding via top-down temporal predictions.

5.3.1 Cortical Tracking of Syllabic Structure: The Case of Speech Rate Variations

Since the coupling of theta band cortical activity to the speech envelope is thought to track syllabic modulations, one way to put it to the test has been to vary the syllable rate of incoming speech. Speech rate variations are ubiquitous in everyday life, and increasing speech rate, either naturally or artificially, is known to hinder speech intelligibility (Gordon-Salant et al., 2014; Guiraud et al., 2018; Janse, 2004; Janse et al., 2003; Park and Jang, 2012). Studies in young normal listeners reported that 50% intelligibility (i.e., 50% of correct identification) was reached when compressed speech was three to four times faster than the normal rate, namely at ~12 syllables/s in German and Dutch and ~16 syllables/s in Mandarin (Meng et al., 2019; Schlueter et al., 2014; Versfeld and Dreschler, 2002). Fast speech therefore increases processing demands, as quantified by higher-rated listening effort and larger pupil dilation with respect to slower speech (Koch and Janse, 2016; Müller et al., 2019). Despite comprehension of accelerated speech being challenging at first, other work showed that listeners are able to adapt relatively quickly (Adank and Devlin, 2010; Adank and Janse, 2009; Golomb et al., 2007; Peelle and Wingfield, 2005). Speech perception indeed improves after short exposure to artificially speeded sentences, and this adaptation is robust to changes in speakers' characteristics (Dupoux and

https://doi.org/10.1017/9781009295888.007 Published online by Cambridge University Press

Green, 1997) and in languages provided they share some phonological and rhythmic properties (Pallier et al., 1998; Sebastián-Gallés et al., 2000).

Neural tracking of the slow envelope fluctuations may be one of the primary mechanisms used by the brain to adapt to speech rate variations, both within and between speakers. Most E/MEG studies that measured brain coupling to syllable rate changes have accelerated – and sometimes decelerated – speech artificially. In their seminal work, Ahissar et al. (2001) reported that the auditory cortex phase-locked to amplitude modulations in English sentences compressed to 75% and 50% of their original duration, that is, at a syllable rate ranging between ~4 and ~6 syllables/s. However, for higher compression ratios (35% and 20%) with syllable rates between ~9 and ~14 syllables/s, neural oscillations no longer coupled to the envelope. These conditions were also the ones showing a drop of sentence intelligibility (see Section 5.3.2). These data seem to fit Ghitza's assumption that speech comprehension is constrained by our capacity to decode syllables within a theta cycle, that is, at a maximal frequency of ~9 Hz (Ghitza, 2011, 2014). The results by Pefkou et al. (2017), however, challenge this view as oscillations in auditory regions tracked the envelope of French sentences time-compressed by a factor of 3 (up to 10–14 Hz), beyond the upper limit of the canonical theta band (see also Nourski et al., 2009, and Section 5.3.2).

Without pushing the limits this far, other studies converge to show that oscillatory activity dynamically adapts to changes in speech rate. In the work by Lizarazu et al. (2019), maximal brain-to-speech coupling in bilateral auditory regions was found at 4.5–7.5 Hz (peaking at 5.6 Hz) for normal-rate sentences in Spanish (~173 words/min). These values slightly decreased to 4–6.5 Hz (peaking at 4.7 Hz) for sentences that were artificially slowed down (~139 words/min), but increased to 5.5–8.5 Hz (peaking at 6.6 Hz) for accelerated sentences (~208 words/min). Note that the syllable rate was unfortunately not provided by the authors, making the comparison with other work rather difficult. Given the coupling frequencies reported, the syllable rate, however, seemed to remain within the theta band boundaries. Theta/gamma coupling also varied with speech rate in this study. The phase of theta oscillations in the three previously mentioned frequency ranges modulated the amplitude of gamma oscillations at 20–37 Hz for normal-rate speech, 18–34 Hz for decelerated speech, and 23–43 Hz for accelerated speech. Such rate-dependent gamma activity was assumed to reflect the encoding of fine acoustic information in speech.

A further step was taken by Kösem et al. (2018) who nicely demonstrated that not only do neural oscillations adjust to speech rate but that coupling persists after the changes, thus biasing subsequent speech comprehension. Echoing the aforementioned behavioral studies on rate normalization (see Section 5.2), their MEG results revealed that listening to a carrier sentence (in Dutch) artificially accelerated at 5.5 Hz entrained neural oscillations at this frequency in a sustained manner, so that an ambiguous vowel in a following normal-rate word was perceived as long (e.g., /a:/). On the contrary, when the syllable rate of the carrier sentence was decelerated to 3 Hz, the subsequent vowel was perceived as short (e.g., /ɑ/). This study exemplifies the capacity of our brain to generate predictions based on the temporal pattern of incoming speech, thus functionally contributing to its decoding.

To the best of our knowledge, studies investigating the neural tracking of natural speech rate changes have been scarce, possibly because of methodological challenges. Yet examining the

https://doi.org/10.1017/9781009295888.007 Published online by Cambridge University Press

case of naturally accelerated (or decelerated) speech may prove highly informative to better grasp the oscillatory dynamic underlying naturalistic speech perception (see Alexandrou et al., 2020, for similar arguments). Naturally speaking faster induces additional spectro-temporal changes in the speech signal compared to time compression, in particular by increasing coarticulation and assimilation between speech segments (Berry, 2011; Janse et al., 2003). This complexifies speech decoding for the listeners (Janse, 2004), and accordingly, processing of natural and artificial speech rate variations may involve different oscillatory mechanisms, at least partly. Alexandrou et al. (2018) compared brain oscillatory coupling to naturally produced speech extracts in Finnish, with syllable rates ranging from ~2.6 (slow) to ~4.7 (normal) to ~6.8 (fast) syllables/s. In the delta (2–4 Hz) and theta (4–7 Hz) bands used for analysis, normal-rate speech elicited stronger synchronization than slower and faster speech in the bilateral superior temporal cortex (for delta) and the right paracentral lobule (for theta). Slow speech increased theta coupling in the right parietal cortex with respect to normal-rate speech. Based on the study by Lizarazu et al. (2019), it is quite surprising that delta tracking was not stronger for slow-rate speech compared to the two other conditions. Similarly, fast-rate speech did not enhance brain coupling in the high theta band as one could have expected. One strength of this MEG study is undeniably the use of connected speech that was spontaneously produced at different rates, whereas most other work that examined neural tracking of speech rate variations included isolated sentences or read-aloud texts. One caveat, however, may be that coupling was measured in fixed, generic frequency bands, potentially preventing the researchers from observing changes in frequency coupling as a function of the syllable rate. The average normal and fast rates in this study were both circumscribed to the theta range, towards the lower (~4.7 Hz) or the upper (~6.8 Hz) band limits, respectively. The use of more flexible frequency windows for analysis might have revealed more specific coupling patterns depending on the two speaking rates (see Keitel et al., 2018). In an MEG study (Hincapié Casas et al., 2021), we used this latter approach to compare neural tracking of French sentences naturally produced at a normal rate (~6.7 syllables/s) or at a fast rate (~9.1 syllables/s), or of sentences time-compressed to the same fast rate. We examined brain-to-speech coupling in two frequency bands defined from the spectral power of the sentence envelopes in each condition, namely 6.25 Hz $\pm$ 1 Hz for normal-rate speech and 8.75 $\pm$ 1 Hz for fast-rate/compressed speech. Our results showed that neural oscillations in right auditory and motor regions were tuned to natural speech rate variations at frequencies specifically matching the syllabic rate. No significant coupling was observed for time-compressed speech in the corresponding frequency band (8.75 $\pm$ 1 Hz), despite the fact sentences had a similar rate as natural fast speech. We suggested a possible preference of cortical oscillations to align to naturally produced speech rather than to artificially accelerated speech, which is not physiologically plausible. To our knowledge, this study is the first to provide evidence for a specific oscillatory signature of natural fast speech compared to artificially accelerated speech (see also Section 5.3.3 for a discussion of the motor cortex involvement).

Altogether, these findings underscore the need for future studies to consider natural speech rate variations in connected speech to refine neurocognitive models of speech perception. A further issue to dig into, and which is actively debated, is whether the rhythmic neural response to the speech envelope truly involves endogenous oscillatory generators or whether it

https://doi.org/10.1017/9781009295888.007 Published online by Cambridge University Press

reflects steady-state evoked responses to acoustic landmarks in the speech stream. Evidence for sustained oscillatory activity after speech offset (Kösem et al., 2018; van Bree et al., 2021) and for neural responses at frequencies that are not physically present in the acoustic signal but that match mentally constructed linguistic units (Ding et al., 2016; Zoefel and VanRullen, 2016) clearly speaks in favor of the oscillatory model (Zoefel et al., 2018, for a review; see also Hincapié Casas et al., 2021, for a lack of power increase accompanying enhanced phase-locking to sentences). Other studies nevertheless reported that cortical coupling to speech is mainly driven by a linear convolution of evoked responses at the frequency of acoustic edges in the input (Oganian and Chang, 2019; Oganian et al., 2023; Novembre and Iannetti, 2018; Zou et al., 2021). Future work should systematically attempt to delineate the contribution of oscillatory versus evoked activity to speech neural tracking, bearing in mind that the two may not be mutually exclusive.

5.3.2 Brain-to-Speech Coupling and Intelligibility: A Bidirectional Relationship?

Increasing speaking rate can obviously reduce speech intelligibility up to a point where the listener is no longer able to comprehend. The question therefore arises as to whether neural tracking of speech varies with intelligibility and contributes to it. A number of E/MEG studies tackled this issue to better characterize the functional role of neural oscillations in speech decoding. The already mentioned work by Ahissar et al. (2001) showed that cortical coupling to speech significantly declined for time-compressed sentences that were not intelligible. The authors hence proposed that speech comprehension depends on neural tracking, which would only be possible if the speech rate is close to the intrinsic oscillatory frequency of the auditory cortex (see Ghitza, 2011, 2014).

According to Ahissar's interpretation, speech becomes unintelligible because neural tracking fails, meaning that effective brain-to-speech coupling is mandatory for successful speech decoding. The reverse question could, however, also be raised: Does oscillatory coupling fail because speech is not intelligible? To put it differently, does this coupling need intelligible speech to occur? One way to address the latter question is to test if neural tracking is also found for unintelligible stimuli, which would be evidence that it does not hinge on speech intelligibility. This would in turn provide insights, at least to some extent, to the first point, namely whether brain coupling is necessary or not for speech decoding. Neural tracking of unintelligible speech would indeed suggest that this process alone cannot fully account for intelligibility. Undertaking this issue also requires examining whether speech can be understood in the absence of neural tracking.

5.3.2.1 Does Neural Tracking Need Intelligible Speech to Occur?

Nourski et al. (2009) reported in an electrocorticography (ECoG) study that Heschl's gyrus only phase-locked to intelligible time-compressed sentences in English, in agreement with Ahissar et al. (2001). However, further analyses revealed that high-frequency (70–250 Hz) power in the same region was modulated by the speech envelope irrespective of compression ratios, that is, even when speech was not understood. This casts doubt on the fact that neural

https://doi.org/10.1017/9781009295888.007 Published online by Cambridge University Press

tracking of speech modulations relies on intelligibility. Reconciling these contradictory findings, another ECoG study (Davidesco et al., 2018) showed that the ability of neural oscillations to synchronize to intelligible and/or unintelligible speech may depend on the cortical site along the processing hierarchy. Whereas low-level early-auditory areas, at the vicinity of the primary auditory cortex, tracked time-compressed speech outside the intelligibility range, higher-order areas such as secondary auditory areas in the superior temporal gyrus and the inferior frontal gyrus only synchronized to intelligible speech. E/MEG studies using time compression, time reversion, or vocoding provided further evidence that the brain tracks envelope modulations in both intelligible and unintelligible speech (Luo and Poeppel, 2007; Pefkou et al., 2017; Zoefel and VanRullen, 2016; see Kösem and van Wassenhove, 2016, for a review). Howard and Poeppel (2010) showed that the theta band phase in the auditory cortex could equally discriminate between spoken English sentences and their time-reversed, non-intelligible counterparts. Converging results were found by Millman et al. (2015), with similar coupling to physically identical sentences that solely differed in intelligibility (see also Baltzell et al., 2017). Kösem et al. (2023) also recently failed to find differential delta/theta neural tracking of noise-vocoded Dutch sentences that were intelligible or not. Altogether, these findings suggest that brain-to-speech coupling does not need intelligible speech to occur, and that it may be driven by acoustic cues rather that by linguistic information per se.

The conclusion is, however, not so straightforward given evidence that neural tracking of speech varies with intelligibility, as reflected by stronger phase-locking to the envelope for intelligible than for unintelligible speech (Dimitrijevic et al., 2019; Ding et al., 2014; Doelling et al., 2014; Pérez et al., 2015). In the MEG study by Gross et al. (2013), delta/theta phase-locking in bilateral auditory regions was enhanced for intact compared to reversed stories in English. Stronger theta phase/gamma amplitude coupling was additionally observed when speech was intelligible. The same year, Peelle et al. (2013) also reported increased theta coupling in the left temporal cortex to intelligible vocoded English sentences more than to unintelligible ones, despite having the same spectral complexity. They suggested that although neural oscillations can be entrained by speech lacking intelligibility, such a process would not only depend on bottom-up synchronization to the acoustic rhythmic fluctuations but also to higher-level linguistic content. Rimmele et al. (2015) reached a somewhat similar conclusion. In a multi-talker paradigm, they showed similar neural tracking (2–8 Hz) of intact and vocoded sentences as long as participants ignored them. When speech was attended, stronger coupling was found only in the intact condition. According to the authors, cortico-acoustic coupling mainly reflects brain responses to slow acoustic modulations in the envelope, but for attended speech, processing of the fine structure can enhance the temporal precision of tracking (see also Zion Golumbic et al., 2013). Linguistic information therefore seems to affect neural oscillatory coupling, a process that may operate nonlinearly, as revealed by Hauswald et al. (2022). In this study, low-frequency (1–7 Hz) coupling in bilateral middle temporal and left frontal regions followed an inverted U-shaped curve. It was maximal for challenging yet intelligible (moderately vocoded) speech in German, reflecting the enhanced use of speech temporal structure, but it decreased to its lowest value for very easily (intact) or barely understandable (strongly vocoded) speech. Note that by using different parameters on the same dataset and by separating periodic from aperiodic components in the coupling spectrum, the research team did not

https://doi.org/10.1017/9781009295888.007 Published online by Cambridge University Press

replicate these results exactly (Schmidt et al., 2023). They instead showed a gradual decrease of neural tracking with the decline in speech intelligibility (in line with Gross et al., 2013; Peelle et al., 2013), an effect yielded by the aperiodic components. The results on the periodic components exhibited a different pattern. As vocoded speech became less intelligible, neural tracking shifted from the syllabic rate (linguistic level) to the modulation rate of the envelope (acoustic level). This was shown by an increase of the coupling frequency and a narrowing of the frequency bandwidth (i.e., frequency tuning) around the higher frequencies of acoustic modulations for more vocoded, less intelligible speech. Accordingly, whereas neural oscillations would track both syllabic and acoustic information in the envelope of intact speech (leading to an increased frequency bandwidth), listeners would preferably rely on acoustic fluctuations when linguistic information such as syllable boundaries is more difficult to extract (see also Verschueren et al., 2022, for acoustic versus linguistic tracking with speech rate increase). Disentangling the syllable rate from the envelope modulation rate in future studies (see Zhang et al., 2023), including those on speech rate variations, may prove efficient to better characterize the complex relationship between brain oscillations and low-level acoustic information on the one hand and higher-level linguistic content on the other.

Coming back to our initial question as to whether intelligibility is a prerequisite for neural tracking, the reviewed evidence is rather negative. Brain-to-speech coupling is also found for unintelligible speech, and it may thus be largely driven by acoustic modulations in the speech signal. Still, the coupling strength can be enhanced by intelligible speech. This suggests that linguistic content plays an important role in the oscillatory dynamic, which cannot be seen as a purely rhythmic acoustic cortical response.

5.3.2.2 Is Neural Tracking Necessary for Successful Speech Decoding?

Determining whether neural coupling is a limiting factor for comprehension may appear a bit more complex. The fact that brain oscillatory activity synchronizes to speech that is not intelligible, as reviewed above, provides some arguments that speech does not become intelligible only because neural oscillations phase-lock to its (quasi-)rhythmic fluctuations. Still, this does not allow to firmly conclude that neural tracking is not a necessary ingredient. We would in fact like to underline that as far as we are aware, no evidence for intelligible speech without concurrent auditory cortical coupling has been reported. Such coupling could thus be essential, albeit not sufficient, for speech decoding.

Convincing evidence for a causal role of neural tracking in speech processing and intelligibility comes from studies using transcranial alternating current stimulation, or tACS (Kadir et al., 2020; Keshavarzi et al., 2020; Riecke et al., 2018; Wilsch et al., 2018; Zoefel et al., 2018, 2020). Zoefel et al. (2018) showed that tACS (~3 Hz) over left auditory regions decreased the functional magnetic resonance imaging (fMRI) response in the bilateral superior and middle temporal gyri for vocoded yet intelligible speech (i.e., sequences of monosyllabic words at ~3 Hz). The behavioral performance in an irregularity detection task was, however, unaffected. A follow-up study (Zoefel et al., 2020) nonetheless revealed that bilateral, but not left-lateralized, tACS disrupted word report accuracy, suggesting that changes in neural oscillatory activity can impair speech perception. Conversely, enhanced speech comprehension has been

https://doi.org/10.1017/9781009295888.007 Published online by Cambridge University Press

reported after stimulation of neural oscillatory activity (Riecke et al., 2018). By applying a current carrying slow envelope modulations (4 Hz) over the bilateral auditory cortex, Riecke et al. could improve Dutch sentence recognition in a two-talker situation. A second experiment confirmed that biasing neural oscillations towards speech temporal structure benefitted speech processing. In a single-talker situation, participants were instructed to identify barely understandable sentences in which low-frequency envelope information (<16 Hz) had been removed. When auditory regions were stimulated with a current shaped with exactly this rhythmic information, sentence intelligibility increased, especially when the current preceded speech stimuli by 375 ms. Using electroencephalography (EEG), the same group then showed that delivering envelope information through vibrotactile stimulation applied to the participants' fingers increased neural speech tracking but failed to improve behavioral outcome during sentence repetition (Riecke et al., 2019; see their discussion for methodological considerations). Positive evidence was yet provided by Guilleminot and Reichenbach (2022): Tactile stimulation delivered at the syllable rate (theta frequency 4.5 Hz) enhanced both brain-to-speech coupling in bilateral auditory regions and intelligibility of speech in noise.

Overall, these findings suggest that neural coupling to speech functionally contributes to and can aid speech perception. In this view, they open new avenues to improve speech processing in hearing-impaired or dyslexic people who show differential oscillatory mechanisms (for hearing impairment, for example: Decruy et al., 2020; Gillis et al., 2023; Petersen et al., 2017; for dyslexia: Goswami, 2011; Lehongre et al., 2013; Lizarazu et al., 2021b). First studies in this direction revealed that tACS at appropriate frequencies can enhance speech-in-noise perception in hearing-impaired, elderly participants (Erkens et al., 2021) as well as phonological processing in dyslexic adults (Marchesotti et al., 2020).

To conclude, the empirical evidence available so far seems to provide arguments that speech comprehension depends on neural tracking, at least to some extent. Brain-to-speech coupling may indeed be necessary, although not sufficient, for intelligibility. On the other hand, as we have seen, neural tracking does not need intelligible speech to occur. Nevertheless, linguistic information can affect the strength of the coupling, suggesting a bidirectional functional relationship between the two.

5.3.3 Motor Regions as Temporal Predictors for Speech Decoding

Low-frequency coupling to speech in auditory regions is also top-down modulated by alpha and beta activity in motor regions. Crucially, the stronger this cortico-cortical coupling, the stronger the auditory neural tracking of the speech input (Assaneo and Poeppel, 2018, for syllable trains at 4.5 Hz; Park et al., 2015, for intact versus reversed speech). Alpha and beta power modulations in fronto-central regions were moreover shown to precede coupling to speech in the auditory cortex, supporting their anticipatory role in lower-level speech processing (Keitel et al., 2017; see also Di Liberto et al., 2018; Hovsepyan et al., 2023). In line with their involvement in predictive coding (Arnal, 2012; Morillon et al., 2014; Schroeder et al., 2010) and temporal processing (Coull, 2015; Schubotz, 2007), motor regions are therefore able to extract the (quasi-)rhythmicity of speech and generate temporal predictions so as to optimize speech sampling.

https://doi.org/10.1017/9781009295888.007 Published online by Cambridge University Press

Motor regions do not only modulate auditory cortex activity; they also synchronize their own oscillations to speech fluctuations at the phrasal and syllabic scales. Keitel et al. (2018) reported stronger delta band (0.6–1.3 Hz) left premotor tracking of speech for correctly identified sentences in English. Delta/beta (13–30 Hz) phase/amplitude coupling within this premotor cluster predicted comprehension as well, emphasizing the role of motor regions in predicting the likely occurrence of upcoming words. At the syllabic level, the right motor cortex was shown to track the isochronous rhythm of Chinese monosyllabic words at 4 Hz (Sheng et al., 2019) as well as the fast syllable rate (~9 syll/s) in French sentences (Hincapié Casas et al., 2021). For the latter, the motor cortex furthermore increased its connectivity with a left fronto-parieto-temporal network, probably belonging to the speech sensorimotor dorsal stream (Hickok and Poeppel, 2007). Such sensorimotor interactions may reflect the encoding of articulatory information as well as the use of internal models for phonological-articulatory mapping in challenging listening situations (see Chapter 6 for sensorimotor interactions during speech processing). Interestingly, listeners' sensorimotor abilities predict the ability of motor oscillations to align to speech syllabic structure (Assaneo et al., 2019). Coupling of bilateral frontal regions to trains of random syllables at 4.5 Hz was indeed stronger in English participants who better synchronized their own syllabic production to the perceived rate. These "high synchronizers" additionally showed microstructural white matter differences, namely stronger left lateralization in a territory of the arcuate fasciculus compared to "low synchronizers." Such anatomical changes could increase information transfer between auditory and motor regions underpinning speech perception and production, and consequently improve neural tracking of speech.

A growing body of evidence therefore converges in favor of a contribution of motor regions to the parsing of speech into syllable- and word/phrase-size packets, either through top-down predictions or direct synchronization to the signal (quasi-)rhythmic fluctuations. More work is nevertheless needed to understand the role of the different motor rhythms, in the delta, theta, and beta bands in particular, in deciphering the temporal structure of speech.

5.4 Concluding Remarks and Future Directions

The reviewed E/MEG findings generally corroborate oscillatory models of speech perception by revealing that low-frequency (theta band) oscillations in the auditory cortex synchronize to syllabic rhythmic information and modulate gamma amplitude, although only a few studies eventually examined this cross-frequency coupling. Theta activity can also adapt to speech rate variations (mostly investigated through artificial manipulation), at least to a certain extent. As we underlined, future work should, however, consider more naturalistic settings, using connected speech naturally produced at different rates and analyzing coupling to global but also local rate changes, to better grasp the underlying oscillatory dynamic and thus advance current neurolinguistic models. It is also striking that, despite some studies assessing neural tracking in bilinguals (Di Liberto et al., 2021; Lizarazu et al., 2021a; Lu et al., 2023; Peña and Melloni, 2012; Pérez et al., 2015; Reetzke et al., 2021), the effects of cross-linguistic differences, regardless of comprehension, have so far been largely ignored in the field. The question as to

https://doi.org/10.1017/9781009295888.007 Published online by Cambridge University Press

whether neural tracking is sensitive to the languages' rhythmic properties per se, and not only to the listeners' native language, indeed remains open. Yet investigating this issue could be highly informative in delineating the correspondence between neural oscillations and the timescales in speech, for instance by comparing brain-to-speech coupling for languages belonging to different rhythmic classes. We are only aware of one recent study that tackled this issue. Using EEG, Özer et al. (2023) examined theta band phase-locking to the speech envelope in English and Spanish adults while they listened to resynthesized incomprehensible sentences from English (a stress-timed language), Spanish (syllable-timed), and Japanese (mora-timed). Neural tracking was weakest for English, intermediate for Spanish, and strongest for Japanese, irrespective of the participants' mother tongue. Hence, brain-to-speech coupling reflected the rhythmic regularities of the heard languages, entailing syllabic complexity (from the least in English to the highest in Japanese), rather than processing of the native language.

When it comes to the functional relationship between brain coupling and speech intelligibility, the picture gets a bit complex with (at least partly) contradictory findings. We again emphasize that no study to the best of our knowledge has reported that speech could be intelligible without auditory oscillatory coupling to its envelope. Neural tracking of the signal could therefore be a necessary though insufficient step to support speech decoding. Bottom-up acoustic coupling in the auditory cortex is indeed modulated by speech intelligibility and occurs in parallel to, or under the influence of, other top-down processes, some of them arising from the motor cortex. By temporally structuring oscillatory activity in auditory regions and by directly synchronizing their own rhythms to speech modulations, motor regions optimize speech decoding. Although we did not address this issue in the present chapter, other high-level predictions generated from lexico-semantic and syntactic knowledge obviously also enable inferences on the incoming speech signal (e.g., Ding et al., 2016; Kaufeld et al., 2020; Lo et al., 2022; Meyer et al., 2019). Examining the dynamic interplay between predictions of different natures, for neural tracking of speech at different frequencies in distributed brain regions, appears crucial to capture how speech units are hierarchically organized and combined to ultimately comprehend spoken language.

Box 5.1 Chapter Overview

Summary

We reviewed evidence that theta range neural oscillations in auditory regions track syllabic rhythm and adapt to speech rate variations. This coupling contributes to speech intelligibility, while at the same time being modulated by linguistic information. By extracting temporal regularities in the signal, motor regions also contribute to this temporal dynamic.

Implications

We draw attention to topics for future research to improve oscillatory models of speech perception. This includes investigating neural tracking of naturally produced connected

https://doi.org/10.1017/9781009295888.007 Published online by Cambridge University Press

Box 5.1 (cont.)

speech at different rates, also in light of cross-linguistic differences, and examining the interplay between motor and auditory rhythms as well as their interactions with higher-level linguistic predictive processes.

Gains

This non-exhaustive review provides a better understanding of the functional relationship between neural oscillations and speech (quasi-)rhythmic information. Considering important issues such as dynamic coupling to speech rate changes, the influence of linguistic information, and the role of motor regions may prove valuable to refine oscillatory models of speech perception.

REFERENCES

Adank, P., and Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *NeuroImage*, 49(1), 1124–1132. https://doi.org/10.1016/j.neuroimage.2009.07.032

Adank, P., and Janse, E. (2009). Perceptual learning of time-compressed and natural fast speech. *The Journal of the Acoustical Society of America*, 126(5), 2649. https://doi.org/10.1121/1.3216914

Ahissar, E., Nagarajan, S., Ahissar, M., et al. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences*, 98(23), 13367–13372. https://doi.org/10.1073/pnas.201400998

Alexandrou, A. M., Saarinen, T., Kujala, J., and Salmelin, R. (2018). Cortical tracking of global and local variations of speech rhythm during connected natural speech perception. *Journal of Cognitive Neuroscience*, 30(11), 1704–1719. https://doi.org/10.1162/jocn_a_01295

Alexandrou, A. M., Saarinen, T., Kujala, J., and Salmelin, R. (2020). Cortical entrainment: What we can learn from studying naturalistic speech perception. *Language, Cognition and Neuroscience*, 35(6), 681–693. https://doi.org/10.1080/23273798.2018.1518534

Arnal, L. H. (2012). Predicting "when" using the motor system's beta-band oscillations. *Frontiers in Human Neuroscience*, 6, 225. https://doi.org/10.3389/fnhum.2012.00225

Assaneo, M. F., and Poeppel, D. (2018). The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. *Science Advances*, 4(2), eaao3842. https://doi.org/10.1126/sciadv.aao3842

Assaneo, M. F., Ripolles, P., Orpella, J., et al. D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nature Neuroscience*, 22(4), 627–632. https://doi.org/10.1038/s41593-019-0353-z

Baltzell, L. S., Srinivasan, R., and Richards, V. M. (2017). The effect of prior knowledge and intelligibility on the cortical entrainment response to speech. *Journal of Neurophysiology*, 118(6), 3144–3151. https://doi.org/10.1152/jn.00023.2017

https://doi.org/10.1017/9781009295888.007 Published online by Cambridge University Press

Berry, J. J. (2011). Speaking rate effects on normal aspects of articulation: Outcomes and issues. *Perspectives on Speech Science and Orofacial Disorders*, 21(1), 15–26. https://doi.org/10.1044/ssod21.1.15

Bosker, H. R. (2017). How our own speech rate influences our perception of others. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(8), 1225–1238. https://doi.org/10.1037/xlm0000381

Bosker, H. R., and Ghitza, O. (2018). Entrained theta oscillations guide perception of subsequent speech: Behavioural evidence from rate normalisation. *Language, Cognition and Neuroscience*, 33(8), 955–967. https://doi.org/10.1080/23273798.2018.1439179

Buzsáki, G., and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304(5679), 1926–1929. https://doi.org/10.1126/science.1099745

Cason, N., and Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia*, 50(11), 2652–2658. https://doi.org/10.1016/j.neuropsychologia.2012.07.018

Christophe, A., Gout, A., Peperkamp, S., and Morgan, J. (2003). Discovering words in the continuous speech stream: The role of prosody. *Journal of Phonetics*, 31(3), 585–598. https://doi.org/10.1016/S0095-4470(03)00040-8

Coull, J. T. (2015). A frontostriatal circuit for timing the duration of events. In A. W. Toga (ed.), *Brain Mapping* (pp. 565–570). Academic Press. https://doi.org/10.1016/B978-0-12-397025-1.00285-2

Davidesco, I., Thesen, T., Honey, C. J., et al. (2018). Electrocorticographic responses to time-compressed speech vary across the cortical auditory hierarchy. BioRxiv. https://doi.org/10.1101/354464

Decruy, L., Vanthornhout, J., and Francart, T. (2020). Hearing impairment is associated with enhanced neural tracking of the speech envelope. *Hearing Research*, 393, 107961. https://doi.org/10.1016/j.heares.2020.107961

Di Liberto, G. M., Lalor, E. C., and Millman, R. E. (2018). Causal cortical dynamics of a predictive enhancement of speech intelligibility. *NeuroImage*, 166, 247–258. https://doi.org/10.1016/j.neuroimage.2017.10.066

Di Liberto, G. M., Nie, J., Yeaton, J., et al. (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., and Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, 21(11), 1664–1670. https://doi.org/10.1177/0956797610384743

Dimitrijevic, A., Smith, M. L., Kadis, D. S., and Moore, D. R. (2019). Neural indices of listening effort in noisy environments. *Scientific Reports*, 9(1), 1. https://doi.org/10.1038/s41598-019-47643-1

Ding, N., Chatterjee, M., and Simon, J. Z. (2014). Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *NeuroImage*, 88, 41–46. https://doi.org/10.1016/j.neuroimage.2013.10.054

Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, 19(1), 158–164. https://doi.org/10.1038/nn.4186

Doelling, K. B., Arnal, L. H., Ghitza, O., and Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, 85, 761–768. https://doi.org/10.1016/j.neuroimage.2013.06.035

Dupoux, E., and Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance*, 23(3), 914–927. https://doi.org/10.1037/0096-1523.23.3.914

Erkens, J., Schulte, M., Vormann, M., Wilsch, A., and Herrmann, C. S. (2021). Hearing impaired participants improve more under envelope-transcranial alternating current stimulation when signal to noise ratio is high. *Neuroscience Insights*, 16, 2633105520988854. https://doi.org/10.1177/2633105520988854

https://doi.org/10.1017/9781009295888.007 Published online by Cambridge University Press

Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, 9(10), 474–480. https://doi.org/10.1016/j.tics.2005.08.011

Fries, P. (2015). Rhythms for cognition: Communication through coherence. *Neuron*, 88(1), 220–235. https://doi.org/10.1016/j.neuron.2015.09.034

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. (2014). Behavioral evidence for the role of cortical Î¸ oscillations in determining auditory channel capacity for speech. *Frontiers in Psychology*, 5, 652. https://doi.org/10.3389/fpsyg.2014.00652

Ghitza, O., and Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, 66(1–2), 113–126. https://doi.org/10.1159/000208934

Gillis, M., Kries, J., Vandermosten, M., and Francart, T. (2023). Neural tracking of linguistic and acoustic speech representations decreases with advancing age. *NeuroImage*, 267, 119841. https://doi.org/10.1016/j.neuroimage.2022.119841

Giraud, A.-L., and Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15(4), 511–517. https://doi.org/10.1038/nn.3063

Golomb, J. D., Peelle, J. E., and Wingfield, A. (2007). Effects of stimulus variability and adult aging on adaptation to time-compressed speech. *The Journal of the Acoustical Society of America*, 121(3), 1701. https://doi.org/10.1121/1.2436635

Gordon-Salant, S., Zion, D. J., and Espy-Wilson, C. (2014). Recognition of time-compressed speech does not predict recognition of natural fast-rate speech by older listeners. *The Journal of the Acoustical Society of America*, 136(4), EL268–EL274. https://doi.org/10.1121/1.4895014

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, 15(1), 3–10. https://doi.org/10.1016/j.tics.2010.10.001

Gross, J., Hoogenboom, N., Thut, G., et al. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, 11(12), e1001752. https://doi.org/10.1371/journal.pbio.1001752

Guilleminot, P., and Reichenbach, T. (2022). Enhancement of speech-in-noise comprehension through vibrotactile stimulation at the syllabic rate. *Proceedings of the National Academy of Sciences*, 119(13), e2117000119. https://doi.org/10.1073/pnas.2117000119

Guiraud, H., Bedoin, N., Krifi-Papoz, S., et al. (2018). Don't speak too fast! Processing of fast rate speech in children with specific language impairment. *PLoS ONE*, 13(1), e0191808. https://doi.org/10.1371/journal.pone.0191808

Hauswald, A., Keitel, A., Chen, Y., Rösch, S., and Weisz, N. (2022). Degradation levels of continuous speech affect neural speech tracking and alpha power differently. *The European Journal of Neuroscience*, 55(11–12), 3288–3302. https://doi.org/10.1111/ejn.14912

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402. https://doi.org/10.1038/nrn2113

Hincapié Casas, A. S., Lajnef, T., Pascarella, A., et al. (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

Hovsepyan, S., Olasagasti, I., Giraud, A.-L., and Affiliation. (2023). Rhythmic modulation of prediction errors: A top-down gating role for the beta-range in speech processing. *PLOS Computational Biology*, 19, e1011595. https://doi.org/10.1371/journal.pcbi.1011595

Howard, M. F., and Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *Journal of Neurophysiology*, 104(5), 2500–2511. https://doi.org/10.1152/jn.00251.2010

https://doi.org/10.1017/9781009295888.007 Published online by Cambridge University Press

Janse, E. (2004). Word perception in fast speech: Artificially time-compressed vs. naturally produced fast speech. *Speech Communication*, 42(2), 155–173. https://doi.org/10.1016/j.specom.2003.07.001

Janse, E., Nooteboom, S., and Quené, H. (2003). Word-level intelligibility of time-compressed speech: Prosodic and segmental factors. *Speech Communication*, 41(2–3), 287–301. https://doi.org/10.1016/S0167-6393(02)00130-9

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., and Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, 32(3), 402–420. https://doi.org/10.1006/jmla.1993.1022

Kadir, S., Kaza, C., Weissbart, H., and Reichenbach, T. (2020). Modulation of speech-in-noise comprehension through transcranial current stimulation with the phase-shifted speech envelope. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(1), 23–31. https://doi.org/10.1109/TNSRE.2019.2939671

Kaufeld, G., Bosker, H. R., ten Oever, S., et al. (2020). Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. *Journal of Neuroscience*, 40(49), 9467–9475. https://doi.org/10.1523/JNEUROSCI.0302-20.2020

Keitel, A., Gross, J., and Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLOS Biology*, 16(3), e2004473. https://doi.org/10.1371/journal.pbio.2004473

Keitel, A., Ince, R. A. A., Gross, J., and Kayser, C. (2017). Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. *NeuroImage*, 147, 32–42. https://doi.org/10.1016/j.neuroimage.2016.11.062

Keshavarzi, M., Kegler, M., Kadir, S., and 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

Koch, X., and Janse, E. (2016). Speech rate effects on the processing of conversational speech across the adult life span. *The Journal of the Acoustical Society of America*, 139(4), 1618–1636. https://doi.org/10.1121/1.4944032

Kösem, A., and van Wassenhove, V. (2016). Distinct contributions of low- and high-frequency neural oscillations to speech comprehension. *Language, Cognition and Neuroscience*, 32(5), 536–544. https://doi.org/10.1080/23273798.2016.1238495

Kösem, A., Dai, B., McQueen, J. M., and 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., Bosker, H. R., Takashima, A., et al. (2018). Neural entrainment determines the words we hear. *Current Biology*, 28(18), 2867–2875. https://doi.org/10.1016/j.cub.2018.07.023

Lamekina, Y., and Meyer, L. (2022). Entrainment to speech prosody influences subsequent sentence comprehension. *Language, Cognition and Neuroscience*, 38(3), 263–276. https://doi.org/10.1080/23273798.2022.2107689

Lehongre, K., Morillon, B., Giraud, A.-L., and Ramus, F. (2013). Impaired auditory sampling in dyslexia: Further evidence from combined fMRI and EEG. *Frontiers in Human Neuroscience*, 7, 454. https://doi.org/10.3389/fnhum.2013.00454

Lizarazu, M., Lallier, M., and Molinaro, N. (2019). Phase–amplitude coupling between theta and gamma oscillations adapts to speech rate. *Annals of the New York Academy of Sciences*, 1453(1), 140–152. https://doi.org/10.1111/nyas.14099

Lizarazu, M., Carreiras, M., Bourguignon, M., Zarraga, A., and Molinaro, N. (2021a). Language proficiency entails tuning cortical activity to second language speech. *Cerebral Cortex*, 31(8), 3820–3831. https://doi.org/10.1093/cercor/bhab051

https://doi.org/10.1017/9781009295888.007 Published online by Cambridge University Press

Lizarazu, M., Scotto di Covella, L., van Wassenhove, V., et al. (2021b). Neural entrainment to speech and nonspeech in dyslexia: Conceptual replication and extension of previous investigations. *Cortex*, 137, 160–178. https://doi.org/10.1016/j.cortex.2020.12.024

Lo, C.-W., Tung, T.-Y., Ke, A. H., and Brennan, J. R. (2022). Hierarchy, not lexical regularity, modulates low-frequency neural synchrony during language comprehension. *Neurobiology of Language*, 3(4), 538–555. https://doi.org/10.1162/nol_a_00077

Lu, L., Deng, Y., Xiao, Z., Jiang, R., and Gao, J.-H. (2023). Neural signatures of hierarchical linguistic structures in second language listening comprehension. *eNeuro*, 10(6), ENEURO.0346-22.2023. https://doi.org/10.1523/ENEURO.0346-22.2023

Luo, H., and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, 54(6), 1001–1010. https://doi.org/10.1016/j.neuron.2007.06.004

Marchesotti, S., Nicolle, J., Merlet, I., et al. (2020). Selective enhancement of low-gamma activity by tACS improves phonemic processing and reading accuracy in dyslexia. *PLoS Biology*, 18(9), e3000833. https://doi.org/10.1371/journal.pbio.3000833

Maslowski, M., Meyer, A. S., and Bosker, H. R. (2019). How the tracking of habitual rate influences speech perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(1), 128–138. https://doi.org/10.1037/xlm0000579

Meng, Q., Wang, X., Cai, Y., et al. (2019). Time-compression thresholds for Mandarin sentences in normal-hearing and cochlear implant listeners. *Hearing Research*, 374, 58–68. https://doi.org/10.1016/j.heares.2019.01.011

Meyer, L., Sun, Y., and Martin, A. E. (2019). Synchronous, but not entrained: Exogenous and endogenous cortical rhythms of speech and language processing. *Language, Cognition and Neuroscience*, 35(9), 1089–1099. https://doi.org/10.1080/23273798.2019.1693050

Millman, R., Johnson, S., and 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, 1–13. https://doi.org/10.1162/jocn_a_00719

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

Müller, J. A., Wendt, D., Kollmeier, B., Debener, S., and Brand, T. (2019). Effect of speech rate on neural tracking of speech. *Frontiers in Psychology*, 10, 449. https://doi.org/10.3389/fpsyg.2019.00449

Nourski, K. V., Reale, R. A., Oya, H., et al. (2009). Temporal envelope of time-compressed speech represented in the human auditory cortex. *Journal of Neuroscience*, 29(49), 15564–15574. https://doi.org/10.1523/JNEUROSCI.3065-09.2009

Novembre, G., and Iannetti, G. D. (2018). Tagging the musical beat: Neural entrainment or event-related potentials? *Proceedings of the National Academy of Sciences of the United States of America*, 115(47), E11002–E11003. https://doi.org/10.1073/pnas.1815311115

Oganian, Y., and Chang, E. F. (2019). A speech envelope landmark for syllable encoding in human superior temporal gyrus. *Science Advances*, 5(11), eaay6279. https://doi.org/10.1126/sciadv.aay6279

Oganian, Y., Kojima, K., Breska, A., et al. (2023). Phase alignment of low-frequency neural activity to the amplitude envelope of speech reflects evoked responses to acoustic edges, not oscillatory entrainment. *Journal of Neuroscience*, 43(21), 3909–3921. https://doi.org/10.1523/JNEUROSCI.1663-22.2023

Özer, E. E., Pereira, S. S., and Sebastian-Galles, N. (2023). Neural entrainment to speech in theta range is affected by language properties but not by the native language of the listeners. bioRxiv. www.biorxiv.org/content/10.1101/2023.07.11.548540v1.abstract

Pallier, C., Sebastian-Gallés, N., Dupoux, E., Christophe, A., and Mehler, J. (1998). Perceptual adjustment to time-compressed speech: A cross-linguistic study. *Memory & Cognition*, 26(4), 844–851. https://doi.org/10.3758/BF03211403

https://doi.org/10.1017/9781009295888.007 Published online by Cambridge University Press

Park, H., and Jang, H. (2012). Recognition of time-compressed words in adults and children. *Audiology*, 8(1), 34–40.

Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., and Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology*, 25(12), 1649–1653. https://doi.org/10.1016/j.cub.2015.04.049

Peelle, J. E., and Davis, M. H. (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., and Wingfield, A. (2005). Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *Journal of Experimental Psychology: Human Perception and Performance*, 31(6), 1315–1330. https://doi.org/10.1037/0096-1523.31.6.1315

Peelle, J. E., Gross, J., and Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, 23(6), 1378–1387. https://doi.org/10.1093/cercor/bhs118

Pefkou, M., Arnal, L. H., Fontolan, L., and Giraud, A.-L. (2017). Theta- and beta-band neural activity reflect independent syllable tracking and comprehension of time-compressed speech. *Journal of Neuroscience*, 37(33), 7930–7938. https://doi.org/10.1523/JNEUROSCI.2882-16.2017

Peña, M., and Melloni, L. (2012). Brain oscillations during spoken sentence processing. *Journal of Cognitive Neuroscience*, 24(5), 1149–1164. https://doi.org/10.1162/jocn_a_00144

Penn, L. R., Ayasse, N. D., Wingfield, A., and Ghitza, O. (2018). The possible role of brain rhythms in perceiving fast speech: Evidence from adult aging. *The Journal of the Acoustical Society of America*, 144(4), 2088–2094. https://doi.org/10.1121/1.5054905

Pérez, A., Carreiras, M., Gillon Dowens, M., and Duñabeitia, J. A. (2015). Differential oscillatory encoding of foreign speech. *Brain and Language*, 147, 51–57. https://doi.org/10.1016/j.bandl.2015.05.008

Petersen, E. B., Wöstmann, M., Obleser, J., and Lunner, T. (2017). Neural tracking of attended versus ignored speech is differentially affected by hearing loss. *Journal of Neurophysiology*, 117(1), 18–27. https://doi.org/10.1152/jn.00527.2016

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as "asymmetric sampling in time." *Speech Communication*, 41(1), 245–255. https://doi.org/10.1016/S0167-6393(02)00107-3

Poeppel, D., and Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, 21(6), 6. https://doi.org/10.1038/s41583-020-0304-4

Quené, H., and Port, R. F. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica*, 62(1), 1.

Reetzke, R., Gnanateja, G. N., and 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

Reinisch, E. (2016). Natural fast speech is perceived as faster than linearly time-compressed speech. *Attention, Perception, & Psychophysics*, 78(4), 1203–1217. https://doi.org/10.3758/s13414-016-1067-x

Reinisch, E., and Sjerps, M. J. (2013). The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context. *Journal of Phonetics*, 41(2), 101–116. https://doi.org/10.1016/j.wocn.2013.01.002

Riecke, L., Formisano, E., Sorger, B., Başkent, D., and Gaudrain, E. (2018). Neural entrainment to speech modulates speech intelligibility. *Current Biology*, 28(2), 161–169. https://doi.org/10.1016/j.cub.2017.11.033

Riecke, L., Snipes, S., van Bree, S., Kaas, A., and Hausfeld, L. (2019). Audio-tactile enhancement of cortical speech-envelope tracking. *NeuroImage*, 202, 116134. https://doi.org/10.1016/j.neuroimage.2019.116134

https://doi.org/10.1017/9781009295888.007 Published online by Cambridge University Press

Rimmele, J. M., Golumbic, E. Z., Schröger, E., and 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

Roncaglia-Denissen, M. P., Schmidt-Kassow, M., and Kotz, S. A. (2013). Speech rhythm facilitates syntactic ambiguity resolution: ERP evidence. *PLoS ONE*, 8(2), e56000. https://doi.org/10.1371/journal.pone.0056000

Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 336(1278), 367–373. https://doi.org/10.1098/rstb.1992.0070

Schlueter, A., Lemke, U., Kollmeier, B., and Holube, I. (2014). Intelligibility of time-compressed speech: The effect of uniform versus non-uniform time-compression algorithms. *The Journal of the Acoustical Society of America*, 135(3), 1541–1555. https://doi.org/10.1121/1.4863654

Schmidt, F., Chen, Y.-P., Keitel, A., et al. (2023). Neural speech tracking shifts from the syllabic to the modulation rate of speech as intelligibility decreases. *Psychophysiology*, 60(11), e14362. https://doi.org/10.1111/psyp.14362

Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., and Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Current Opinion in Neurobiology*, 20(2), 172–176. https://doi.org/10.1016/j.conb.2010.02.010

Schubotz, R. I. (2007). Prediction of external events with our motor system: Towards a new framework. *Trends in Cognitive Sciences*, 11(5), 211–218. https://doi.org/10.1016/j.tics.2007.02.006

Sebastián-Gallés, N., Dupoux, E., Costa, A., and Mehler, J. (2000). Adaptation to time-compressed speech: Phonological determinants. *Perception & Psychophysics*, 62, 834–842. https://doi.org/10.3758/BF03206926

Sheng, J., Zheng, L., Lyu, B., et al. (2019). The cortical maps of hierarchical linguistic structures during speech perception. *Cerebral Cortex*, 29(8), 3232–3240. https://doi.org/10.1093/cercor/bhy191

Sidiras, C., Iliadou, V., Nimatoudis, I., Reichenbach, T., and Bamiou, D.-E. (2017). Spoken word recognition enhancement due to preceding synchronized beats compared to unsynchronized or unrhythmic beats. *Frontiers in Neuroscience*, 11, 415. https://doi.org/10.3389/fnins.2017.00415

van Bree, S., Sohoglu, E., Davis, M. H., and Zoefel, B. (2021). Sustained neural rhythms reveal endogenous oscillations supporting speech perception. *PLoS Biology*, 19(2), e3001142. https://doi.org/10.1371/journal.pbio.3001142

Varela, F., Lachaux, J.-P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience*, 2(4), 4. https://doi.org/10.1038/35067550

Verschueren, E., Gillis, M., Decruy, L., Vanthornhout, J., and Francart, T. (2022). Speech understanding oppositely affects acoustic and linguistic neural tracking in a speech rate manipulation paradigm. *Journal of Neuroscience*, 42(39), 7442–7453. https://doi.org/10.1523/JNEUROSCI.0259-22.2022

Versfeld, N., and Dreschler, W. (2002). The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners. *The Journal of the Acoustical Society of America*, 111, 401–408. https://doi.org/10.1121/1.1426376

Wilsch, A., Neuling, T., Obleser, J., and 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

Zhang, Y., Zou, J., and Ding, N. (2023). Acoustic correlates of the syllabic rhythm of speech: Modulation spectrum or local features of the temporal envelope. *Neuroscience & Biobehavioral Reviews*, 147, 105111. https://doi.org/10.1016/j.neubiorev.2023.105111

Zion Golumbic, E. M., Ding, N., Bickel, S., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail Party." *Neuron*, 77(5), 980–991. https://doi.org/10.1016/j.neuron.2012.12.037

https://doi.org/10.1017/9781009295888.007 Published online by Cambridge University Press

Zoefel, B., and 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

Zoefel, B., Archer-Boyd, A., and Davis, M. H. (2018). Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. *Current Biology*, 28(3), 401–408. https://doi.org/10.1016/j.cub.2017.11.071

Zoefel, B., Allard, I., Anil, M., and Davis, M. H. (2020). Perception of rhythmic speech is modulated by focal bilateral transcranial alternating current stimulation. *Journal of Cognitive Neuroscience*, 32(2), 226–240. https://doi.org/10.1162/jocn_a_01490

Zou, J., Xu, C., Luo, C., et al. (2021). θ-band cortical tracking of the speech envelope shows the linear phase property. *Eneuro*, 8(4), ENEURO.0058-21.2021. https://doi.org/10.1523/ENEURO.0058-21.2021

https://doi.org/10.1017/9781009295888.007 Published online by Cambridge University Press