UNIVERSITÄT ZU LÜBECK

Aus dem Institut für Psychologie I

der Universität zu Lübeck

Direktor: Prof. Dr. Nico Bunzeck

Speech comprehension through the lens of auditory and motor brain rhythms

Inauguraldissertation

zur

Erlangung der Doktorwürde

der Universität zu Lübeck

Aus der Sektion Naturwissenschaften

vorgelegt von

Christina Anna-Lena Lubinus

aus Aurich

Lübeck, 2024

1. Berichterstatter/Berichterstatterin: Prof. Dr. Jonas Obleser

2. Berichterstatter/Berichterstatterin: Prof. Dr. Nathan Weisz

Tag der mündlichen Prüfung: 11.10.2024

Zum Druck genehmigt. Lübeck, den 08.05.2025

Speech comprehension through the lens of auditory and motor brain rhythms

Acknowledgements

I am deeply thankful to my supervisor, Johanna Rimmele, who has been an incredible support throughout this PhD journey, always had an open door, and provided structure to my thoughts when needed. I am grateful for this relationship we established, balancing mentorship and friendship, and for the many things I was able to learn from you.

I wish to express my gratitude to David Poeppel for giving me the opportunity to pursue my PhD at the Max Planck Institute for Empirical Aesthetics and for being such an inspiration, to Jonas Obleser for formally supervising me and offering valuable advice, to Anne Keitel for being an incredible co-author and mentor who feels like a friend, and to my examiners for agreeing to read and evaluate this thesis.

I am thankful to my colleagues at the MPI who shared the daily business of this journey, countless coffees and many lunches, especially Camila Bruder, Anna Czepiel, Ece Kaya, Yuranny Cabral-Calderin, Simone Franz, Dominik Welke, and Frank Sierra; to Arne Nagels, Jona Sassenhagen, Phillip Alday, Miriam Steines, and Bianca van Kemenade, my early academic supporters at the Philipps-University of Marburg, who encouraged undergrad-me to pursue a PhD; to my beloved friends Alke Rickels, Johanna Vey, Lisa-Marie Smit, Corinna Hess, and Franzi Rüsch for being the incredible, smart, kind-hearted and supportive women they are.

This achievement would not have been possible without my family, who always believed in me and unconditionally supported me in following my path. Thank you to my parents, Hanne and Oltmann Lubinus, my brother Oliver, my stepmom Heidi Wende, and my grandfather Johann van Lengen, who was so looking forward to calling me Dr. Christina.

And most importantly, I extend my deepest gratitude to Lars Krautkrämer, my rock, best friend, and constant source of inspiration. I am grateful for all the little and big things you did on my behalf while I was knee-deep in writing my thesis.

Christina Anna-Lena Lubinus, Frankfurt a.M., Mai 2024

Contents

1 General introduction

Communication is a fundamental aspect of human behavior, with speech being a pivotal acoustic signal within our perceptual environment. Speech is dynamic and unfolds in time, fluctuating at similar timescales as the ongoing activity in our brains. These endogenous brain rhythms are posited to partake in temporally segmenting speech into discrete units, such as syllables, thereby contributing to the hierarchical process underlying speech comprehension (Giraud and Poeppel, 2012). However, while there is convincing evidence that the brain synchronizes to acoustic speech signals, empirical substantiation regarding an engagement of endogenous brain rhythms in syllabic segmentation remains limited. The engagement of endogenous brain rhythms that fluctuate in the absence of external stimulation would endorse the proposal of a proactive oscillatory process (Zoefel et al., 2018b; Obleser and Kayser, 2019). Communication is an intricate interplay of acoustic processing and motor production, exemplifying an instance of ecologically relevant auditory-motor interaction in human behavior. In the brain, speech comprehension and production are conventionally associated with auditory and speech-motor brain regions, respectively (Hickok and Poeppel, 2007). However, not only during production but also during comprehension, these regions may interact tightly, possibly enhancing perceptual processes.

This thesis aims to deepen our understanding of the neural underpinnings of speech processing by examining the putative involvement of oscillatory mechanisms. Specifically, it investigates the contribution of endogenous brain rhythms to speech processing, with a particular emphasis on the temporal dynamics of the speech-motor system and its interactions with auditory processing. Additionally, other higher-level influences, such as linguistic predictability and working memory, on the proposed oscillatory model of speech perception are investigated. Comprising three studies, this thesis capitalizes on the neuroplasticity of brain rhythms, individual differences in behavior, and a multimodal approach integrating behavioral and electrophysiological methods to assess effects of individual differences in endogenous brain rhythms and auditory-motor interactions on speech processing.

1.1 Endogenous brain rhythms in speech perception

1.1.1 Origins of endogenous brain rhythms

Rhythmic neural activity is a prominent feature of the brain, observed at multiple timescales throughout the brain (Singer, 2018). These brain rhythms can be observed both during task-related and resting states, with resting activity likely reflecting intrinsic or “endogenous” rhythms. First identified in humans by Hans Berger (1929), endogenous brain rhythms have since been measured non-invasively using electro- (EEG) and magnetoencephalography (MEG) or intracranially in patients

using electrocorticography (ECoG) or depth electrodes. Brain rhythms are commonly grouped into the canonical frequency bands which can be observed from such recordings (Lopes da Silva, 2013; Cole and Voytek, 2017): delta (0.2-3.5 Hz), theta (4-7.5 Hz), alpha (8-13 Hz), beta (14-30 Hz), gamma (30-90 Hz), and high-gamma (>90 Hz). Interestingly, rhythmic activity at similar timescales is conserved across species (Buzsáki et al., 2013), suggesting that these timescales reflect behaviorally relevant biophysical properties, likely underlying sensory, motor and cognitive functions.

Rhythmic activity can originate from various sources across the neural hierarchy. At the single-cell (microscopic) level, it emerges from fluctuations of currents that destabilize the membrane resting potential of neurons (Hutcheon and Yarom, 2000). At the macroscopic level, brain rhythms are observed as population phenomenon, originating from transmembrane currents of neighboring neural assemblies (Schroeder et al., 1995; Schroeder and Lakatos, 2009; for review: Singer, 2018). In this scenario, synchronized alternations of inward and outward current flow accumulate (Lakatos et al., 2005; Rimmele and Keitel, 2023), generating electromagnetic fields large enough to be detected outside the brain by noninvasive methods such as MEG and EEG (Singer, 2013; Gross, 2019). As first proposed by Bishop (1932), fluctuations in neural activity are believed to reflect changes in neural excitability, where inward and outward currents correspond to depolarization and hyperpolarization, i.e. states in which cells are more or less likely to fire, respectively (Lakatos et al., 2005).

Functionally, brain rhythms may underlie distinct forms of neural communication and information transfer. Firstly, the resonance properties of brain rhythms naturally associate them with a filtering function, allowing inputs that match their frequency characteristics to pass through while attenuating others (Buzsáki and Draguhn, 2004). Secondly, synchronization of rhythmic activity is believed to facilitate binding of information, such as individual features of visual objects. Early work suggested that such a synchronization of neural discharges reflects neurons encoding related information (Singer and Gray, 1995). A similar account hypothesized that synchronization in the phase space (i.e. phase-locking) enables information transfer between distant brain areas (Fries, 2005). Specifically, fluctuations in excitability are proposed to reflect temporal windows of opportunity offering “natural temporal frames for grouping, or 'chunking', neural activity into cell assemblies […] for the effective exchange of information among cortical networks” (Buzsáki et al., 2013 p. 752). Thus, synchronized neural populations may coordinate the reception and transmission of information, promoting rapid communication, while out-of-phase activity does not interfere (Fries, 2005; for criticism of this account see: Schneider et al., 2021).

Rhythmic brain activity is often conceptualized as neural oscillations (Izhikevich et al., 2003), a fundamental phenomenon observed across many biological and non-biological systems that is well-defined in physical science. Neural oscillations are characterized by a few key parameters including phase, amplitude, and preferred frequency. Phase contains temporal information of an oscillation,

denoting a point in time within its oscillatory cycle, often represented as phase angle. The frequency of an oscillation corresponds to cycles per second in units of Hertz (Hz)(Jones, 2019). Preferred frequencies, observed in both single and groups of neurons, reflect either the frequency of spontaneous neural activity in the absence of input or resonance responses, i.e. the stimulation frequency eliciting the strongest reaction. Such preferences are often measured via electro/magnetic or rhythmic sensory stimulation (Hutcheon and Yarom, 2000; Rosanova et al., 2009). The ability of driving neural activity through external stimulation underscores that, in addition to facilitating internal neural communication, rhythmic brain activity is modulated by external (e.g. sensory) signals originating in the environment. Preferred frequencies have been revealed in the alpha, beta, and gamma ranges for the visual, parietal, and frontal cortices, respectively (Rosanova et al., 2009). Steady-state evoked responses (SSER) provide further insights into these preferences, with strongest SSER observed at the preferred frequency of the stimulated system (Narici and Romani, 1989; Herrmann, 2001; Norcia et al., 2015), particularly adding value in studying the auditory system (Galambos et al., 1981; Pastor et al., 2002). SSER studies have elucidated preferences for auditory input frequencies around 40 Hz (Picton et al., 2003) and at lower frequencies (~5 Hz), prevalent in speech acoustics (i.e. amplitude envelope) and essential for speech comprehension (Rodenburg et al., 1972; Picton et al., 1987; Liegeois-Chauvel et al., 2004; Will and Berg, 2007). These lower frequencies in the theta frequency range, crucial for speech perception, are of particular interest in the current thesis. Additionally, spectral profiles (sometimes referred to as spectral fingerprints) provide an alternative method to characterize endogenous brain rhythms. This approach has been used to identify predominant spectral signatures across the brain or in individual brain regions (Groppe et al., 2013; Gunasekaran et al., 2023). Intriguingly, this research uncovered a posterior-to-anterior spectral gradient across cortex and revealed that brain regions can display multiple spectral clusters, which fluctuate in dominating over time (Keitel and Gross, 2016; Komorowski et al., 2018).

Recent research revealed that rhythmic brain activity is not only crucial for proper function of and communication between (distant) neural populations, but also interacts with environmental stimuli optimizing sensory selection (Schroeder and Lakatos, 2009). These brain rhythms can be recruited in a task-specific manner during sensory, motor and higher-cognitive tasks, linking intrinsic brain rhythms to cognitive functions (Engel et al., 1992, 2001; Singer and Gray, 1995; Lakatos et al., 2008, 2013; Holcombe, 2009; Giraud and Poeppel, 2012; Heusser et al., 2016; Portoles et al., 2018; Rimmele et al., 2018; Schroeder et al., 2018; VanRullen, 2018). These findings led to the notion that “neural oscillations provide the temporal dimensions of how the mind works” (Doelling et al., 2023 p. 6). Given the intrinsic temporal properties of endogenous brain rhythms, they may play a pivotal functional role in processing (quasi)rhythmic sensory signals.

1.1.2 Brain rhythms in rhythmic auditory perception

Rhythmic sensory events facilitate behavior, as demonstrated by early behavioral work showing improved processing of stimuli embedded in rhythmic contexts (i.e. sequence of events), even when the context is task-irrelevant (Jones et al., 1981, 1982, 2002; Kidd et al., 1984). The Dynamic Attending Theory (DAT) postulates that our environment is highly temporally structured, with these "world patterns" being processed by "perceptual rhythms", influencing perception, memory, and attention (Jones, 1976, 2019). DAT suggests that rhythmic patterns in auditory stimuli are the basis for generating temporal expectations (Boltz, 1993), supposedly enhancing processing of sensory events. Importantly, DAT – despite being grounded in behavioral work – already hypothesized that the perceptual rhythms may be implemented as rhythmic brain activity (Jones, 2019). As such the theory predicts that behavioral facilitation of rhythmic stimuli is due to optimal alignment between ongoing brain rhythms and a rhythmic sensory stimulus, termed neural entrainment. Specifically, entrainment describes the unidirectional coupling of an endogenous brain rhythm's phase to the phase of an external signal, e.g. a sequence of tones. As such it is similar to but different from the neural synchronization outlined in the previous section, wherein two (supposedly) oscillatory signals phase-align bidirectionally (Lakatos et al., 2019). In the past few decades, extensive research investigated the neurophysiological aspects of rhythmic processing, yielding many results that substantiate the suggested role of brain rhythms in rhythmic perception.

"The most straightforward prediction of rhythmic perception is that the outcome of sensory processing should depend on the exact phase of the critical rhythm at around the time of stimulus presentation" (VanRullen, 2016 p. 725). As outlined in section 1.1.1, the instantaneous phase of brain rhythms reflects the excitability level of the underlying neural population at any given moment in time. As such, brain rhythms are believed to naturally sample sensory inputs rhythmically causing sensory events to be processed differently (enhanced vs. attenuated), depending on the phase in which they arrive (high vs. low excitable). Importantly, this should also apply to non-rhythmic stimuli. The evidence for sensory sampling in the visual system is compelling (for review: VanRullen, 2016), while only few studies showed such effects in audition. Behaviorally, perceptual detection of non-rhythmic auditory events exhibits modulation at low (delta to theta) frequencies (Henry et al., 2016; Ho et al., 2017; Kayser, 2019; Plöchl et al., 2022). Furthermore, participants' target detection was linked to the phase of the neural delta (ten Oever et al., 2015) and theta (Ng et al., 2012) rhythms, or a more complex relation of nested slow-frequency oscillatory phases (i.e. microstates), governed by the phase of the delta rhythm (Henry et al., 2016). These results contributed to the notion that brain rhythms generate perceptual cycles (VanRullen, 2016).

Beyond spontaneous sensory sampling, numerous studies demonstrated that brain rhythms track incoming rhythmic stimuli, affecting both perceptual processing and post-stimulus performance.

Using different kinds of rhythmic auditory stimuli and paradigms these studies reported phase-locking between low-frequency (delta, theta, alpha) brain activity and the temporal structure of rhythmic stimuli (Henry and Obleser, 2012; Doelling et al., 2014; Henry et al., 2014, 2016; Doelling and Poeppel, 2015; Lakatos et al., 2016), with the optimal phases likely being subject to individual differences (Assaneo et al., 2021; Cabral-Calderin and Henry, 2022; Cabral-Calderin et al., 2024). The concept of neural entrainment is very appealing as it provides the foundation for temporal predictions as "any rhythmic input can, by definition, be predicted" (Zoefel et al., 2018b p. 1). Although consistent with the notion of oscillatory processing, the studies above cannot discern whether the underlying neural activity generating these effects is genuinely oscillatory or arises from a series of evoked potentials—a pertinent consideration given that any auditory event elicits an event-related response (Zoefel et al., 2018b). Some studies demonstrated phase-alignment with subthreshold acoustic stimuli, indicating that tracking may not solely reflect evoked responses, but may also involve an anticipatory mechanism (ten Oever et al., 2017). However, conclusive evidence for genuine entrainment remains elusive. To establish this, it would be necessary to demonstrate: (1) the presence of an endogenous rhythm existing without stimulation (2) that synchronizes to a speech signal but (3) regresses to its endogenous rate (4) within a few cycles after stimulus offset (Obleser and Kayser, 2019).

Brain rhythms likely facilitate rhythmic perception by generating temporal predictions based on the intrinsic temporal patterns of sensory events, with potential enhancement through various top-down influences. Stefanics et al. (2010) tested the influence of top-down influences by presenting auditory streams with varying predictability of event onsets, revealing stronger delta phase consistency (and faster reaction times) for predictable event onsets. This finding was recently replicated by Herbst et al. (2022). Other studies provide additional evidence for top-down effects on tracking, showing adjustments based on expectations without changes in the rhythmic signal. For instance, participants' temporal judgments have been shown to depend on the delta phase in the pre-target period, with targets presented in phase (compared to out-of-phase) with rhythmic tones associated with higher accuracy (Arnal et al., 2015). Similarly, Kösem et al. (2014) reported that perceptual changes were accompanied by shifts in neural synchronization in an audio-visual lag-adaptation paradigm, despite constant stimuli. As participants' simultaneity perception adapted to the audio-visual lag, the entrained delta phase underwent systematic shifts as well, suggesting phase adaptation. Overall, these studies suggests that the phase of brain rhythms does not only mimic the rhythm of exogenous rhythms automatically but actively anticipates it.

1.1.3 Brain rhythms in speech perception

The mechanisms believed to underlie rhythm perception are proposed to extend to speech processing (Schroeder et al., 2010; Zion Golumbic et al., 2012). However, defining rhythm in speech remains highly debated amongst researchers. While periodicity, i.e. the regular occurrence of

events, is the most intuitive notion of rhythm, strict isochrony is not required for sequences to qualify as rhythms. For instance, another notion of rhythm cognition encompasses hierarchical pattern processing (Obleser et al., 2017; Kotz et al., 2018; Doelling et al., 2022). Speech has a complex structure, characterized by nested information at multiple timescales and linguistic levels of organization such as phonemes, syllables, and phrases (Rosen, 1992; Gross et al., 2013; Ding et al., 2017). Successful speech comprehension thus requires integrating this nested and interdependent information (Poeppel, 2003; Poeppel et al., 2008). Although speech is not perfectly rhythmic, it exhibits quasi-rhythmicity at the syllabic scale (Ding, 2017), proposed to enable an oscillatory mechanism to temporally anticipate sensory events (Giraud and Poeppel, 2012; Gross et al., 2013). Nevertheless, the lack of isochrony in speech challenges the predictions from the DAT framework and neural entrainment models, as simple predictions may not suffice to precisely anticipate the onset of quasi-rhythmic events. Instead, oscillatory processing at the syllabic scale may be modulated by top-down effects to compensate for the quasi-rhythmicity, leveraging the complex temporal information in speech (Haegens and Zion Golumbic, 2018; Rimmele et al., 2018). This process is hypothesized to involve hierarchical nesting of lower-level units into one cycle of a higher-level unit (Cummins and Port, 1998; Haegens and Zion Golumbic, 2018; Rimmele et al., 2018), such as phonemes nested into syllables (Ghitza, 2011; Giraud and Poeppel, 2012).

The temporal structure of speech. The temporal structure of speech is underpinned by several acoustic features, including the amplitude envelope (Rosen, 1992), arguably the most relevant acoustic feature for oscillatory models of speech comprehension. The amplitude envelope reflects amplitude modulations over time and is associated with the perceived speech rhythm and syllabic rate of speech (Rosen, 1992; Greenberg et al., 2003; Goswami and Leong, 2013; Tilsen and Arvaniti, 2013; Ding et al., 2017; Varnet et al., 2017), although it is not a perfect correlate (Zhang et al., 2023). Beyond prosodic cues, it conveys low-level segmental information, including manner of articulation, voicing, and vowel identity, which can be inferred from acoustic features such as loudness, length, and attack. Such parallelism between linguistic units and speech acoustics has been observed at other timescales. For instance, in addition to the acoustic envelope, speech contains fast temporal modulations in the temporal fine structure, ~30-50 Hz, corresponding to phonetic features including formant transitions and voicing information (Rosen, 1992; Giraud and Poeppel, 2012).

The pseudo-rhythmic amplitude modulations of the speech envelope are evident in its spectrum, the modulation spectrum, which quantifies the rate of amplitude fluctuations over time (Ding et al., 2017). The modulation spectrum of speech is dominated by energy in the low frequencies, with a prominent peak around 4.5 Hz (Houtgast and Steeneken, 1973; Houtgast et al., 1980; Greenberg and Arai, 2001; Pellegrino et al., 2011; Ding et al., 2017; Varnet et al., 2017), aligning with the average syllabic rate (Greenberg, 1999). Despite slight variations between languages (Tilsen and Arvaniti, 2013;

Varnet et al., 2017), speakers (Tilsen and Arvaniti, 2013), and situational aspects (Sobin and Alpert, 1999), syllable duration typically averages around 200-250 ms, corresponding to a syllabic rate of 4-5 Hz (Greenberg, 1999; Strauß and Schwartz, 2017). This “nominal” frequency emerges in speech signals irrespective of the type of speech (i.e. audiobooks, interviews) or the language (Ding et al., 2017), suggesting acoustic regularities within speech, despite cross-linguistic variations in syntax, phonology etc.

Both the amplitude envelope and its modulation spectrum have straightforward neurophysiological interpretations. Analogous to the computation of the envelope, the cochlea functions as a bank of passband filters, decomposing incoming sounds spectrally and extracting the amplitude envelope (Yang et al., 1992). The ascending auditory pathway likely operates as a low-pass filter, with cells in the inferior colliculus tuned to frequency modulations ranging from 50 to 500 Hz, while the primary auditory cortex primarily responds to frequencies between 3 and 30 Hz (Giraud et al., 2000). Notably, both sub-cortical auditory nuclei and primary auditory cortex have been observed to follow speech envelopes at frequencies below 30 Hz (Shamma, 2001; Joris et al., 2004). Accordingly, the modulation spectrum is thought to approximate the spectrum of the auditory nerves and sub-cortical auditory nuclei (for review see: Ding et al., 2017).

The amplitude envelope is crucial for intelligibility. Beyond being an acoustic feature of speech, influential psychophysical studies emphasize the importance of the amplitude envelope in speech intelligibility, with envelope frequencies between 4 and 16 Hz being most critical (Drullman et al., 1994a, 1994b; Arai et al., 1996; Van Der Horst et al., 1999; Smith et al., 2002). Drullman et al. (1994b) demonstrated that low-pass filters with a cut-off frequency of 4 Hz heavily disrupt intelligibility, whereas high-pass filters with a cut-off frequency of 4 Hz have no effect. Complementarily, increasing the cut-off frequency of the low-pass filter beyond 16 Hz did not notably enhance intelligibility, while speech became incomprehensible if only frequencies above 16 Hz were present. Speech-noise chimeras yielded complementary results (Smith et al., 2002), indicating that, in contrast to manipulations to the temporal structure, intelligibility can withstand strong spectral alterations and degradation of the speech signal (Shannon et al., 1995; Greenberg et al., 1998). These results underscore the relevance of temporal modulation of the speech fine structure in speech comprehension, challenging the traditional emphasis on spectral aspects of speech (Poeppel and Assaneo, 2020).

Another line of research highlighted the behavioral relevance of an intact temporal structure of speech using time-compression, supporting an oscillatory account of speech comprehension. Time-compression modulates the temporal structure while largely preserving the spectral information (Ghitza and Greenberg, 2009). Behavioral experiments revealed that while a certain degree of time-compression is tolerable, speech becomes unintelligible at rates faster than ~9-10 Hz (Ahissar et

al., 2001; Doelling et al., 2014). Intelligibility presumably is not constrained by the syllabic rate per se (i.e. encoding the signal) but rather by the capacity to decode the speech signal. Specifically, sentences were unintelligible when compressed by a factor greater than 3 (i.e. 9 syllables per second), however, inserting periodic silence to restore the original syllabic (or theta) rhythm also restored intelligibility (Ghitza and Greenberg, 2009). This suggests that failures in intelligibility may result from a mismatch between the syllabic rhythm and the endogenous rate of the auditory brain rhythms. The authors and others proposed that the alignment between neural theta rhythm and syllabic rhythm may provide a mechanism for the brain to segment the speech signal into syllable-sized segments (Ghitza and Greenberg, 2009; Ghitza, 2012; Giraud and Poeppel, 2012; Gross et al., 2013).

Segmentation is a crucial early subroutine in the cascaded processing hierarchy of speech comprehension (Poeppel, 2003; Ghitza, 2012; Giraud and Poeppel, 2012; Meyer, 2018; Gwilliams, 2020). The primary objective of speech comprehension is to decode the intended meaning of an utterance, encapsulated in discrete "higher-order" linguistic units like morphemes, words, and phrases (Gwilliams and Davis, 2022). As such, a listener's brain faces the difficult task of segmenting the continuous acoustic waveform into discrete, lower-level units, including phonetic features and phonemes, which can be mapped onto higher-order representations carrying conceptual meaning (Teng et al., 2019; Gwilliams and Davis, 2022). According to the above proposal, theta brain rhythms may solve the problem of segmentation by identifying syllable boundaries and facilitating the subsequent phonetic readout.

Neural processing of temporal features of speech: auditory theta brain rhythms track the amplitude envelope. The importance of temporal modulations in shaping speech intelligibility invites the question: does the brain leverage these temporal modulations, exploiting the quasi-rhythmic information embedded within the speech signal? The convergence between linguistic, acoustic, and neural timescales is striking. Consistent with neural processing of rhythmic auditory stimuli, endogenous theta oscillations in auditory cortex are proposed to entrain to quasi-periodic modulations in the speech acoustics (Poeppel, 2003; Giraud and Poeppel, 2012; Peelle and Davis, 2012; Ghitza, 2013), reflecting "neural entrainment in the narrow sense" (Lakatos et al., 2008, 2019; Obleser and Kayser, 2019). This framework proposes a neural mechanism wherein phase-resets of endogenous theta oscillations serve a fundamental role in speech processing, underlying syllabic speech segmentation. The phase-reset, likely triggered by syllable edges (Doelling et al., 2014)—or syllable nuclei (Ghitza, 2013; Oganian and Chang, 2019)— is believed to align excitable phases of the theta cycle with relevant speech information. Due to a nested theta-gamma structure, phase-resets affect gamma activity indirectly. It is theorized that theta and gamma oscillations discretize syllabic and phonemic information, respectively, to group spikes into the necessary temporal packages. "In

summary, speech onsets trigger cycles of neural encoding at embedded syllabic and phonemic scales" (Giraud and Poeppel, 2012 p. 512). "Therefore, according to this hypothesis, speech entrainment does not only passively track acoustic features but also reflects the language-based packaging of speech into syllable size chunks" (Ding and Simon, 2014 p. 3).

Compelling evidence aligns with the oscillatory speech processing framework. In pioneering work, Ahissar et al. (2001) demonstrated that rhythmic cortical activity tracks the speech amplitude envelope, revealing two crucial findings: Firstly, the spectra of the speech envelope were strongly correlated with the MEG responses, peaking in the theta band (3-7 Hz). Secondly, brain-to-speech alignment was contingent on intelligibility, with intelligibility decreasing for more time-compressed speech. Emphasis has since been placed on the role of the endogenous theta rhythm in auditory cortex. Luo et al. (2007) demonstrated that the phase coherence of the auditory theta (and delta) rhythms was greater for intelligible than distorted speech and could discriminate between different sentences. The significance of the theta phase was further underscored by a study showing that removing temporal information between 2 and 9 Hz led to decreased envelope tracking and speech intelligibility, both of which were restored by substituting temporal information at the theta timescale (Doelling et al., 2014). The findings suggests that (edges in) the speech envelope may reset the phase of endogenous auditory brain rhythms to align with temporal fluctuations at the syllabic level.

Several studies support this finding of a positive association between intelligibility and speech tracking, demonstrating stronger brain-to-speech alignment for intelligible than unintelligible speech. For example, speech tracking is enhanced for forward vs. backward speech (Gross et al., 2013), clear (or slightly degraded) vs. degraded speech (Peelle et al., 2013; Chen et al., 2023), and intact speech vs. speech without temporal fluctuations (Doelling et al., 2014). Additionally, manipulating the theta phase of auditory brain rhythms using stimulation techniques has been shown to affect speech intelligibility (Riecke et al., 2018; Wilsch et al., 2018; Zoefel et al., 2018a). However, the relationship between speech tracking and intelligibility appears to be complex, with some studies reporting an inverse (Song and Iverson, 2018) or inverted U-shape (Hauswald et al., 2022) relationship. Furthermore, speech tracking has been observed for non-speech stimuli (see above, chapter 1.1.2), time-reversed speech (Howard and Poeppel, 2010), and non-native speech (Peña and Melloni, 2012). This may suggest a nuanced effect of intelligibility, as speech tracking reportedly is reduced but not diminished for unattended vs. attended speech (Zion Golumbic et al., 2013; Rimmele et al., 2015) and for perceptually irrelevant vs. relevant speech (Keitel et al., 2018). Overall, considerable evidence supports the synchronization of low-frequency brain rhythms in auditory cortex to the speech envelope, which may aid the temporal segmentation of speech. However, the relationship between intelligibility and tracking remains less clear.

1.2 Optimal temporal processing regimes for speech comprehension

Human speech displays pronounced amplitude modulations between 3 and 5 Hz, likely corresponding to syllables, suggesting temporal regularity (at least) at the syllabic rate of speech (Tilsen and Arvaniti, 2013; Ding et al., 2017; Varnet et al., 2017). However, despite the apparent rhythmic stability, there exists considerable variation in the temporal regularity of speech. While the average syllabic duration is 200-250 ms (Greenberg, 1999), these numbers may result from averaging across slower and faster syllables (Strauß and Schwartz, 2017). Furthermore, speech rate varies between languages (Tilsen and Arvaniti, 2013; Ding et al., 2017; Varnet et al., 2017), situational contexts (Sobin and Alpert, 1999), and speakers (Miller et al., 1984; Tsao and Weismer, 1997; Tilsen and Arvaniti, 2013). Thus, effective processing of such sensory input demands the auditory system's capacity to adapt to changes in speech rate.

Auditory cortex likely exhibits two distinct preferred frequency regimes: a slower one in the delta-theta range (~2-8 Hz) and a faster one in the gamma range (~30-40 Hz)(Boemio et al., 2005; Giraud et al., 2007; for review see: Edwards and Chang, 2013). Research on the temporal sensitivity of the auditory system has used (spectro-)temporal modulation transfer functions, denoting a listener's sensibility to rate modulations in sounds (Viemeister, 1979). These studies revealed a band-pass pattern of behavioral sensitivity with an optimal range for amplitude modulations between 2 and 5 Hz for speech (Elliott and Theunissen, 2009) and several types of amplitude- and frequency-modulated tones (Rodenburg, 1977; Viemeister, 1979; Chi et al., 1999). Neuroimaging studies further support these findings, demonstrating enhanced frequency tuning of auditory cortex within this 2-5 Hz behavioral sensitivity range (Giraud et al., 2000; Tanaka et al., 2000; Harms and Melcher, 2002). Additionally, neural tracking studies highlight a preference for auditory features in the gamma range, with decreased sensitivity for intermediate frequencies (Teng et al., 2017; Giroud et al., 2020; Teng and Poeppel, 2020).

While auditory cortex demonstrates temporal sensitivity to delta-theta and gamma frequencies, the system likely exhibits some degrees of flexibility with respect to the optimal range. Research using time compression revealed the systems' ability to track stimuli faster than the average modulation frequency of speech. However, this flexibility is constrained, with both speech tracking and comprehension declining beyond syllabic rates of approximately ~8-9 Hz (Ahissar et al., 2001; Brungart et al., 2007; Ghitza and Greenberg, 2009; Hincapié Casas et al., 2021). This implies that while the auditory system can adapt to varying speech rates, it is constrained by its preferred rate. This phenomenon aligns with general properties of oscillators which can synchronize to periodic signals within a specific frequency range, as described by the concept of the Arnold tongue. The Arnold tongue proposes that synchronization of an oscillator to a periodic stimulation weakens with increasing dissimilarity between the oscillators preferred and the stimulation rates, whereas its modulated

by the intensity of the stimulation (Fröhlich, 2015; Notbohm et al., 2016). Such principles may also account for the interaction of auditory and motor cortex. We hypothesize that the degree of flexibility of auditory cortex neural oscillators may be influenced by other modalities and factors, such as its interactions with the motor cortex.

1.3 Motor dynamics during speech perception

At its essence, speech perception relies on auditory processing. Therefore, when studying the neurobiological underpinnings of speech perception, the primary focus naturally is on the (extended) auditory system. However, a growing body of literature emphasizes the potential significance of the motor system in speech processing. While research has predominantly studied the motor system's involvement in speech production, its contributions to and specific functions in speech perception are less understood. Nevertheless, recent findings support an engagement of the motor system in speech perception. A prominent conceptual framework proposes "that the motor system may play a key role in perception by 'predictively' modulating sensory excitability via oscillatory coupling mechanisms" (Arnal, 2012).

For successful predictive modulation of the auditory system during speech perception, speech-motor cortex and auditory cortex must interact. In fact, the auditory and motor brain regions believed to underlie speech perception and production are structurally connected via multiple fiber bundles. As proposed by popular dual-stream models (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009), there exist two mayor connections: a ventral stream connecting temporal areas to orbitofrontal areas via anterior temporal cortex and a dorsal stream connecting superior temporal gyrus via parietal cortex to the (posterior) frontal brain (Catani et al., 2005; Saur et al., 2008). Functionally, the dorsal pathway has been linked to auditory-motor transformations (Warren et al., 2006; Hickok and Poeppel, 2007) and, according to predictive coding theories, is implicated in generating predictions of "how" and "when" something will happen (Friston, 2005).

Functionally, the motor system "engages" in speech perception. For instance, early neuroimaging research revealed coactivation of the motor cortex alongside auditory brain areas during passive speech listening (Wilson et al., 2004; Pulvermüller et al., 2006). Intracranial studies, offering very high spatial and temporal precision, further support these findings, revealing evoked neural responses to speech stimuli in frontal and motor cortex (Edwards et al., 2010; Cogan et al., 2014; Cheung et al., 2016). Cheung et al. (2016) specified this coactivation by examining the representational content in sensory motor cortex (SMC) during syllable listening. They found that ventral SMC represented acoustic categories (manner of articulation: i.e. fricative, voiced) similar to auditory brain regions (i.e. superior temporal gyrus), whereas articulatory categories (place of articulation, i.e. labial, alveolar) observed in SMC during syllable production were not represented during listening.

Moreover, speech tracking studies revealed entrainment of not only the auditory cortex but also the motor cortex to the amplitude envelope of speech (for motor tracking in beat perception, see: Fujioka et al., 2012; Keitel et al., 2018; Assaneo et al., 2019; Chalas et al., 2022). While Chalas et al. (2022) demonstrated widespread speech tracking in temporal, parietal, and frontal brain areas, Keitel et al. (2018) observed speech tracking in (left) premotor cortex, especially when perceptually relevant. In line with findings in speech, passive listening to rhythmic (non-speech) stimuli elicits activity in motor brain areas (i.e. the supplementary motor area and the premotor cortex)(Grahn and Brett, 2007; Zatorre et al., 2007; Chen et al., 2008a, 2008b).

Beyond structural connectivity and functional recruitment, brain rhythms in auditory and motor regions likely interact during speech perception, supported by evidence indicating influences from both the auditory-to-motor and motor-to-auditory systems (Scott et al., 2009; Park et al., 2015, 2018; Assaneo and Poeppel, 2018; Assaneo et al., 2019; Abbasi et al., 2023). Auditory perceptual processes can shape speech motor production, as demonstrated in psychophysical experiments manipulating real-time auditory feedback. A classic example is the delayed auditory feedback paradigm, wherein delaying auditory feedback results in stuttering-like disfluencies and speech rate reductions (Lee, 1950; Black, 1951; Bradshaw et al., 1971; Fabbro and Daro, 1995). Similarly, participants compensating for acoustically modified utterances in subsequent speech instances demonstrate real-time adjustments, seen with changes to formant or fundamental frequencies (Houde and Jordan, 1998; Jones and Munhall, 2000) and other articulatory parameters (Mitsuya et al., 2014). Alongside these behavioral findings, neurophysiological research has shown feed-forward information flow from auditory to motor brain areas during speech perception (Fontolan et al., 2014; Sedley et al., 2016; Chao et al., 2018). Reverse influences, that is from the motor to the auditory system, have also been observed. In a behavioral protocol, Assaneo, Rimmele et al. (2021) demonstrated that discrimination of syllables can be influenced by one's own periodic syllable production. Similarly, transcranial magnetic stimulation targeting the speech motor cortex interferes with the discrimination of phonemes during speech perception (D'Ausilio et al., 2009). Such a modulatory effect of the motor system on speech perception is also supported by neuroimaging studies, for instance, in studies demonstrating sensory suppression to self-elicited sensory events (Blakemore et al., 1998, 2000). This suppression phenomenon has also been observed for speech, as one's own speech signal results in lower neural activity compared to an external but acoustically identical or similar stimulus (Curio et al., 2000; Houde et al., 2002; Flinker et al., 2010; Abbasi and Gross, 2020), even at the single-trial level (Flinker et al., 2010). Moreover, directed connectivity measures revealed a top-down effect from motor to auditory cortex, particularly in theta and beta frequency ranges (Abbasi and Gross, 2020; Abbasi et al., 2023), with enhanced brain-to-speech synchronization linked to the strength of this top-down influence (Park et al., 2015).

Temporal characteristics of the motor cortex and rate-restricted auditory motor coupling. Speech perception, being a temporal task (Kotz and Schwartze, 2010), requires high temporal precision and anticipation at multiple timescales (Rosen, 1992). One functional proposal of the motor system's top-down influence during speech perception pertains to temporal predictions. Specifically, it has been hypothesized that top-down effects of the extended sensorimotor system on auditory cortex modulate the entrainment to speech (and other auditory) signals in auditory cortex (Schroeder et al., 2010; Arnal, 2012), likely supporting optimal alignment of non-periodic signals through temporal predictions (Morillon et al., 2016).

When listening to temporally regular auditory sequences, individuals automatically synchronize their movement to the rhythm, like clapping along at a concert, illustrating the motor system's sensitivity to temporal structure (Schubotz, 2007; Arnal et al., 2015). Some proposed that the motor system can be conceptualized as an amodal network involved in timing processes and time perception, including timescales relevant to speech processing (Merchant and Yarrow, 2016). Particularly, brain rhythms in the beta frequency band, the default frequency regime of the motor system (Engel et al., 2001; Brookes et al., 2011; Keitel and Gross, 2016), have been associated with the representation of temporal information (Iversen et al., 2009; Fujioka et al., 2012; Kulashekhar et al., 2016). For instance, Fujioka et al (2012) observed that the motor system tracks isochronous tones during passive listening. Moreover, the beta rebound (resynchronization) triggered by stimulus presentation was modulated by the stimulus rate and maximal prior to a forthcoming sound, suggesting a predictive function of sensorimotor cortex in auditory perception (Arnal, 2012; Fujioka et al., 2012).

These results, alongside others, gave rise to theoretical accounts proposing that the motor system's function is mainly predictive (Grahn and Rowe, 2009; Arnal, 2012; Arnal and Giraud, 2012), capitalizing on its temporal characteristics. Combining its timing and predictive abilities, the motor system likely generates temporal predictions, providing these in a feed-forward fashion to the sensory cortices. Mechanistically, this predictive top-down influence is believed to modulate auditory processing by phase-resetting ongoing auditory brain rhythms (Park et al., 2015; Morillon and Baillet, 2017), potentially serving as mechanism for covert active sensing (Schroeder et al., 2010; Morillon et al., 2015; Rimmele et al., 2018).

The interaction between auditory and motor brain rhythms is proposed to favor certain frequencies. While the beta band dominates the motor system both at rest and during motor action (Brookes et al., 2011; Hillebrand et al., 2012; Groppe et al., 2013), recent work highlights the relevance of the delta and theta frequency regimes in auditory-motor coupling during speech perception (Park et al., 2015; Assaneo and Poeppel, 2018; Assaneo et al., 2019; Morillon et al., 2019; He et al., 2023). These frequencies align with the preferred frequencies of auditory cortex (Giraud et al., 2007; Keitel and Gross, 2016). Assaneo and Poeppel (2018) reported optimal auditory-motor coupling at 4.5 Hz

when listening to syllables at the same rate (4.5 syllables/s), compared to slower or faster (2.5, 3.5, 5.5, 6.5 syllables/s) rates. This rate-restricted interaction between the auditory and motor systems, together with computational modelling results (Assaneo et al., 2021), suggests that speech perception may be supported by an oscillatory auditory-motor coupling process.

Overall, the findings discussed in this section suggest that the auditory and motor systems are highly integrated during speech perception. One hypothesis is that top-down influences from the motor system phase-reset ongoing rhythmic activity in auditory cortex during speech perception. This phase-reset is believed to enhance perceptual acuity, providing temporal information about the onset of forthcoming sensory events (Arnal, 2012; Haegens and Zion Golumbic, 2018; Rimmele et al., 2018). Likely, the influence is optimized at frequencies matching the natural statistics of speech (~4.5 Hz).

1.4 Plasticity in endogenous rhythms

Behaviorally, individuals vary in their temporal abilities and this variability likely affects their speech perception. However, it is unclear how such behavioral variability comes about. To test the validity of the proposed mechanism of speech entrainment, the current thesis aims to assess the plasticity of endogenous brain rhythms, particularly in auditory cortex, and how these relate to both speech tracking and comprehension. If brain rhythms and entrainment are important pillars of speech comprehension, individual variation in behavioral performance should be reflected in the neural underpinnings and vice versa. In the context of speech perception, a few parameters are plausible candidates giving rise to behavioral variation, for instance flexibility of brain rhythms. This thesis approaches flexibility in two ways by (1) studying how brain rhythms adapt in the case of brain reorganization and (2) capitalizing on individual differences. Particularly, the current thesis focuses on the preferred frequency of motor cortices and the synchronization between rhythms in auditory and motor brain areas as estimate of coupling strength. Higher endogenous theta frequencies in auditory cortex may be beneficial, as they may allow for better auditory sampling and, hence better phase-alignment to faster speech. Similarly, stronger interactions between auditory and motor cortex may facilitate information flow, potentially allowing for better temporal anticipation of auditory events. Such interaction should be facilitated if frequencies in both systems are similar (Fröhlich, 2015; Notbohm et al., 2016).

1.4.1 Spectral plasticity due to early reorganization and life-long learning

One approach for investigating the flexibility of endogenous brain rhythms involves studying how they are altered in the case of brain reorganization and behavioral adaptations. Congenitally blind individuals represent an ideal population for studying these phenomena. Study 1 of the present thesis will follow this approach.

In congenital blindness, individuals experience total loss of vision, prompting adaptive behavior and neural reorganization. Neuroimaging studies indicate altered structural and functional cortical properties in the brains of congenitally bling individuals such as decreased surface and volume (Noppeney et al., 2005; Noppeney, 2007), as well as increased thickness (Jiang et al., 2009; Park et al., 2009). Functional changes include the repurposing of visual areas during various non-visual tasks, *crossmodal* plasticity, (Burton, 2003; Pascual-Leone et al., 2005; Bedny et al., 2011; Voss and Zatorre, 2012; Gudi-Mindermann et al., 2018; Rimmele et al., 2019) and reorganization of the intact cortices (Roeder et al., 1996; Elbert et al., 2002), *intramodal* plasticity (Roeder and Neville, 2003). Furthermore, functional interactions between visual cortex and other cortical areas may be altered (Liu et al., 2007; Yu et al., 2008; Klinge et al., 2010; Bedny et al., 2011; Burton et al., 2014; Pelland et al., 2017).

Whether the neural reorganization in congenital blindness is reflected also in altered brain rhythms remains largely unknown. While a reduced or absent visual alpha rhythm is consistently observed in congenitally blind individuals (Adrian and Matthews, 1934; Noebels et al., 1978; Kriegseis et al., 2006; Hawellek et al., 2013; Schubert et al., 2015), studies that investigate brain rhythms beyond alpha rhythms and the visual cortex are scarce. One such study reported increased connectivity in the delta and gamma ranges within visual cortex, alongside preserved alpha connectivity between visual cortex and other cortical areas, despite decreased visual alpha power (Hawellek et al., 2013). Furthermore, beta-band connectivity involving visual cortex has been suggested to be increased (Gudi-Mindermann et al., 2018; Rimmele et al., 2019). Together, these results suggest that sensory deprivation may induce spectral plasticity, but the specific brain areas and spectral bands affected remain mostly uncertain.

The notion of plastic brain rhythms is intriguing considering the behavioral adaptations seen in congenitally blind individuals. Congenitally blind individuals demonstrate advantages in auditory (Bull et al., 1983; Lessard et al., 1998; Roeder et al., 2004; Gougoux et al., 2005; Stevens and Weaver, 2005; Foecker et al., 2012), tactile (Hoetting et al., 2004; Roeder et al., 2004) and higher-level cognitive tasks (Roeder et al., 2001, 2007; Amedi et al., 2003; Roeder and Roesler, 2003; Rimmele et al., 2019), compared to normally sighted controls. Notably, they exhibit remarkable abilities in understanding ultra-fast speech (Trouvain, 2007; Hertrich et al., 2009, 2013; Dietrich et al., 2013), beyond syllabic rates typically understood. This observation challenges the mechanistic relevance of the theta rhythm for speech segmentation with preferred frequency around 4.5 Hz, for which synchronization with a stimulus at 20 Hz should be challenging.

One hypothesis is that congenitally blind individuals have altered theta rhythms with higher endogenous frequencies. To isolate endogenous characteristics of brain rhythms from their evoked responses to stimulation, resting state measurements provide a good opportunity. Such resting-state

signatures can then be linked to task-induced activity, directly testing the relationship between resting and evoked activity. Few studies have done this so far, however, these studies demonstrated predictive power of endogenous brain rhythms on task-related processing (Keitel and Gross, 2016; Becker and Hervais-Adelman, 2023; Gunasekaran et al., 2023).

1.4.2 Individual differences in endogenous rhythms, auditory-motor synchronization, and behavior

A more nuanced approach to comprehending the influence of the preferences of endogenous rhythms and auditory-motor synchronization in speech comprehension entails studying individual differences in both preferred behavioral and endogenous neural rhythms. Studies 2 and 3 will pursue this approach, with Study 2 emphasizing behavioral measures and Study 3 combining behavioral and neurophysiology measures.

Preferred auditory and motor rates can be measured behaviorally. Preferred perceptual auditory rates have been quantified using preference tasks, ratings, or as the rate of best performance in different auditory tasks (McAuley et al., 2006; Kaya et al., 2023). In the motor domain, preferred rates are often determined as natural production rates such as tapping, drumming, or walking (Drake et al., 2000; McAuley et al., 2006). Speech rate is another indicator of natural motor production rates. The average speaker produces speech at a rate of 4-5 syllables per second, reflected in the modulation spectrum of the speech acoustics peaking at 4-5 Hz (Pellegrino et al., 2011; Varnet et al., 2017; Poeppel and Assaneo, 2020). There is, however, considerable speech rate variability between (and even within) individuals. Generally, behavioral preferred rates are believed to mirror the endogenous rates of underlying neural oscillators and are presumed to be reliably measurable across different time points. Such test-retest reliability has been demonstrated in tapping (Kaya and Henry, 2022). However, tapping rates are subject to individual variation, as they change across the lifespan (McAuley et al., 2006; Kaya et al., 2023) and with musical training (Drake et al., 2000).

According to dynamics system theory, preferred rhythms are linked to optimal behavior, with a system's preferred rate acting as an attractor toward which individuals tend to gravitate back to over time (Strogatz and Stewart, 1993). Empirically, preferred motor rates consistently predict auditory-motor synchronization, as individuals synchronize best at a rate close to their own preferred motor rate (Loehr and Palmer, 2011; McPherson et al., 2018; Scheurich et al., 2018). While this optimal synchronization is often interpreted as indicative of enhanced perception, empirical evidence supporting this notion is limited. However, recent work demonstrated such a correlation between preferred motor (but not auditory) rhythms and auditory duration discrimination (Kaya et al., 2023).

Consistent with behavioral variations, individual differences in endogenous neural frequencies can affect neural processing and behavioral performance. A well-studied example is the alpha band, the endogenous rhythm of the visual system. The frequency of the alpha rhythm demonstrates variability

across individuals and is recognized as a stable and heritable neurophysiological trait (Lykken et al., 1974; Gasser et al., 1985; Kondacs and Szabó, 1999; Van Beijsterveldt and Van Baal, 2002; Smit et al., 2006). The alpha rhythm is generally implicated in various cognitive and perceptual tasks (Klimesch, 1999), potentially influencing the segmentation of visual inputs and thereby determining the temporal resolution of visual perception (VanRullen and Koch, 2003). Moreover, higher individual alpha frequencies predict enhanced flash discrimination, suggesting superior temporal resolution (Cecere et al., 2015; Samaha and Postle, 2015), while alteration of individual alpha frequency via transcranial alternating current stimulation can modify the temporal integration window (Cecere et al., 2015). For frequency bands other than the alpha band, there has been limited exploration of the relationship between individual peak frequency of endogenous rhythms and behavioral performance. Focusing on the delta band, Gunasekaran et al. (2023) investigated whether endogenous delta peaks were linked to preferred rates of finger tapping and covert counting. However, no correlation between delta peak frequency residuals and behavioral rates was found. Endogenous theta rhythms in Heschl's gyrus reportedly are modulated in amplitude during speech listening, with increased amplitude during listening (Keitel and Gross, 2016). While Keitel and Gross (2016) did not establish a link to behavioral performance (i.e. passive listening task), syllabic tracking in right STG and left middle temporal gyrus demonstrated a positive association with individual speech-in-noise comprehension (Becker and Hervais-Adelman, 2023). However, in the latter study, the authors did not assess a link to endogenous rhythms. Overall, the relationship between endogenous (theta) rhythms and speech comprehension remains unclear, and there is a lack of comprehensive studies integrating endogenous theta rhythms, speech tracking, and behavioral performance. The present thesis seeks to address this gap in the literature, focusing particularly on the role of individual differences.

Individual differences in preferred perceptual rates and endogenous (theta) peaks in the auditory domain are still relatively unexplored. As discussed above (see chapter 1.1.2), the motor system may provide temporal predictions during speech processing and operate optimally at similar frequencies as the auditory system. With respect to individual differences, recent work suggests significant variability in auditory-motor synchronization across the population (Assaneo et al., 2019). Using a simple behavioral protocol, the spontaneous speech synchronization (SSS) test, Assaneo et al. (2019) suggest that half the population spontaneously synchronizes their own motor output (whispered syllables) to an ongoing speech stimulus (stream of syllables), while the other half does not. This dichotomy between "high and low synchronizers" was reflected in functional and structural brain connectivity and various behavioral paradigms, with better behavioral performance for "high synchronizers" (Assaneo et al., 2021; Kern et al., 2021; Barchet et al., 2024). Interestingly, "high synchronizers" (as measured with the SSS-test) tend to have more musical training (Rimmele et al., 2022). This finding implies that one of the reasons for the motor system's contribution to individual

differences in auditory perception may be related to its susceptibility to training. This hypothesis is supported by the observation that musicians, compared to non-musicians, display enhanced functional and structural connectivity at rest (Bangert et al., 2006; Oechslin, 2010; Herholz and Zatorre, 2012; Luo et al., 2012; Zamorano et al., 2017) and during task-relevant processing (Chen et al., 2008b; Grahn and Rowe, 2009; Du and Zatorre, 2017; Li et al., 2021).

1.5 Research questions and hypotheses

The auditory system tracks the temporal dynamics of speech and other acoustic signals. Yet, whether speech tracking reflects “entrainment in the narrow sense” (Obleser and Kayser, 2019), engaging endogenous brain rhythms, remains uncertain (Haegens and Zion Golumbic, 2018; Zoefel et al., 2018b; Lakatos et al., 2019; Obleser and Kayser, 2019). The primary objective of this thesis (see Figure 1.1) is to advance our understanding of whether speech processing engages endogenous theta rhythms in auditory cortex. Specifically, it examines whether endogenous theta peak frequencies of the auditory and motor systems predict speech comprehension and tracking. While alpha band peak frequencies are known to vary across individuals, similar research in the theta band is limited despite its importance for oscillatory speech perception models. Thus, a secondary yet pivotal goal is to delineate the individual variation in endogenous theta peak frequencies of the auditory and motor systems, alongside their corresponding preferred behavioral rates.

Specifically, this thesis seeks to advance this ongoing debate by addressing the following questions: (1) Do individual differences in endogenous theta brain rhythms account for behavioral and neural differences among individuals? More specifically, do spectral profiles change with brain reorganization, particularly in populations exhibiting adaptive behaviors such as congenitally blind individuals? (2) Do we observe individual differences in the peak frequency of preferred auditory and motor rhythms and in auditory-motor synchronization in normally sighted individuals? Are speech comprehension abilities affected by the individual preferred rates of the auditory and motor systems? (3) Are the effects of endogenous auditory and motor cortices on speech comprehension related to modulations of speech tracking?

Study 1 aims to investigate potential modifications in brain rhythms in congenital blindness. Congenital blindness is characterized by brain reorganization and behavioral adaptations, including heightened auditory acuity and the ability to understand heavily accelerated speech. We hypothesized that if speech tracking relies on endogenous theta brain rhythms, congenital blindness may prompt adaptations to the spectral profiles, particularly in auditory cortex. Therefore, we posited that congenitally blind individuals would exhibit higher endogenous theta frequencies in auditory cortex, possibly enhancing their ability to comprehend fast speech. While we were particularly interested in

auditory cortex, this study adopts a whole-brain approach, providing the first comprehensive characterization of spectral profiles in congenital blindness.

Study 2 quantifies the temporal characteristics of individuals' auditory and motor systems, their impact on speech comprehension, and how comprehension may be influenced by motor-based predictions and higher-level factors. In two behavioral experiments, we measured preferred auditory and motor rates, auditory-motor synchronization, and speech comprehension across various syllabic rates. Our hypothesis posited that individuals with higher preferred auditory and motor rates would demonstrate enhanced comprehension of faster speech, owing to a superior temporal resolution. We further expected that stronger auditory-motor synchronization would predict better speech comprehension, particularly at faster syllabic rates, as it may facilitate the transfer of temporal predictions. Additionally, we examined the contributions of individual auditory working memory and sentence predictability, higher-level factors known to influence speech comprehension.

Study 3 integrates elements of Studies 1 and 2 to ascertain whether the associations between behavioral preferred rates, auditory-motor synchronization, and speech comprehension observed in Study 2 are reflected in corresponding neural dynamics. Specifically, as speech tracking is believed to reflect segmentation (Giraud and Poeppel, 2012), this study aims to elucidate whether the preferred frequencies of the auditory and motor systems, and auditory-motor synchronization, contribute to comprehension at the level of syllabic segmentation. We hypothesized that faster endogenous peak frequencies in the auditory and motor cortex, along with stronger auditory-motor coupling, would enhance speech tracking, supporting an oscillatory model of speech processing. To address this hypothesis, we conducted a combined behavioral and MEG protocol, encompassing a purely behavioral session, a behavioral task during MEG acquisition, and a resting-state MEG measurement.

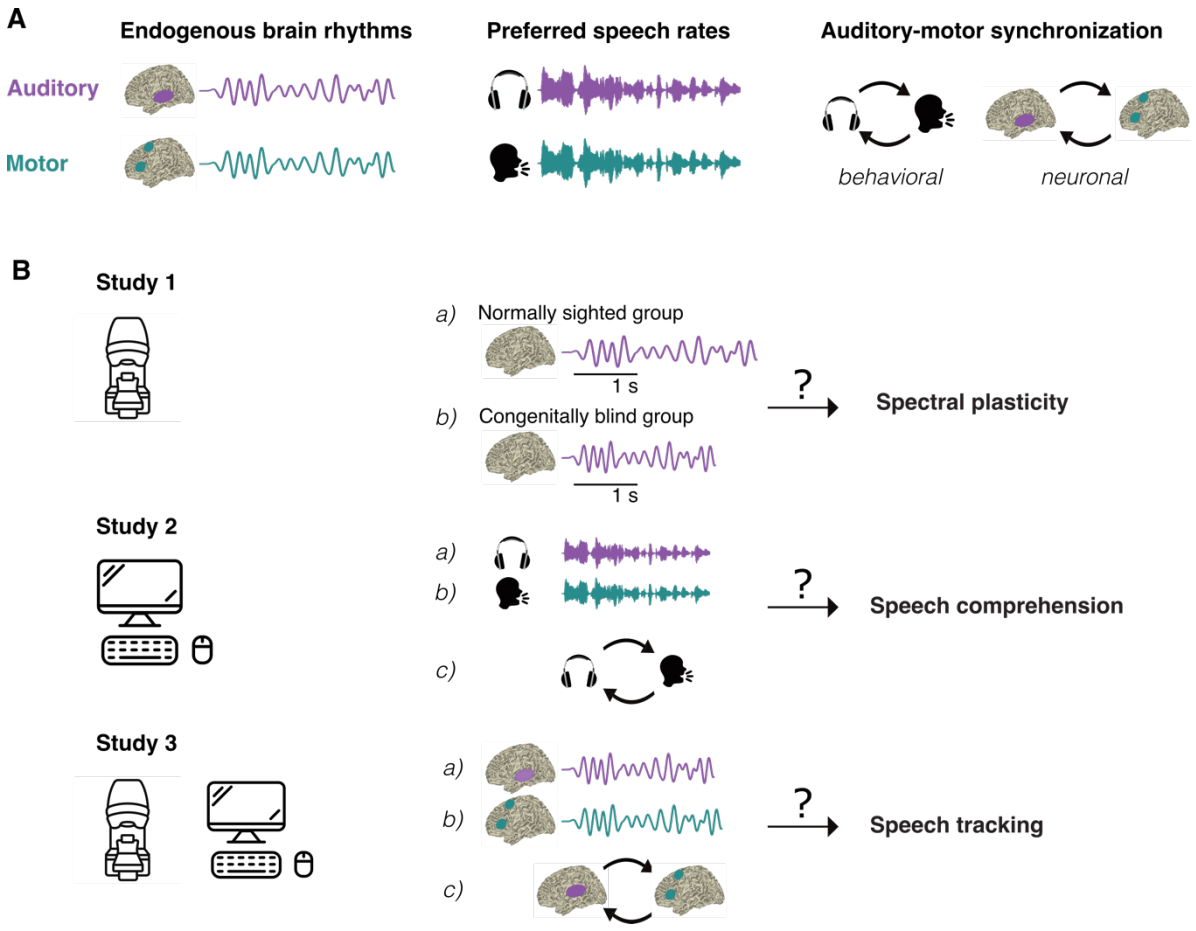


Figure 1.1. Conceptual overview of the three studies presented in this thesis. The upper row (**A**) illustrates the three key phenomena hypothesized to affect speech comprehension and neural processing of speech: endogenous brain rhythms in auditory and motor cortex, behavioral preferred auditory and motor rates (preferred perceptual speech rate and spontaneous motor production rate), and behavioral and neural auditory-motor synchronization. (**B**) Rows display the central research question of each study. Study 1 (top) explored whether endogenous brain rhythms, measured as spectral profiles, undergo plasticity in congenital blindness, with a focus on potential adaptations in the theta rhythm in auditory cortex. Study 2 (middle) investigated whether behavioral preferred rhythms and auditory-motor synchronization predict speech comprehension. Study 3 (bottom) examined whether speech comprehension is affected by endogenous auditory and motor brain rhythms and auditory-motor coupling at the level of syllabic segmentation (i.e. speech tracking), or a later process.

2 General methods

2.1 Speech materials

(More) Naturalistic speech stimuli. In line with traditional linguistic research, the cognitive neuroscience of speech (and language) has predominantly studied how the brain processes speech from the perspective of isolated syllables, words or simple sentences (Brennan, 2022). These stimuli are typically specifically designed for experiments, controlling various linguistic variables (Hamilton and Huth, 2020). Controlling stimuli and experiments is general practice in many fields of cognitive neuroscience, aiming to minimize potentially confounding variables and isolate the brain activity underlying the cognitive process of interest (Nastase et al., 2020). However, there is growing interest in using more naturalistic stimuli (Hamilton and Huth, 2020; Nastase et al., 2020) as research suggests that findings from experiments using controlled and simplified stimuli may not seamlessly translate to naturalistic settings (Hamilton and Huth, 2020). This underscores the need of employing more naturalistic stimuli to understand brain and behavior beyond controlled settings.

To incorporate more naturalistic stimuli, the experiments in this thesis included sentences extracted from books and audiobooks. Consequently, these sentences were neither specifically created for the experiment, nor controlled along any linguistic dimensions. Instead, they varied widely in topic, vocabulary, and memory load, owing to differences in sentence length. However, in speech comprehension tasks, the sentences were time-compressed, randomized and presented individually, rather than as continuous narratives. While such stimuli could be encountered in everyday experiences, it is unlikely that participants would listen to them in the same context, i.e. a laboratory, or mode, i.e. time compressed. Thus, while prioritizing engaging and naturalistic materials, we balanced ecological validity with the requirements of the experimental designs to address our research questions.

Stimulus selection. Studies 2 and 3 investigate how speech comprehension and tracking change with increasing syllabic rates. Speech rate variations are often investigated by manipulating the degree of time compression, with findings indicating a decline in comprehension as compression (and hence speech rate) increases (Ahissar et al., 2001; Brungart et al., 2007; Ghitza and Greenberg, 2009). Time-compressed speech undergoes a linear reduction in duration (Janse, 2004). As such, comparing conditions differing in degree of time compression generates acoustically different experimental conditions. To avoid systematic acoustic differences between conditions, here we controlled the degree of compression across syllabic rate conditions.

For uncompressed speech, the modulation spectrum displays a prominent spectral peak around 4.5 Hz (Ding et al., 2017; Varnet et al., 2017; Poeppel and Assaneo, 2020). To ensure consistent degrees of compression across different syllabic rate conditions in our experiment, speech

recordings needed to exhibit maximal variation in speech rate before compression. To achieve this, in addition to sampling sentences from existing audiobooks, we conducted speech recordings with instructions for speakers to vary their speech rate. For these recordings, we recruited native speakers of North American English (Study 2) and German (Study 3).

Sentences used in Study 2 (experiment 2) and Study 3 were sourced from freely available books and audiobooks on platforms such as Zeno, Librivox, and Lit2Go. Eligible sentences, screened for semantic correctness and grammatically simplicity with a maximum of one embedded clause (for similar selection criteria: Hervais-Adelman et al., 2015; Pefkou et al., 2017), were extracted from audiobooks. Recorded sentences were extracted from text and recorded in a sound-attenuated recording booth. During recording, sentences were displayed on a screen and speakers first read them silently for familiarization, then aloud at a normal pace and, finally, as quickly or slowly as possible while maintaining clear articulation. Speakers were encouraged to repeat recordings if errors occurred or if they believed they could increase/decrease speaking pace, receiving additional feedback and requests for repetitions as needed. After recording, pauses longer than 300 ms were removed from all files to avoid wrong speech rate estimates. To yield the desired speaking rate at which the files were to be presented in the experiments, the original files were time-compressed and time-expanded using the Pitch Synchronous Overlap and Add (PSOLA; Moulines and Charpentier, 1990) algorithm. This algorithm alters the formant patterns and other spectral properties of speech in duration while preserving the fundamental frequency (‘pitch’) contour.

2.2 Magnetoencephalography

Brain rhythms can be measured using various neurophysiological methods, including magnetoencephalography (MEG). MEG is a non-invasive method capturing magnetic fields emitted from the brain with both high temporal and solid spatial resolution, using superconductive sensors positioned in a helmet surrounding the participant’s head (Hari and Puce, 2023).

MEG capitalizes on the fundamental physical principle that any electric current is associated with magnetic induction (Baillet, 2017). As introduced above (chapter 1.1.1), rhythmic neural activity arises from fluctuations in membrane potentials, creating electric currents. The neurophysiological currents predominantly depicted by MEG signals are postsynaptic potentials of pyramidal cells in the neocortex (Lopes da Silva, 2013; Gross, 2019). While currents from single neurons are too weak to be detected outside the scalp using non-invasive methods, synchronized electrical current flow across neighboring cells produces a magnetic field of sufficient strength (i.e. in the femto-tesla range) for detection by MEG sensors (Hämäläinen et al., 1993). Thus, MEG signals reflect fluctuations in the membrane potential of thousands of neurons (Murakami and Okada, 2006), with the

signal strength influenced by the number of activated neurons, the extent of their synchrony, and their proximity (Gross, 2019).

MEG provides high temporal resolution, capturing the rapid temporal dynamics of neural processes on the order of milliseconds. Due to the minimal distortion of magnetic fields by head tissues, MEG offers superior spatial resolution compared to other electrophysiological measures like EEG, especially when combined with magnetic resonance imaging (MRI), facilitating the interpretation of MEG signals (Baillet, 2017). Source reconstruction methods enable the estimation of the neural sources underlying the signals measured by sensors outside the brain. Source reconstruction is based on two fundamental concepts: the inverse and the forward problem, with the inverse problem being ill-posed as there is no unique solution to it. Forward modelling involves estimating the magnetic field distribution of a known neural source. In contrast, inverse modelling aims to estimate unknown neural sources generating the magnetic field measured at the MEG sensors. To refine solutions to the inverse problem, forward models are combined with an inverse modelling technique that imposes mathematical constraints (Gross, 2019). One such modelling technique is beamforming, also known as spatial filtering, which involves predefining (putative) source locations, for each of which a set of weights is computed. These weights are applied to sensor data, such that the resulting source estimate is a weighted contribution of each sensor (Westner et al., 2022). While there exist many different types of beamformers (Kim and Davis, 2021), the one used in the current work is the linearly-constrained minimum variance (LCMV) beamformer (Van Veen et al., 1997).

MEG data, being a time-series, lends itself to analyses in both the time and frequency domains. Time-domain analysis, rooted in EEG research, focuses on the temporal response of the brain evoked by presented events or performed actions in experimental paradigms. Assuming that the signal contains both noise and signal of interest, event-related fields are computed by averaging over multiple trials, aiming to isolate consistent brain responses time-locked to events and eliminate task-irrelevant noise. Typically, such an analysis includes the comparison of component parameters, such as relative latency, topography, and amplitude between conditions or experimental groups (Gross, 2019; Hari and Puce, 2023). However, the explanatory power of event-related fields for neurophysiological mechanisms is limited (Cohen, 2014).

In contrast, spectral analyses offer valuable insights into brain rhythms and potential oscillatory dynamics, offering more straightforward interpretations of neurophysiology. Spectral analyses transform time-domain MEG data into the frequency or time-frequency domain, where signals are represented as a linear combination of oscillatory basis functions. Common spectral analyses methods, including the Fourier transform, wavelet transform, and Hilbert transform, yield a complex-valued representation of the data in the time-frequency plane. From these complex-valued representations, multiple indices can be derived, such as the amplitude and phase of oscillatory

brain components (Gross, 2014), which are of great interest for the research on neural oscillations. The work presented in the current dissertation particularly focusses on the analysis of phase information in both speech and brain signals and thus, primarily employs spectral analyses. These spectral analyses are combined with clustering (Study 1 and 3) and information-theoretical (Study 3) approaches. Lastly, one caveat of spectral analysis methods is the assumption of signal stationarity because brain data are highly non-stationary. This poses challenges for the successful identification on neural oscillations and calls for cautious interpretations (Gross, 2014).

3 Study 1[1]: Data-driven classification of spectral profiles reveals brain region-specific plasticity in blindness

3.1 Abstract

Congenital blindness has been shown to result in behavioral adaptation and neural reorganization, but the underlying neural mechanisms are largely unknown. Brain rhythms are characteristic for anatomically defined brain regions and provide a putative mechanistic link to cognitive processes. In a novel approach, using magnetoencephalography resting state data of congenitally blind and sighted humans, deprivation-related changes in spectral profiles were mapped to the cortex using clustering and classification procedures. Altered spectral profiles in visual areas suggest changes in visual alpha-gamma band inhibitory-excitatory circuits. Remarkably, spectral profiles were also altered in auditory and right frontal areas showing increased power in theta-to-beta frequency bands in blind compared to sighted individuals, possibly related to adaptive auditory and higher-cognitive processing. Moreover, occipital alpha correlated with microstructural white matter properties extending bilaterally across posterior parts of the brain. We provide evidence that visual deprivation selectively modulates spectral profiles, possibly reflecting structural and functional adaptation.

[1] This section is highly adopted from a published article (**Lubinus, C.**, Orpella, J., Keitel, A., Gudi-Mindermann, H., Engel, A. K., Roeder, B., & Rimmele, J. M., 2021, Cereb Cortex). CL and JR designed the research question; CL preprocessed and analyzed all MEG data; JR and AK supervised the analysis; CL wrote the first draft of the manuscript; HGM and JR recorded the data in a previous project under supervision of AE and BR; JO performed the DTI analysis. All authors revised the manuscript and approved its final version.

3.2 Introduction

Brain rhythms occur ubiquitously across cortex (Buzsáki, 2004; Buzsáki et al., 2013; Singer, 2018) and have been related to cognitive functions. Investigating changes in brain rhythms in congenital blindness can yield crucial insights into the neural mechanisms underlying the behavioral changes in this population. Congenitally blind individuals outperform sighted controls in various auditory, tactile, and higher-cognitive tasks (Bull et al., 1983; Lessard et al., 1998; Roeder et al., 2001; Amedi et al., 2003; Gougoux et al., 2005; Foecker et al., 2012). This includes tasks where temporal processing is essential, such as temporal order processing, musical meter processing or ultra-fast speech comprehension (Hoetting et al., 2004; Roeder et al., 2004, 2007; Stevens and Weaver, 2005; Moos and Trouvain, 2007; Trouvain, 2007; Hertrich et al., 2009; Dietrich et al., 2013; Lerens et al., 2014; Carrara-Augustenborg and Schultz, 2019; Zhang et al., 2019). Enhanced behavioral performance in auditory or tactile tasks among congenitally blind individuals is accompanied by cross-modal plasticity, wherein the “visual” cortex is engaged during non-visual tasks, potentially aiding compensatory task performance (Burton, 2003; Pascual-Leone et al., 2005; Bedny et al., 2011; Voss and Zatorre, 2012; Gudi-Mindermann et al., 2018; Rimmele et al., 2019). Additionally, intramodal plasticity occurs in congenitally blind individuals, with altered (mostly increased) neural activity in the auditory (Elbert et al., 2002; Stevens and Weaver, 2005; Gougoux et al., 2009; Schepers et al., 2012; Hertrich et al., 2013; Watkins et al., 2013) and somatosensory (Pascual-Leone and Torres, 1993; Roeder et al., 1996; Sterr et al., 1998a, 1998b) cortices during task processing. Intramodal plasticity has been related to early postnatal sensory experiences shaping the functional development of cortical representations, as shown in animal research (Zhang et al., 2001; for review: Rauschecker, 2008). For instance, reorganization of primary somatosensory and auditory areas has been reported in cats deprived of vision from birth (Rauschecker et al., 1992; Korte and Rauschecker, 1993). Despite these observed behavioral and neural adaptations, the mechanisms underlying visual deprivation-related plasticity remain poorly understood.

Studying brain rhythms in congenitally blind individuals may offer deeper insights into the mechanisms of neural plasticity, as these rhythms reflect neural circuitry activity and have been related to various sensory, motor and higher-cognitive processes. Brain rhythms are believed to reflect the synchronization (phase-alignment) of oscillatory activity across neural populations, supporting the formation of both local assemblies and large-scale functional networks (Engel et al., 1992; Hipp et al., 2011; Raichle, 2011; Siegel et al., 2012; Singer, 2013). Anatomical areas exhibit characteristic spectral signatures, so called spectral profiles (Giraud et al., 2007; Keitel and Gross, 2016), assumed to reflect intrinsic brain rhythms (Keitel and Gross, 2016). Intrinsic brain rhythms, measured during resting state, reflect the brain’s functional organization and default architecture recruited during task-related processes (Smith et al., 2009; Deco et al., 2011; Raichle, 2011, 2015; Engel et al., 2013; Keitel and Gross, 2016; Sormaz et al., 2018). Importantly, these brain rhythms oscillate

in a characteristic way even when not engaged in task-related processing (Brookes et al., 2011; Deco et al., 2011). For example, Keitel and Gross (2016) demonstrated a characteristic pattern of delta-, theta-, alpha-, and beta-band activity in the primary auditory cortex during both rest and speech comprehension. Although the overall spectral profile of primary auditory cortex was similar during speech comprehension and rest, the relative contribution of individual spectral clusters (i.e., their spectral power) was altered by the condition. During speech comprehension, compared to rest, the alpha band cluster was suppressed, indicating reduced inhibition of sensory cortices (Jensen and Mazaheri, 2010) while the other bands associated with several levels of speech processing were enhanced.

Various studies have observed that brain rhythms are recruited in a task-specific manner during sensory, motor and higher-cognitive tasks, linking intrinsic brain activity to cognitive functions (Engel et al., 1992, 2001; Singer and Gray, 1995; Lakatos et al., 2008, 2013; Holcombe, 2009; Giraud and Poeppel, 2012; Heusser et al., 2016; Portoles et al., 2018; Rimmele et al., 2018; Schroeder et al., 2018; VanRullen, 2018). For instance, in sighted individuals, the role of the alpha rhythm (~8–12 Hz) in cognition is well-established, with alpha- and gamma-band oscillations in occipital cortex coupling in excitatory-inhibitory cycles (Klimesch et al., 2007; Buffalo et al., 2011; Jensen et al., 2012). Specifically, the alpha rhythm inhibits task-irrelevant neural circuits, while gamma-band oscillations (30–100 Hz) allow for feedforward processing of task-relevant visual information (van Kerkoerle et al., 2014; Michalareas et al., 2016). In congenitally blind individuals, a reduced or absent occipital alpha rhythm has repeatedly been observed (Adrian and Matthews, 1934; Noebels et al., 1978; Kriegseis et al., 2006; Hawellek et al., 2013; Schubert et al., 2015). The lack of visual input during a sensitive period of brain development in congenitally blind individuals presumably results in atrophy or reorganization of posterior alpha generators (Ptito et al., 2008; Wang et al., 2013; Aguirre et al., 2016; Reislev et al., 2016). Posterior alpha generators have been located within the cortical and thalamo-cortical visual pathways (Lopes da Silva, 1991; Lőrincz et al., 2009) and posterior gamma generators within the cortical visual pathways (Bastos et al., 2014; Marshall et al., 2018).

In sighted individuals intrinsic theta-band oscillations (~4–7 Hz) in auditory cortex (Lakatos et al., 2005; Keitel and Gross, 2016) likely support the temporal segmentation of speech (Luo and Poeppel, 2007; Giraud and Poeppel, 2012; Gross et al., 2013; Doelling et al., 2014; Pittman-Polletta et al., 2021), music, and environmental sounds (Henry et al., 2014; Doelling and Poeppel, 2015). Congenitally blind individuals reportedly can comprehend ultra-fast speech—at rates where speech comprehension fails in sighted individuals—which may be linked to accelerated theta brain rhythms in auditory cortex, as well as the recruitment of additional brain areas such as the visual cortex (Trouvain, 2007; Hertrich et al., 2009, 2013; Dietrich et al., 2013; see also: Van Ackeren et al.,

2018). Furthermore, other brain rhythms outside of visual cortex might also be altered in congenitally blind compared to sighted individuals, for example beta- and gamma-band rhythms during working memory or semantic categorization tasks (Schepers et al., 2012; Gudi-Mindermann et al., 2018; Rimmele et al., 2019). However, the systematic understanding of alterations to spectral profiles outside of visual cortex due to congenital visual deprivation remains limited, and no clear hypothesis can be derived beyond the occipital alpha rhythm.

The goal of the present study is to characterize alterations in spectral profiles in congenitally blind as compared to sighted individuals in order to map visual deprivation-related spectral changes to brain areas. By studying a broad range of frequencies, instead of narrow frequency bands, our research provides the means towards linking changes in spectral profiles across the whole brain to adaptive behavior in congenitally blind individuals in future research. Standard analyses of brain rhythms face several problems (see Keitel and Gross, 2016), which possibly explains the lack of a comprehensive account of spectral changes in congenitally blind individuals beyond the alpha rhythm. Predominant neural activity in the alpha and super-low frequency ranges (1/f) complicates the analysis of spectral properties at lower frequencies. Here, we have employed a novel analysis pipeline, introduced by Keitel and Gross (2016), which reveals brain area-specific spectral profiles. The analysis pipeline overcomes the limitations of standard analyses of brain rhythms by using segment-based clustering (of source-localized Fourier spectra) to disentangle spectral properties in the lower frequency ranges and by taking the time course of activity into account. We extended the pipeline (Fig. 3.1) to quantify differences in spectral profiles across cortical brain areas between congenitally blind and sighted individuals using a cross-classification approach. This approach allowed us to identify brain areas with changed spectral profiles as a result of congenital blindness.

First, we expected to replicate Keitel and Gross (2016) by showing that spectral profiles in sighted adults are region-specific, enabling the classification of brain regions based on spectral profiles. Second, we hypothesized that within a group of congenitally blind individuals, spectral profiles would similarly follow specific patterns and enable the classification of brain regions. Third, visual deprivation-related plasticity was predicted to result in altered spectral profiles in congenitally blind individuals as compared to the sighted, particularly for brain regions for which visual deprivation-related reorganization (cross- and intra-modal plasticity) has previously been shown, such as for sensory cortices. A specific hypothesis regarding the changes in brain rhythms in congenitally blind individuals was only derived for the alpha rhythm, which was expected to be reduced in occipital cortex. Finally, we expected a relationship of microstructural white matter properties and spectral profiles of occipital cortex.

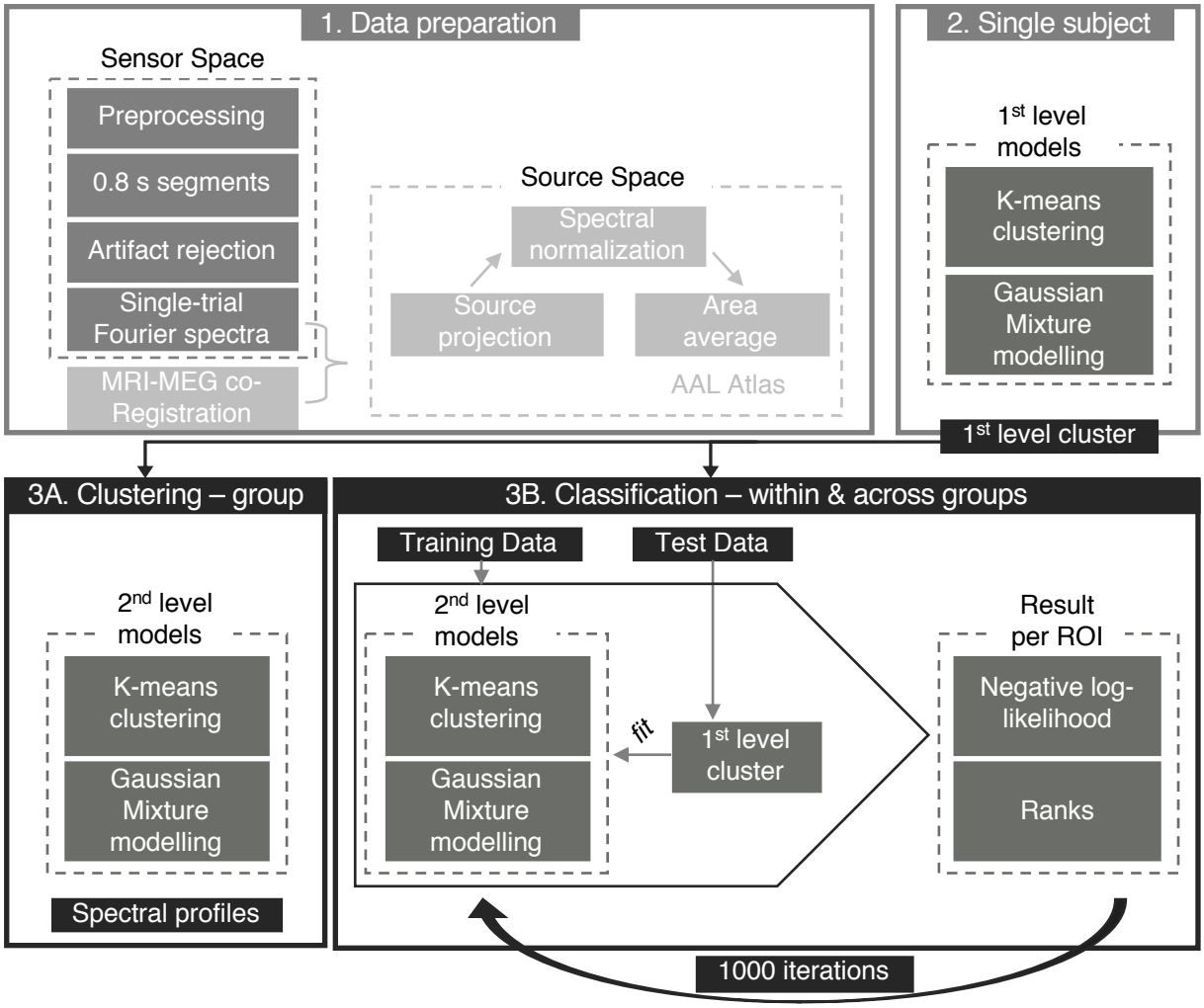


Figure 3.1. Analysis pipeline adapted from Keitel and Gross (2016)**. (1)** Resting-state MEG was preprocessed and segmented into trials of 0.8 s length. Complex Fourier spectra were computed for each trial separately and projected into source space using previously defined beamforming (LCVM) coefficients. The data were spatially normalized, dividing each voxel's power by the mean power of all trials and voxels and parcellated according to the AAL atlas, averaging power across voxels within an area (N = 115). **(2)** In the 1st-level analysis, power matrices were clustered into 9 spectral clusters per participant and brain region using k-means and Gaussian Mixture model algorithms. **(3A)** In the 2nd-level analysis, 1st-level individual clusters were subjected to k-means clustering and GMMs to establish region-specific spectral clusters consistent at the group level, referred to as spectral profile. The optimal number of group-level clusters per anatomical area was defined by the Silhouette Criterion. **(3B)** For the classification procedure the experimental group was divided into training and test set. The fit between 1st-level clusters of the test group and group-level clusters of the training set was calculated for each brain region. For each anatomical region, this resulted in a negative log-likelihood value for all regions (i.e., 115 values per region), indicating its similarity to all brain regions based on the spectral clusters. This training-test fitting procedure was repeated 1000 times with new group assignment (training vs. test) on each iteration. The interindividual variance within test and training groups of each iteration was controlled for with an additional 100 iterations within the respective sets.

3.3 Materials and Methods

The data reported in this study were recorded in the context of a larger working memory training project (Gudi-Mindermann et al., 2018; Rimmele et al., 2019) in the scope of which the participants

underwent behavioral and electrophysiological recordings in multiple sessions. Here, we exclusively analyze resting-state MEG data. The study was approved by the German Psychological Association.

Participants. Fifty-four German native speakers took part in this study, twenty-seven congenitally-blind adults and twenty-seven normally-sighted adults, matched in gender, age and education. To reduce eye movements during MEG data collection, the congenitally blind (CB) and normally-sighted (S-BF) participants were blindfolded. In addition, the sighted control group underwent a second MEG resting state measurement with open eyes (S-EO). A few participants were excluded due to corrupted MEG files (one subject for the CB, one subject for the S-EO) or lack of structural MRI scan (three subjects for the S-BF and S-EO). Thus, the final sample comprised of 26 participants for the CB group (12 females; mean age: 37.8 years; SD: 10.2 years; age range: 22–55 years), 24 for the S-BF group (11 females; mean age: 36.8 years; SD: 10.1 years; age range: 21–55 years) and 23 for the S-EO group (11 females; mean age: 37.3 years; SD: 9.8 years; age range: 21–55 years).

All participants reported normal hearing and no history of psychiatric or neurological disorders. As assessed by self-report, the sighted participants had normal or corrected-to-normal vision. In the blind, vision loss was due to a variety of peripheral (pre)natal conditions: retinopathy of prematurity: n=9; genetic defect, n=5; congenital optic atrophy: n=2; Leber's congenital amaurosis: n=2; congenital cataracts, glaucoma: n=2; congenital retinitis: n=2; binocular anophthalmia: n=2; retinitis pigmentosa: n=1; congenital degeneration of the retina, n=1. Vision loss resulted in total congenital blindness (n=8) or minimal residual light perception (n=17). All participants gave written informed consent prior to the participation and received monetary compensation.

MRI and MEG Data Acquisition. For all participants T1-weighted structural MRI scans and DWI-MRI scans were obtained with a 3T scanner (Siemens Magnetom Trio, Siemens, Erlangen, Germany). For the T1-weighted images we used the following parameters: TE = 2.98 ms, TR = 2300 ms, flip angle = 9, and isotropic 1 mm3 voxels, 256 sagittal slices. The MEG data were recorded in a magnetically shielded room using a 275-channel whole-head system (Omega, 2000, CTF Systems Inc.), while participants sat in an upright position. The data were acquired with a sampling rate of 1200 Hz. Prior to each experiment, the head position was measured relative to the MEG sensors and during the recording the head position was tracked.

Data Analysis. Some analyses in this study are adopted from the analysis pipeline proposed by Keitel and Gross (2016). The modifications to the analysis pipeline and the novel analyses are stated in detail below. All analyses were carried out using Matlab R2018a (version 9.4.0.813654, The Math Works, Inc.), the Fieldtrip Toolbox (version 20181104), and SPM12.

Preprocessing. During preprocessing, the MEG signal was downsampled to 250 Hz, denoised and detrended. To better capture the dynamically changing spectral properties of the brain, the continuous signal was segmented into trials of 0.8 s. Trials were declared as noisy and excluded when their z-score was higher than 2. On average, seven trials were excluded, resulting in a mean of 340.3 trials (SD= 34.7) per subject (S-EO: mean = 346.7, SD = 37.4; S-BF: mean = 336.8, SD = 34.2; CB: mean = 338, SD 33.2). Due to shorter recordings in the present study, trial duration was slightly shorter, relative to the 1 s duration used in Keitel and Gross (2016), to increase the amount of trials. MEG channels were labeled as noisy and rejected when the ratio between their noise level (in SD) and that of the neighboring sensors (in SD) exceeded a value of 0.5 ([Sensor SD – Neighbor SD] / Neighbor SD; mean number of excluded channels = 1.22, SD = 1.34). Finally, using independent component analysis (ICA), data was cleaned from heartbeat, eye blinks and eye movements-related artifacts (components were identified based on their time-course, topography and variance across trials).

Source Localization. To prepare the source projection of the Fourier spectra, beamformer coefficients were obtained. For this purpose, we applied co-registration of individual T1-weighted MRI scans and the MEG coordinate system, realignment, segmentation and normalization to Montreal Neurological Institute (MNI) space. A forward model was created using a single-shell model and linearly constrained minimum variance (LCMV) beamformer coefficients (Van Veen et al., 1997) were calculated for the MEG time series for each individual voxel on the 10 mm regular grid.

Spectral Analysis in Sensor Space. The analyses were performed separately for the three groups (CB, S-EO, S-BF; for an overview, see Fig. 3.1). First, Fourier spectra were calculated on 0.8 s long trials for each subject, using a multitaper approach (3 tapers) and zero-padding (length of 2 s). We slightly reduced the trial duration to match the overall number of segments used by Keitel and Gross (2016), i.e., 1 s, while accounting for our shorter MEG recordings. Second, using the previously computed LCMV coefficients, the complex Fourier spectra were projected into source space. Neural power, the squared amplitude of the Fourier coefficient, was computed for each time segment, voxel, and frequency. Neural power was ratio normalized, dividing by the mean power across all voxels and trials. This ratio normalization highlighted voxel-specific spectral properties with values above/below 1, compared to the mean spectral power across all voxels separately at each frequency. All values were subtracted by 1 (leading to values above/below zero), to facilitate the identification of changes in power (decreases/increases). We additionally conducted a control analysis without the normalization, performing the clustering and classification procedures on the non-normalized data.

k-Means Clustering and Gaussian Mixture Modelling of Source-Localized Spectral Activity. To identify region-specific spectral clusters in the individual subject, the brain was parcellated

according to the Automated Anatomical Labelling (AAL) atlas (Tzourio-Mazoyer et al., 2002; 116 regions of interest, ROIs). For one anatomical region (cerebellum 3L), however, the interpolation between the AAL atlas and the source model was not successful. Thus, this region was excluded and all analyses are based on the remaining 115 anatomical areas. For each of the ROIs, power spectra were averaged across voxels. Within a brain area, neural populations might express several distinct intrinsic brain rhythms, which together constitute the spectral profile of an area. Clustering algorithms were employed to identify spectral clusters, reflecting intrinsic brain rhythms. First, trial-by-frequency matrices were subjected to a k-means algorithm (MacQueen, 1967) which established spectral clusters by partitioning the n observations (0.8 s temporal segments) into k clusters. For the 1st-level analysis, the k was set to 9, based on the Silhouette criterion evaluation (Rousseeuw, 1987). Second, for each subject and ROI, Gaussian Mixture Models (GMMs; Reynolds and Rose, 1995) were fitted to the 9 clusters obtained from the k-means analysis (1st-level GMM). Next, in order to identify the optimal number of clusters per brain region across all subjects for the 2nd-level group analysis, the 1st-level GMMs were evaluated using the Silhouette criterion. Silhouette values were computed for cluster solutions in the range from 1 to 15, the fitting was repeated 1000 times. At the group level, k-means clustering was applied to the 1st-level clusters in order to disclose consistent patterns across subjects. The optimal number of clusters per brain area, as assessed by the Silhouette criterion evaluation (Rousseeuw, 1987), was used as k-parameter for the algorithm. As before, k-means results were fed into GMM revealing the final clusters per brain region (2nd-level GMM).

Clusters were considered for visualization only if they were reflective of the majority of participants. To facilitate reading of the spectral plots, group-level clusters were color-coded according to the frequency of the maximum amplitude of the cluster (peak frequency) (delta: 1–3.5 Hz, red; theta: 4–8 Hz, green; alpha: 8.5–12.5 Hz, blue; beta: 14–30.5 Hz, yellow; gamma: 33.5–100 Hz, magenta). Furthermore, we computed the relevance of each cluster per brain region by analyzing the amount of single subject trials during which a cluster was present. Group clusters (Fig. 3.1, step 3) were traced back to single subject clusters and the amount of trials that contributed to a single subject cluster (Fig. 3.1, step 2) was calculated and expressed as percentage. Percentages were averaged across subjects.

Automatic Within-Group Classification. A classifier was employed to test the specificity of region-specific spectral profiles. After splitting each group into half (training and test group), group-level clusters were calculated for the training group for all anatomical regions using k-means and GMM clustering. For each brain region and participant of the test group, the similarity of spectral profiles was assessed as compared to all brain regions of the 2nd-level group clusters of the training group by computing the negative log-likelihood for all pairs of regions. This procedure, i.e., group

assignment and classification, was repeated 1000 times (note that for the S-EO one subject was left out in every iteration to yield an even number of participants in training and test groups). On each iteration, an additional loop (N = 100) controlled for interindividual noise within a group by randomly drawing the adequate number of subjects (i.e., $N_{S\text{-}EO} = 11$, $N_{S\text{-}BF} = 12$, $N_{CB} = 13$) from the group with replacement, allowing a subject to enter multiple times or not at all. Put differently, within one iteration (N = 1000) each participant belonged to either the training or the test group. To account for individual differences, the group clusters were calculated 100 times with a different subset chosen from the respective group each time and finally averaged to obtain a robust group estimate. Based on the mode of clusters identified per brain region in the 2nd-level cluster analysis, the optimal number of clusters for the classification analysis was k = 2. Likelihood values were ranked and averaged across iterations (20% trimmed mean). For further comparisons, only corresponding ROIs (e.g., how is the Heschl ROI in the test set ranked based on the training set Heschl ROI) were considered.

In addition to the descriptive report of the classification performance, here we tested whether a specific ROI (of the test set) was classified significantly better by the corresponding area of the training set as compared to all other 115 ROIs. This allowed us to exclude the possibility that classification performance was caused by unspecific effects—that is, generic fingerprints (i.e. profiles). To this end, each region's mean rank (averaged across iterations) was tested against a distribution of classification ranks generated from all other ROIs (null distribution).

Automatic Cross-Group Classification. In order to identify differences in region-specific spectral properties between the CB and S-BF, we performed a cross-group classification. We employed the same classification procedure as before; however, the classifier was trained on one group (S-BF), while the other (CB) was utilized as the test set. As before, the classification procedure was repeated 1000 times, drawing a subset of N = 12 per group on every iteration. The randomization of subjects chosen on each iteration was identical to the one used for the within-group classification in the S-BF (this is the reason why N = 12, instead of using all subjects of both groups). Thus, differences in the classification, as reflected by the ranks, could not be caused by the training set per se. In order to understand whether some brain areas in the CB were not classified well based on the S-BF spectral profiles (i.e., whether the classification of brain regions was different in the cross-group condition as compared to the within-S-BF classification), we tested the cross-group classification mean ranks against the distribution of ranks from the same ROI from the S-BF group (here the null distribution). The distributions were generated by taking the classification rank of a corresponding area from training and test set (i.e., Calcarine) across all iterations. We calculated the 95th percentile of the distribution and tested whether the cross-group mean rank of the current region fell above (significant) or below (not significant) this threshold.

To further assess the spectral profiles of brain areas that were significantly different in the cross-group classification, post-hoc permutation statistics were applied to the raw, normalized region-specific spectra (i.e., Fourier spectra without clustering procedure). This additional analysis was performed to evaluate the cross-group differences between the sighted and the congenitally blind groups by using a simpler, more established approach. The analysis confirmed the findings from the spectral clustering approach. The spectral analysis was calculated as in the main analysis (see above). For all significant brain regions separately, power was averaged across voxels and segments, resulting in a single power value per frequency and per subject. Based on frequency by subject matrices for the CB and the S-BF, group differences in spectral power were tested against a distribution where the group assignment (CB vs. S-BF) was randomly permuted (N = 1000). To control for multiple comparisons, we used false discovery rate (FDR, Q = 0.05).

Microstructural White Matter Properties. DW-MRI data were acquired together with T1-weighted structural scans described above. We used an echo planar imaging (EPI) sequence optimized for DWI-MRI of white matter covering the whole brain (64 axial slices; bandwidth = 1502 Hz/Px, 104 × 128 matrix, TR, 8,200 ms; TE, 93 ms; flip angle, 90°; slice thickness, 2 mm; voxel size, 2 × 2 × 2 mm3). The protocol comprised three acquisitions yielding a total acquisition time of 9 min 51 s. This resulted in a total of 120 diffusion-weighted volumes with six interleaved non-diffusion-weighted volumes (b values of 1,500 s/mm2). DWI-MRI scans were acquired from a subset of the original sample of 16 blind and 12 sighted participants.

DTI-MRI preprocessing and analysis. Diffusion data processing initially corrected for eddy current distortions and head motion by using FMRIB's Diffusion Toolbox (FDT; FMRIB Software Library; FSL 5.0.1; http://www.fmrib.ox.ac.uk/ fsl/; Jenkinson et al., 2012). For a more accurate estimate of diffusion tensor orientations, the gradient matrix was rotated to correct for head movement, using the fdt_rotate_bvecs program in FSL. We then used the Brain Extraction Tool (Smith, 2002) for brain extraction, also part of the FSL distribution. Analysis continued with the reconstruction of the diffusion tensors using FSL's DTIFIT program. Fractional anisotropy (FA) and Radial diffusivity (RD) maps for each participant were calculated using the eigenvalues extracted from the diffusion tensors. Note that FA maps are required in the early stages of TBSS to compute the registrations to MNI standard space and subsequently create the diffusion skeletons. However, we focused our analysis on RD, as this is a more specific measure of diffusivity in white matter than FA or mean diffusivity. Indeed, although several factors can contribute to producing particular RD values, including the number of axons and axon packing and diameter, RD has been most consistently related to myelin content along axons, with increased RD values reflecting higher demyelination (Song et al., 2002, 2005; Klawiter et al., 2011; Zatorre et al., 2012). In animal studies, directional

measures such as RD, unlike summary parameters such as mean diffusivity or FA, provide better structural details of the state of the axons and myelin (Aung et al., 2013).

Voxel-based analyses of RD maps were performed with TBSS (Smith et al., 2006). Participants' FA maps (necessary to calculate the registrations to MNI standard space and create the RD skeletons) were registered to the FMRIB58_FA template (MNI152 space and 1 × 1 × 1 mm3) using the nonlinear registration tool (Andersson et al., 2007). These registered FA maps were first averaged to create a mean FA volume. A mean FA skeleton was then produced, representing the centers of all white matter tracts common to all participants in the study. Each participant's aligned FA data were then projected onto this skeleton by searching for the highest FA value within a search space perpendicular to each voxel of the mean skeleton. This process was repeated for the RD maps by applying the transformations previously calculated with the FA maps. This resulted in individual RD skeletons for each participant. Finally, to assess white matter differences between CB and sighted participants, independent-samples t-tests were performed on the RD skeleton. Significant results are reported at FWE-corrected $p < 0.05$ using threshold-free cluster enhancement (Smith and Nichols, 2009) and a nonparametric permutation test with 5,000 permutations (Nichols and Holmes, 2002). Significant cluster results were averaged and a mean value per participant, reflecting individual microstructural differences, was obtained.

For the mapping between RD values and the standard probabilistic atlases of white matter pathways, all voxels that differed significantly in RD values between the CB and sighted were included. We report only tracts that showed an overlap with these voxels, with the tracts from the probabilistic atlas thresholded at 0.95 probability.

Spearman correlations were used to analyze the correlation between RD values (across groups) and the spectral profiles of brain areas in the occipital cortex. For these areas, peak power values were retrieved for all clusters (i.e., lines showing clear power peak) of the complex spectral profile for each subject, and correlated with the RD values. Multiple comparisons were controlled for by using FDR ($Q = 0.05$).

3.4 Results

Spectral Profiles Replicate

In our sample of sighted adults with open eyes, we successfully replicated the classification of individual brain regions by their spectral profiles, as reported by Keitel and Gross (2016). Mean classification performance was high across all 115 ROIs, indicated by the classification ranks (Fig. 3.2A). Classification ranks refer to the probability of a region being identified by the classifier: For example, a mean rank of 1 indicates that a region was correctly assigned (i.e., highest probability

among all areas) on every iteration; a mean rank of 2 means that the assignment was correct in many but not all of the iterations (i.e., had the second highest probability among all areas). Here, the mean rank across all iterations and brain areas was 2.70 (range of ranks: 1–12.7, Keitel mean rank = 1.8), or 2.32 when considering identification of homologue (left/right hemisphere) areas (Keitel homologue mean rank = 1.4). Mean ranks of all ROIs are depicted in Fig. 3.2A. We statistically quantified the classification performance using permutation tests. The mean classification rank of an area (e.g., right calcarine) was tested against a distribution of classification ranks of all brain areas (except the current one, e.g., right calcarine) accumulated across all iterations (N = 1000). This analysis revealed that for 97% of all areas classification was significantly better when identifying themselves compared to all other regions.

Furthermore, while the average optimal number of clusters per anatomical area was lower in our sample (3.4 +/– 2.3 clusters per area) compared to Keitel and Gross (2016)(4.1 +/– 1.86 (M + SD)), the clustering approach revealed comparable spectral profiles between the studies (see Fig. 3.4 and Supplementary Fig. 3.1, Supplementary Data). Interestingly, for deeper subcortical brain structures (e.g., thalamic and limbic areas), the clusters in our data were less characteristic (i.e., only a few clusters per area with less specific shapes and high classification ranks; see Supplementary Fig. 3.1), possibly due limitations in the signal-to-noise ratio of the MEG system used.

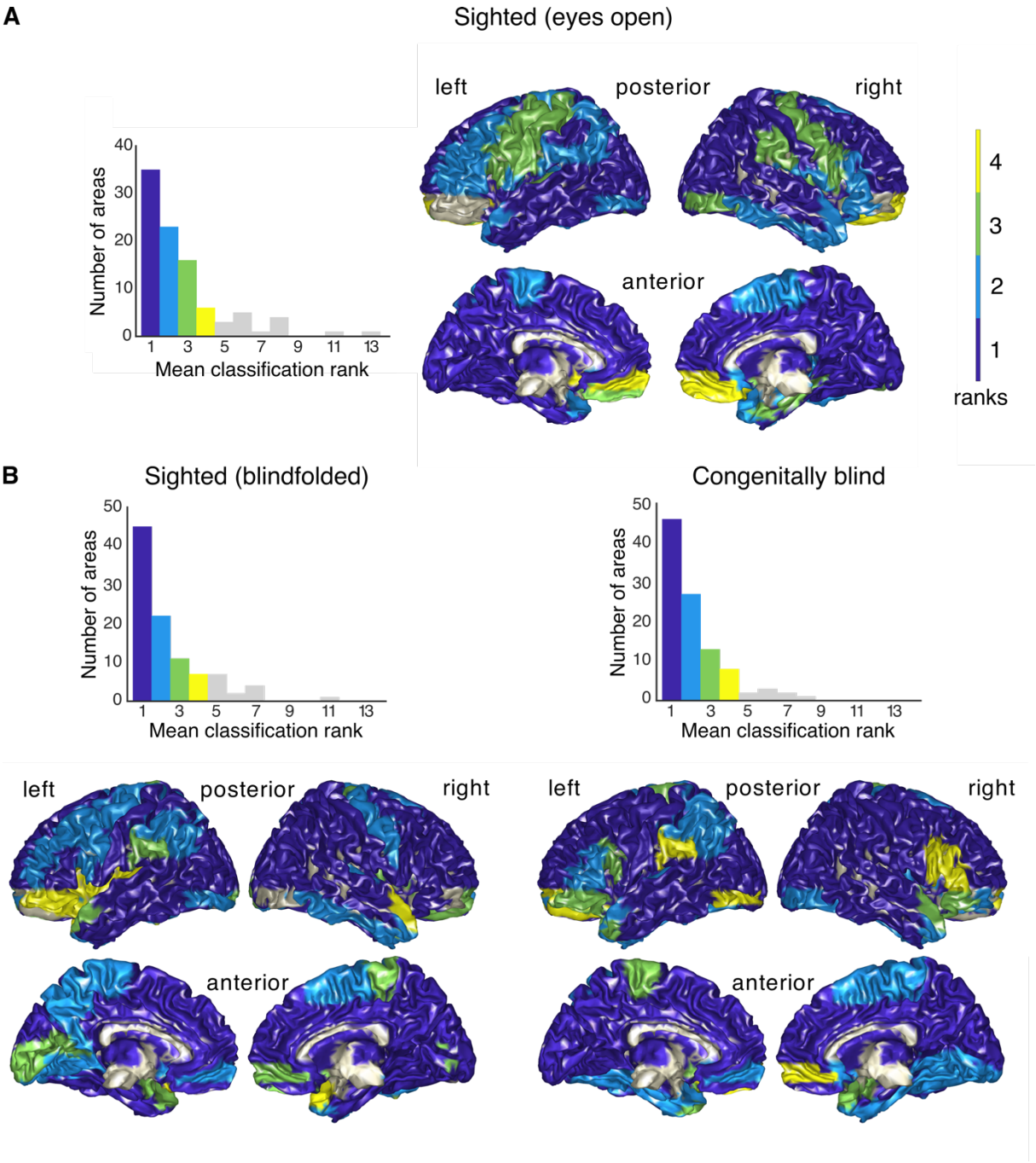


Figure 3.2. Classification results for all experimental groups. A. Sighted group with open eyes (replication sample). Left. Histogram of mean ranks in classification across all 115 brain areas. 84.5% of the brain regions obtain a mean rank between 1 and 4, while 15.5% of regions were assigned ranks up to 13. (in Keitel and Gross (2016) all mean ranks were between 1 and 4). Right. Topography of mean ranks (colors match ranks from the histogram). **B.** Sighted group with eyes closed (left) and group of congenitally blind individuals (right). For both groups, the histogram of mean ranks in classification across all 115 brain areas is displayed in the upper panel and the topography of mean ranks (colors match ranks from histogram) is displayed in the lower panel. Bin width is 1 for all subplots.

Good Classification Within Sighted and Congenitally Blind Groups

Our second hypothesis posited that, among congenitally blind individuals, anatomical areas exhibit specific spectral profiles, although possibly altered compared to those of sighted individuals. We conducted the classification procedure for the CB group and observed good classification ranks, similar to those seen in the S-EO group (Fig. 3.2B; mean rank = 2.51, range = 1–10.3, homologue mean rank = 2.10, percent significant ROIs = 100%). This indicates that spectral profiles across brain areas were consistent in the congenitally blind participants. Likewise, when applied to the data of the S-BF group, this analysis yielded similarly good classification ranks (S-BF: mean rank = 2.64, range = 1–11.4, homologue mean rank = 2.17, percent significant ROIs = 98%) compared to both the S-EO and CB. Ensuring robust within-group classification in both the CB and the S-BF was an important prerequisite for consecutive between-group analyses, as it aimed to mitigate the risk of group differences stemming from large within-group variation. Furthermore, the findings revealed a comparable distribution of mean ranks across the cortical surface for both groups, with the majority of brain regions (~70 out of 115) identified as either best or second-best.

Spectral Changes in Sensory and Right Frontal Regions in Congenital Blindness

Drawing from the literature on intra- and cross-modal plasticity in the CB, we hypothesized that spectral properties would vary between the congenitally blind and normally sighted individuals. To investigate the differences in spectral properties across brain regions between these two groups, we implemented a cross-group classification, drawing samples from the S-BF for training and samples from the CB for testing. This analysis resulted in a mean rank of 5.3 (range = 1.09 - 27.17) (Fig. 3.3A).

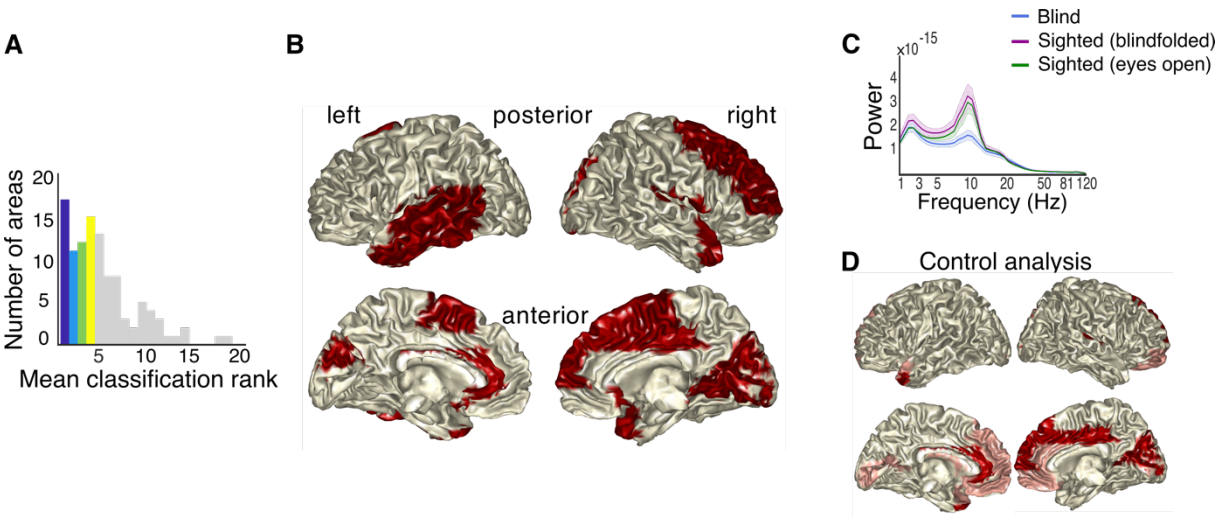


Figure 3.3. Cross-group classification results. **A.** Histogram of classification ranks. Bin width is 1. **B.** Topographic distribution of significantly different classification ranks in the cross-group classification, highlighted in red, as determined by a permutation procedure. **C.** The normalization spectrum used in the normalization procedure (i.e., the average power spectrum across segments and voxels), averaged across participants for each experimental group. Shaded error bars reflect the standard error of the mean across participants. Frequency is displayed on the x-axis, scaled logarithmically from 1 to 120 Hz; power on the y-axis. For visualization, power was interpolated at 50 Hz to remove line noise. **D.** A control analysis on the non-normalized data was performed to assess the impact of the normalization procedure on the cross-group classification findings (lower panel). Confirming the main analysis, significant findings were observed in various areas overlapping with the main analysis (highlighted in dark red). However, we also observed additional areas with significant group differences (highlighted in light pink), and some areas that, in contrast to the main analysis, did not show significant group differences in the control analysis (see Supplementary Table 3.2 for a full list). Only brain areas showing group differences in both analyses (i.e., with and without normalization procedure) will be considered for discussion.

In the cross-group analysis, 54.8% of 115 areas obtained classification ranks between 1 and 4, while the automatic identification of the remaining regions was less precise (see Fig. 3.3A). Beyond these descriptive procedures, we conducted statistical comparisons between the cross-group classification and the within S-BF group classification. As depicted in Fig. 3.3B, cross-group classification ranks were significantly worse for several sensory, as well as for right frontal areas and an extended network of left temporal brain regions. This suggests that spectral profiles in these brain regions were different between the CB group and the sighted group (see Table 3.1 for all ROIs showing significantly worse classification compared to the null distribution). A control analysis was performed to ensure that group differences were not introduced by the normalization procedure (i.e., reflecting differences in baseline power, Fig. 3.3C, D). While this analysis largely corroborates our findings, some of the brain areas did not show significant group differences in the control analysis, and several additional areas with significantly worse classification ranks were identified (see Supplementary Table 3.2 for a full list). The differences between the analyses partly reflect the advantages of the normalization procedure, which down-weights the neural activity (1/f power) predominant across the brain and not specific to a given brain area (Supplementary Fig. 3.4). In the following, we focus on interpreting findings at ROIs that exhibited significant group differences in both cross-group classification analyses (i.e., with and without normalization procedure). These areas comprise: right Heschl's gyrus, Calcarine gyrus, Cuneus, superior medial frontal gyrus, middle cingulate cortex and left middle temporal pole and anterior cingulate cortex (highlighted in Table 3.1).

Table 3.1. Table of all brain areas (out of 115) with significant classification differences. Brain areas where classification ranks were significantly different between the congenitally blind and the sighted (blindfolded) are listed. The areas highlighted in bold showed significant group differences in the cross-group classification analyses (i.e., with normalization procedure) and in the control analysis (i.e., without normalization procedure; see Supplementary Fig. 3.4) and will be the focus of interpretation.

Coarse region	Hemisphere	
	Left	Right
Frontal		Superior frontal gyrus Middle frontal gyrus **Superior medial frontal** Rolandic Operculum
Visual	Cuneus	**Calcarine gyrus** **Cuneus** Superior occipital gyrus Lingual gyrus
Temporal	Superior temporal gyrus Middle temporal gyrus **Middle temporal pole** Inferior temporal gyrus	**Heschl's gyrus** Middle temporal pole Superior temporal pole
Sensorimotor	Supplementary Motor Area	Supplementary Motor Area
Non-cortical	Olfactory **Anterior cingulate cortex** Cerebellum lobule VI Cerebellar vermis 3 Cerebellar vermis 9	Olfactory **Middle cingulate cortex** Cerebellum lobule X Cerebellar vermis 3 Cerebellar vermis 9

Interestingly, the brain areas showing group differences in the spectral profiles in the cross-group classification were characterized by clusters with peaks showing increased power at higher frequencies in the CB as compared to the S-BF participants (for a representative selection of or all brain areas with significant effects, see Fig. 3.4A and Supplementary Fig. 3.1, respectively). This pattern was evident for the auditory regions, with more power in the alpha and beta band in the CB compared to the sighted participants, as well as in right frontal areas, which showed higher power in the theta band. In visual brain areas, power peaks were reduced in the alpha band for the CB compared to the S-BF participants, whereas power was increased in the low-gamma band. Post-hoc permutation tests were performed to validate these findings. To assess differences in spectral signatures between the S-BF and CB, raw Fourier spectra (i.e., without applying the spectral clustering) were extracted und subjected to permutation statistics. Participants' group assignment (S-BF vs. CB) was randomly permuted (1000 permutations) ($Q = 0.05$; false discovery rate [FDR] corrected p-value = .033; p-values < .033) (see Fig. 3.4B and Supplementary Fig. 3.2). The differences in low-gamma power in Calcarine gyrus between the CB and the S-BF were not significant in the post-hoc tests (see Supplementary Fig. 3.5).

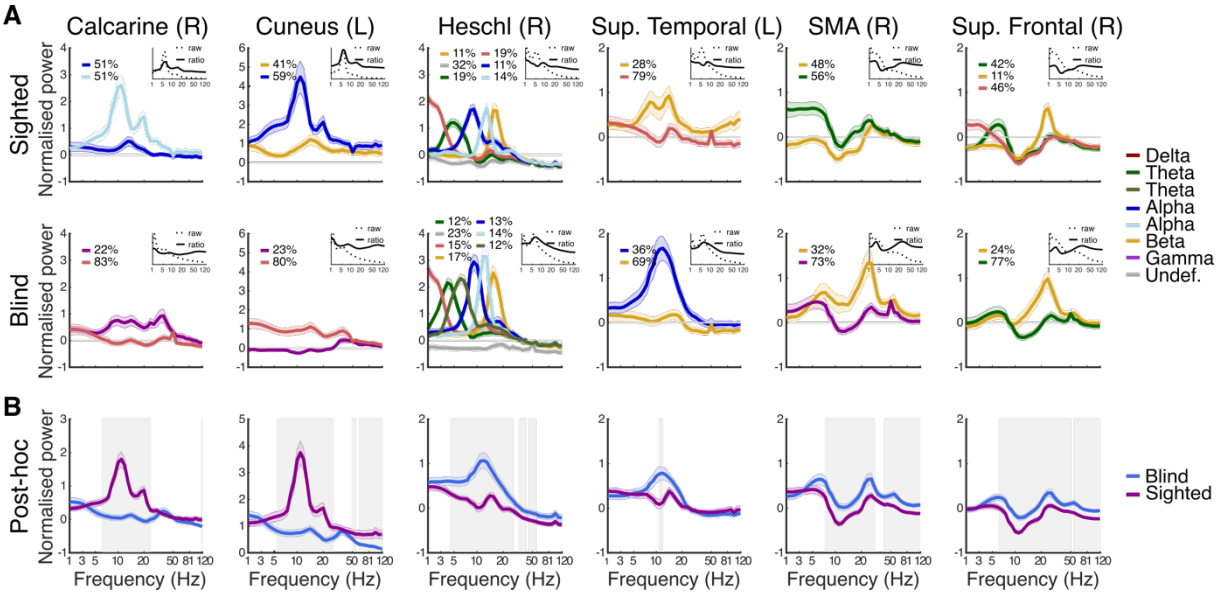


Figure 3.4. Spectral profiles differ between sighted and blind participants. A. Selection of spectral profiles of brain areas (columns) showing significantly different cross-group classification, separately for the sighted (S-BF) and congenitally blind (CB) participants (rows). Clusters are color-coded by the frequency of the peak amplitude (legend on the right). The insets display normalized (solid lines) and unnormalized mean power spectra (dashed lines) (i.e., both without applying the spectral clustering). **B.** Post-hoc analyses of spectral differences for selected brain regions (columns). Spectra represent normalized mean power (i.e., without applying the spectral clustering). Frequencies with group differences in power are indicated by grey boxes (permutation test: Q = 0.05; FDR corrected p-value = .033; p-values < .033). The groups are color-coded (legend on the right). In all panels: Shaded lines represent standard error of the mean.

Spectral Changes Correlate with Structural Group Differences

We examined the relation between brain structure and the spectral differences in occipital brain areas observed between the groups in the classifier analyses using diffusion-tensor imaging (DTI) in a subsample of participants. The TBSS analysis revealed significantly higher Radial Diffusivity (RD) values in a bi-hemispheric spatial cluster in posterior brain regions (i.e., a cluster of voxels; Supplementary Fig. 3.3B) for the CB group compared to the sighted participants, indicating reduced white matter structural integrity in the CB group ($N_{sighted}$ = 12, N_{CB} = 16; family-wise error (FWE)-corrected at the peak voxel, two-sided p = 0.05; see Supplementary Fig. 3.3B). We used a probabilistic atlas of white matter pathways in MNI space (Thiebaut de Schotten et al., 2011) to evaluate the overlap of the spatial cluster with known white matter tracts. With the probabilistic atlas thresholded at 0.95, the TBSS cluster overlapped significantly with the posterior corpus callosum, the posterior inferior longitudinal fasciculus (bilaterally), the posterior inferior fronto-occipital fasciculus (bilaterally), and the optic radiations (also bilaterally). This indicates a 95% probability that the reduced white matter properties identified in the CB group primarily affect these white matter tracts.

Finally, RD values belonging to this spatial cluster were extracted and correlated with the power of individual cluster peaks from the spectral profile of occipital brain areas (Spearman correlations, FDR correction, Q = 0.05, corrected p-value = .0065, Supplementary Fig. 3.3C). This analysis revealed a negative correlation between RD and alpha power in both right (rho = –0.55, p = .005) and left (rho = –0.81, p > 0.001) calcarine, as well as left cuneus (rho = –0.68, p > .001). The findings suggest that reduced white matter properties in occipital areas were accompanied by reduced neural alpha power. Alpha and gamma power of the remaining occipital regions did not correlate with the RD values (all ps > .0065; cf. Supplementary Table 3.1).

3.5 Discussion

Congenitally blind individuals often exhibit enhanced behavioral performance in tasks involving the intact senses (Lessard et al., 1998; Roeder et al., 2000, 2003, 2004; Gougoux et al., 2004, 2005; Hoetting et al., 2004; Hertrich et al., 2009; Dietrich et al., 2013), alongside structural and functional reorganization of brain areas associated with the intact sensory systems (Pascual-Leone and Torres, 1993; Sterr et al., 1998a, 1998b; Roeder et al., 1999; Elbert et al., 2002; Stevens and Weaver, 2005; Gougoux et al., 2009; Schepers et al., 2012; Hertrich et al., 2013; Watkins et al., 2013), as well as in brain regions predominantly associated with the visual system (Elbert et al., 2002; Burton, 2003; Roeder and Neville, 2003; Noppeney et al., 2005; Pascual-Leone et al., 2005; Noppeney, 2007; Bedny et al., 2011; Voss and Zatorre, 2012). However, the impact of these behavioral and neural adaptations on brain rhythms beyond occipital brain regions has remained unclear.

Here, we investigated spectral alterations in congenitally blind individuals to deepen our mechanistic understanding of both blindness-related adaptations and the functional relevance of brain rhythms. Confirming our first two hypotheses, spectral clustering and classification procedures performed exceptionally well across all groups (97–100% of areas were classified correctly in each group), highlighting consistent brain area-specific spectral properties within both sighted and, demonstrated for the first time, congenitally blind individuals. Crucially, cross-classification highlighted changes in spectral profiles in sensory (auditory and visual) and right-frontal brain areas, left middle temporal pole and bilateral parts of the cingulate cortex, associated with visual deprivation. This was reflected in increased power in the theta-to-beta frequency bands in the right primary auditory cortex and right-frontal brain regions. While the sighted displayed a distinct alpha rhythm commonly observed during rest in visual cortex, the congenitally blind individuals showed reduced alpha power alongside a gamma (~40 Hz) peak. Additionally, a correlation analysis of individual spectral clusters in occipital cortex and microstructural white matter properties revealed a positive

correlation between alpha (but not gamma) power and white matter integrity in an extended bihemispheric spatial cluster in posterior parts of the brain.

Robust Classification of Brain Areas Based on Spectral Profiles

Spectral profiles measured during resting state have revealed intrinsic brain rhythms that are believed to be involved in task-related processing (Smith et al., 2009; Hipp et al., 2012; Engel et al., 2013; Raichle, 2015; Keitel and Gross, 2016; Sormaz et al., 2018). Here, we replicated the observation that brain areas express characteristic spectral profiles, with similar spectral profiles, classification ranks, and distributions of classification ranks across the cortex in the sighted (S-EO) group as reported by Keitel and Gross (2016). Occipital regions exhibited the typical alpha rhythm (~10 Hz), often associated with attentional inhibition (Klimesch et al., 2007; Buffalo et al., 2011; Jensen et al., 2012). Moreover, spectral peaks in the beta band (~20 Hz) were prominent across frontal and central brain areas, aligning with previously reported natural frequencies of these regions (Rosanova et al., 2009; Ferrarelli et al., 2012; Keitel and Gross, 2016), which may contribute to temporal motor processing (Fujioka et al., 2012; Arnal et al., 2015; Morillon and Baillet, 2017).

While the spectral profiles of most brain areas resembled those reported by Keitel and Gross (2016), some differences were noted in our study (see Supplementary Fig. 3.1). Notably, deep subcortical brain areas, in contrast to Keitel and Gross (2016), were not classified well. This discrepancy may arise from the slightly shorter segment duration used here, reducing frequency resolution at lower frequencies and potentially affecting the detection of low-frequency clusters. Alternatively, lower signal-to-noise ratio in deeper brain areas in our study, possibly due to different MEG systems, could account for this variation. Thus, it is conceivable that both the recording system used and/or the sample of participants had a differing impact on the specific profiles of certain brain areas. A comprehensive examination using a large dataset from various recording sites, comprising hundreds of recordings, will be necessary to elucidate which spectral modes generalize across individuals in a larger population. Importantly, within our sample, the spectral profiles were consistent across individuals, with reported group-level clusters comprising contributions from at least ~70% of participants, and on average ~97% for the S-EO and ~94% for the S-BF group. In summary, our results underscore the robustness of brain area-specific spectral profiles, suggesting that these profiles represent characteristic properties reflecting the intrinsic brain rhythms of cortical regions.

On the premise that changes in spectral profiles are functionally related to adaptive plasticity in congenitally blind individuals, consistence rather than randomness is expected within this group. A novel discovery of our study that aligns with this hypothesis is the reliable identification of brain regions based on their spectral clusters in both congenitally blind and sighted groups, suggesting spectral consistencies across individuals (Fig. 3.2B, right). Specifically, only group clusters were

reported where at least ~69% of participants, and on average ~95 % contributed to each spectral cluster. This result implies that cortical adaptation to visual deprivation leads to specifically and homogenously altered spectral profiles in congenitally blind individuals.

Selective Spectral Plasticity Across the Brain

Our study aimed to create a comprehensive (in space and frequency), data-driven model of spectral profiles in congenitally blind compared to sighted individuals, shedding light on the neural mechanisms underlying congenital blindness-related plasticity. Given the cross-group classification, the participant groups were well matched (cf. methods section) to isolate effects related to visual deprivation. While cross-group classification performed well for many brain areas (i.e., low ranks; Fig. 3.3A), spectra related to sensory (auditory and visual), right-frontal, left middle temporal pole and bilateral parts of the cingulate cortex were classified notably less accurately compared to within-sighted classification (Table 3.1; Fig. 3.3B). These findings suggest systematic alterations in spectral properties due to visual deprivation-related plasticity in certain brain areas. Previous fMRI research has highlighted a non-monotonic relationship between plasticity and stability across cortex (Haak and Beckmann, 2019), with plasticity observed in neural activity being related to gene expressions (Ortiz-Terán et al. 2017). Specifically, plasticity likely decreases from early visual to mid-level cortex, but increases again further in the visual cortical hierarchy (Haak and Beckmann, 2019).

Spectral Plasticity in Sensory Areas

Altered spectral profiles in the congenitally blind individuals reflect changes in intrinsic brain rhythms, potentially linked to cognitive processes. Our findings underscore changes in spectral properties of the auditory and visual cortices due to neuroplasticity resulting from visual deprivation. These findings confirm previous results of both cross-modal reorganization in visual cortex (Burton, 2003; Pascual-Leone et al., 2005; Bedny et al., 2011; Voss and Zatorre, 2012; Gudi-Mindermann et al., 2018; Rimmele et al., 2019) and intra-modal reorganization in auditory cortex (Roeder and Neville, 2003) in blind humans. Importantly, our findings extend these results by offering evidence for genuine changes in the processing mode of these regions, as indicated by changes in their spectral characteristics, thereby contributing to a more mechanistic understanding.

Visual brain areas classified as spectrally different between the sighted and congenitally blind individuals, both in the normalized and non-normalized control analysis, included the right primary visual cortex (calcarine sulcus) and cuneus (Table 3.1). In the sighted, these areas exhibited two alpha-band clusters indicating distinct alpha brain rhythms. The posterior alpha rhythm, typically observed in visual areas during resting state, likely reflects rhythmic inhibition of task-irrelevant neural circuits (Haegens et al., 2011). Our data revealed one cluster with a clear visual alpha peak around 10 Hz and a smaller peak in the beta band (~20 Hz). A second alpha cluster was

characterized by a smaller amplitude, corroborating previous work suggesting distinct visual alpha rhythms (Barzegaran et al., 2017). In contrast, in the congenitally blind individuals, different spectral patterns emerged in these typical visual areas. One cluster displayed a peak at higher frequencies (low gamma, ~40 Hz) with a strongly reduced alpha power, while a second cluster showed low power in the alpha band (Fig. 3.4 A, B). This observation aligns with previous findings reporting reduced or absent alpha rhythms in the visual system of blind individuals (Adrian and Matthews, 1934; Noebels et al., 1978; Kriegseis et al., 2006; Hawellek et al., 2013). The diminished alpha brain rhythm is believed to reflect atrophy or reorganization of the cortico-thalamic and cortical pathways, where posterior alpha generators are located (Lopes da Silva, 1991; Lőrincz et al., 2009). Synchronized posterior gamma activity may be modulated by alpha (de-)synchronization, indicating increased neural gain for feedforward sensory processing (van Kerkoerle et al., 2014; Michalareas et al., 2016). The gamma rhythm in sighted individuals typically increases in response to a visual stimulus and decreases during rest. In blind individuals, the gamma rhythm may be preserved, emerging locally in the occipital cortex (Bastos et al., 2014; Marshall et al., 2018) and being less impacted by atrophy of the thalamo-cortical pathway. The presence of gamma activity in blind individuals during rest may reflect altered alpha-gamma excitation-inhibition balance, with the reduced alpha rhythm potentially leading to a disinhibition of visual cortex, thereby increasing gamma band activity (Ossandón et al., 2023). Functionally, the reduced alpha rhythm in the congenitally blind may reflect cortical and thalamo-cortical atrophy, while the gamma-band rhythm could be associated with the compensatory reorganization of visual cortex, such as its recruitment during non-visual tasks (Voss and Zatorre, 2012; Striem-Amit et al., 2015; Bedny, 2017; Voss, 2019). This conjecture is supported by our correlation analysis of spectral profiles with microstructural white matter properties (Supplementary Fig. 3.3, Supplementary Table 3.1), revealing that only the alpha cluster correlated with the deteriorated white matter properties.

Additionally, we identified altered spectral profiles in right auditory cortex (Heschl's gyrus; Table 3.1) characterized by increased power in specific frequency bands. Specifically, blind individuals exhibited increased power in the theta-to-beta frequencies compared to sighted participants (Fig. 3.4A, 4B). Importantly, Keitel and Gross (2016) provided evidence for the functional relevance of intrinsic delta, theta and beta brain rhythms observed during rest for speech processing in auditory cortex. Speech tracking in the theta band, where auditory cortex oscillatory activity phase-locks to speech acoustics, is tightly linked to syllabic segmentation and predicting upcoming stimuli (Giraud and Poeppel, 2012; Gross et al., 2013; Haegens and Zion Golumbic, 2018; Rimmele et al., 2018), as well as for segmentation of sound and music (Doelling and Poeppel, 2015; Morillon and Baillet, 2017). Interestingly, congenitally blind individuals can comprehend ultra-fast speech at syllabic rates beyond the ability of sighted individuals (i.e., syllablic rates faster than the theta range; Brungart et al., 2007; Ghitza and Greenberg, 2009), which has been linked to accelerated theta

brain rhythms in right auditory cortex (see also: Trouvain, 2007; Hertrich et al., 2009, 2013; Dietrich et al., 2013; Van Ackeren et al., 2018). This suggests that the altered spectral profiles in congenitally blind individuals reflect intrinsic brain rhythms that enhance auditory temporal segmentation skills. However, further research is needed to confirm this hypothesis by investigating the relationship between changes in brain rhythms and behavioral performance.

Spectral Plasticity Beyond Sensory Cortices

In addition to spectral reorganization in sensory cortices in congenital blindness, our data suggest adaptive changes in right-hemispheric frontal brain regions, bilateral parts of the cingulate cortex (left anterior, and right middle) and the left middle temporal pole. Spectral clusters of right medial superior frontal gyrus and cingulate cortices were classified as different between blind and sighted individuals, with increased power in the theta-band in the blind. Medial superior frontal gyrus has been linked to theta-band brain rhythms (Başar-Eroglu et al., 1992; Kubota et al., 2001; Onton et al., 2005; Cohen et al., 2008; Cohen, 2011; Itthipuripat et al., 2013; Hsieh and Ranganath, 2014; Töllner et al., 2017; Eschmann et al., 2018) and various cognitive processes, including cognitive control (Cavanagh and Frank, 2014), cognitive control of timing (Lewis and Miall, 2003a), decision making (Schmidt et al., 2018), working memory (Itthipuripat et al., 2013; Hsieh and Ranganath, 2014), temporal order processing (Chander et al., 2016), or language processing (Binder et al., 2009). Some of these processes are likely altered in congenital blind individuals, such as language processing in a frontotemporal network (Roeder et al., 2002; Lane et al., 2017), and temporal order processing (Roeder et al., 2004; Stevens and Weaver, 2005).

The cingulate cortex has been associated with various processes including emotion, cognitive control and pain processing (Shackman et al., 2011). Similarly, the left temporal pole, part of the anterior temporal lobe, is implicated in semantic (Binder et al., 1999; Visser and Lambon Ralph, 2011) and conceptual (Baron and Osherson, 2011) processing. Altered brain responses in both the cingular (Ortiz-Terán et al., 2016) and anterior temporal (Striem-Amit et al., 2018) have been observed in congenitally blind individuals, suggesting adaptive behavior. However, due to limited understanding of their functionality in congenital blindness, particularly the role of endogenous brain rhythms, we refrain from further interpreting these findings.

It is worth noting that while we observed differences in spectral profiles for many brain areas in the cross-group classification analysis (normalized data), not all of these differences were evident in the control analysis (on the non-normalized data; Table 3.1; Fig. 3.4C, Supplementary Table 3.2). We applied a previously established procedure (Keitel and Gross, 2016). One drawback of the procedure is the potential introduction of spurious activity due to group differences in the normalization spectrum, as it normalizes the spectra across the brain. However, an advantage is that it down-weights the neural activity (1/f power) that is predominant across the brain and not specific

to particular areas, thereby simplifying the spectral analysis. Thus, without normalization, the noise-level in the data is increased due to individual differences in overall power across the spectrum and across the brain. While our rationale justifies this approach, as a precautionary measure, we only interpreted effects that were consistent across both analyses.

Spectral profiles related to microstructural white matter properties

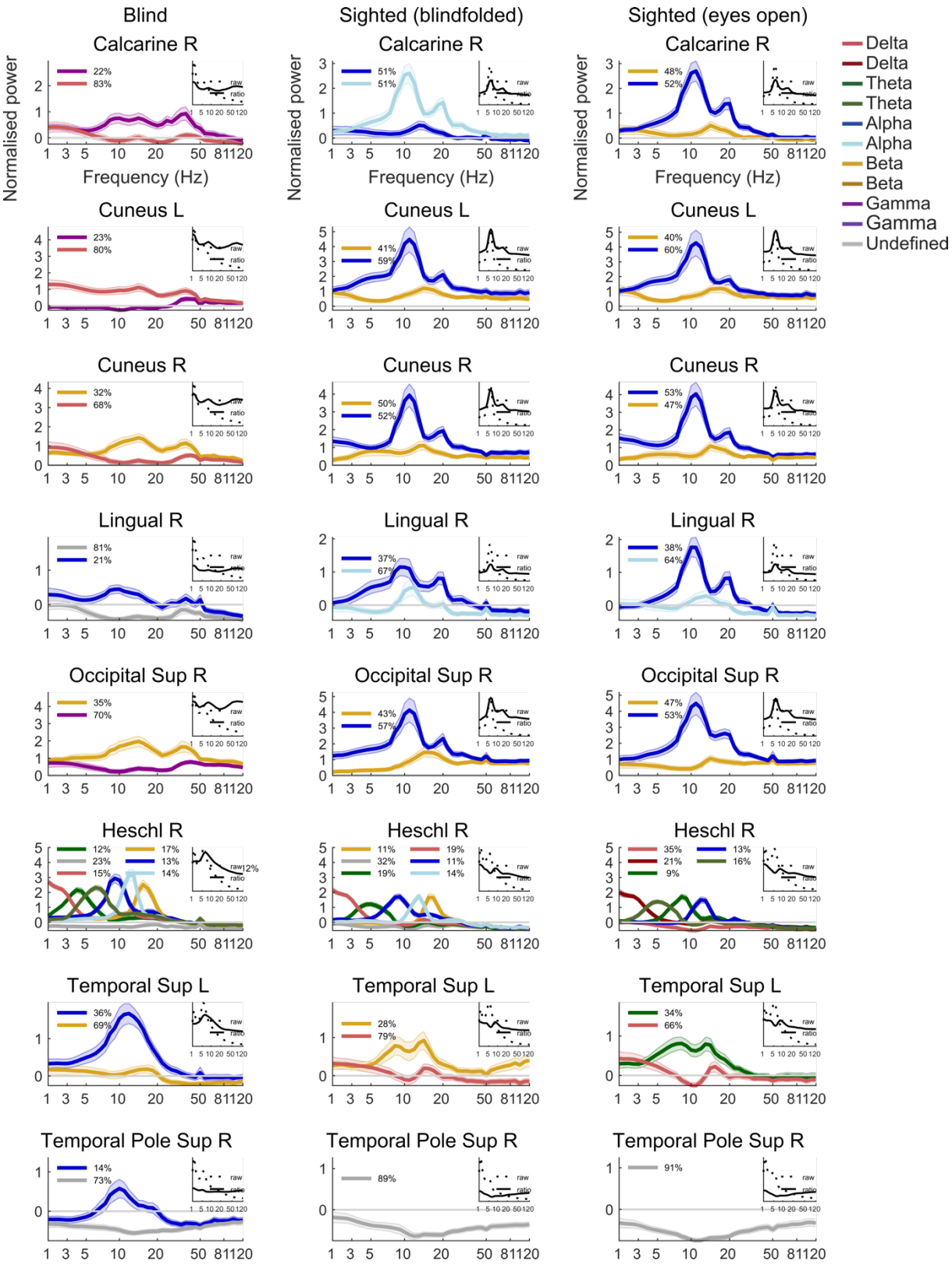
A relevant question arises regarding the association between altered spectral profiles in occipital cortex of congenitally blind individuals and structural differences. Our DTI analysis revealed compromised white matter integrity in congenitally blind individuals within visual association tracts comprising the ventral visual stream. These tracts included the bilateral inferior fronto-occipital fasciculus, connecting occipital and frontal brain areas and the bilateral inferior longitudinal fasciculus, connecting occipital and anterior temporal cortices, which have been conjectured to play a role in reading, writing and language semantics (Catani and Mesulam, 2008); and the bilateral optic radiations, linking the visual thalamus to the primary visual cortex. Additionally, compromised white matter integrity was observed in the posterior corpus callosum, facilitating homologous visual cortices (Restani and Caleo, 2016). Reduced white matter integrity is a well-documented finding in early (Shimony et al., 2006; Park et al., 2007), late (Wang et al., 2013; Hofstetter et al., 2019), and congenitally blind individuals (Ptito et al., 2008; Wang et al., 2013; Aguirre et al., 2016; Reislev et al., 2016), particularly affecting visual pathways (e.g., optic radiations and the corpus callosum). As myelination is proportional to the degree of neural activation during brain development (Demerens et al., 1996; Stevens et al., 2002; Fields, 2004; Ishibashi et al., 2006; Gautier et al., 2015; Wake et al., 2015), the lack of visual processing during development in congenital blindness likely causes the reduced white matter density in visual pathways (Hill et al., 2014; Dietz et al., 2016; Restani and Caleo, 2016; Anurova et al., 2019). Our findings show that reduced spectral power in the visual alpha rhythm coincided with reduced white matter integrity in pathways overlapping with visual cortical pathways and thalamo-cortical projections (i.e., negative correlation between alpha power and RD; Supplementary Fig. 3.3B). This confirms previous accounts suggesting that diminished visual alpha power originates from structural impairment (Adrian and Matthews, 1934; Noebels et al., 1978; Kriegseis et al., 2006; Hawellek et al., 2013).

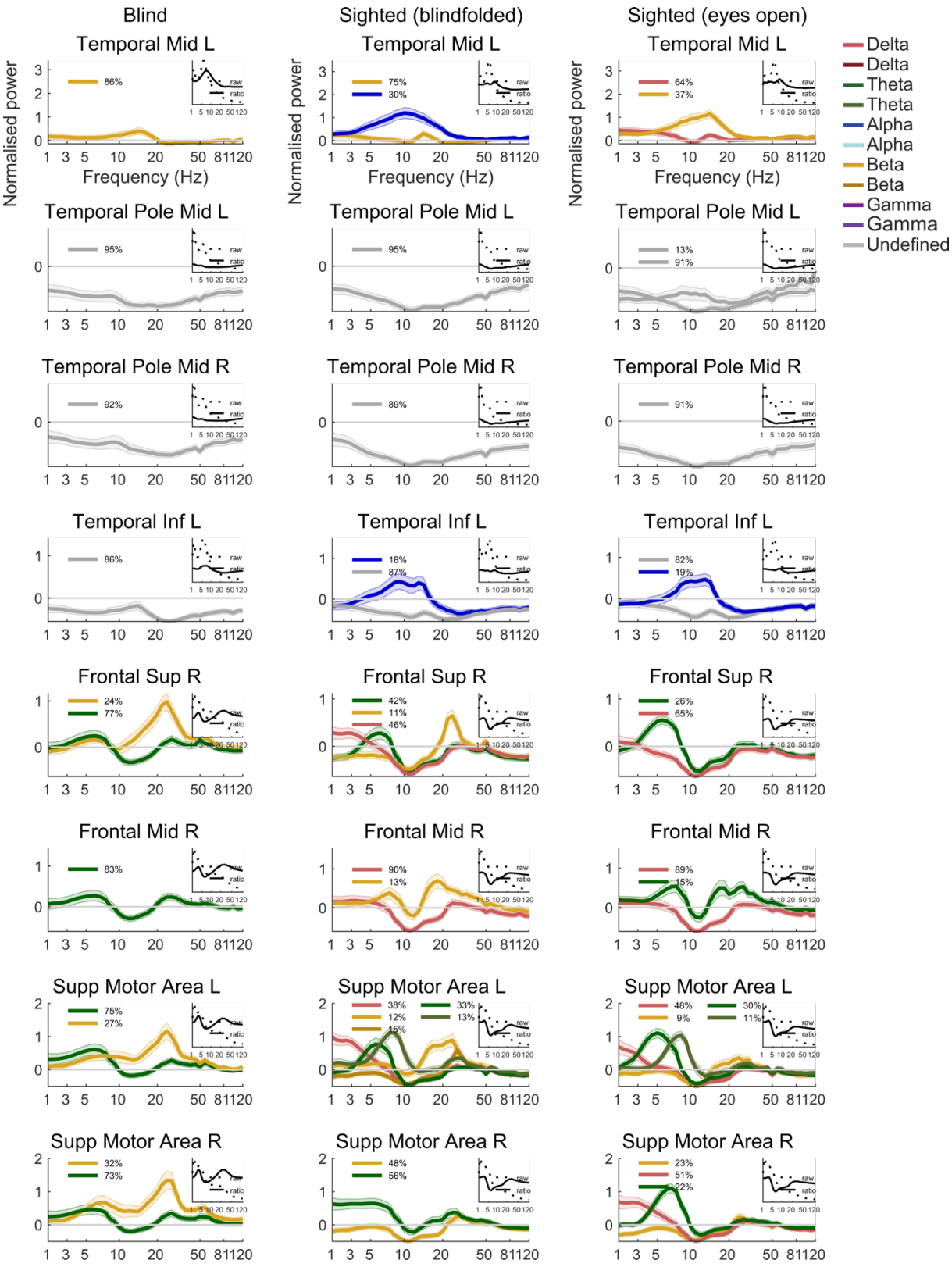
3.6 Conclusion

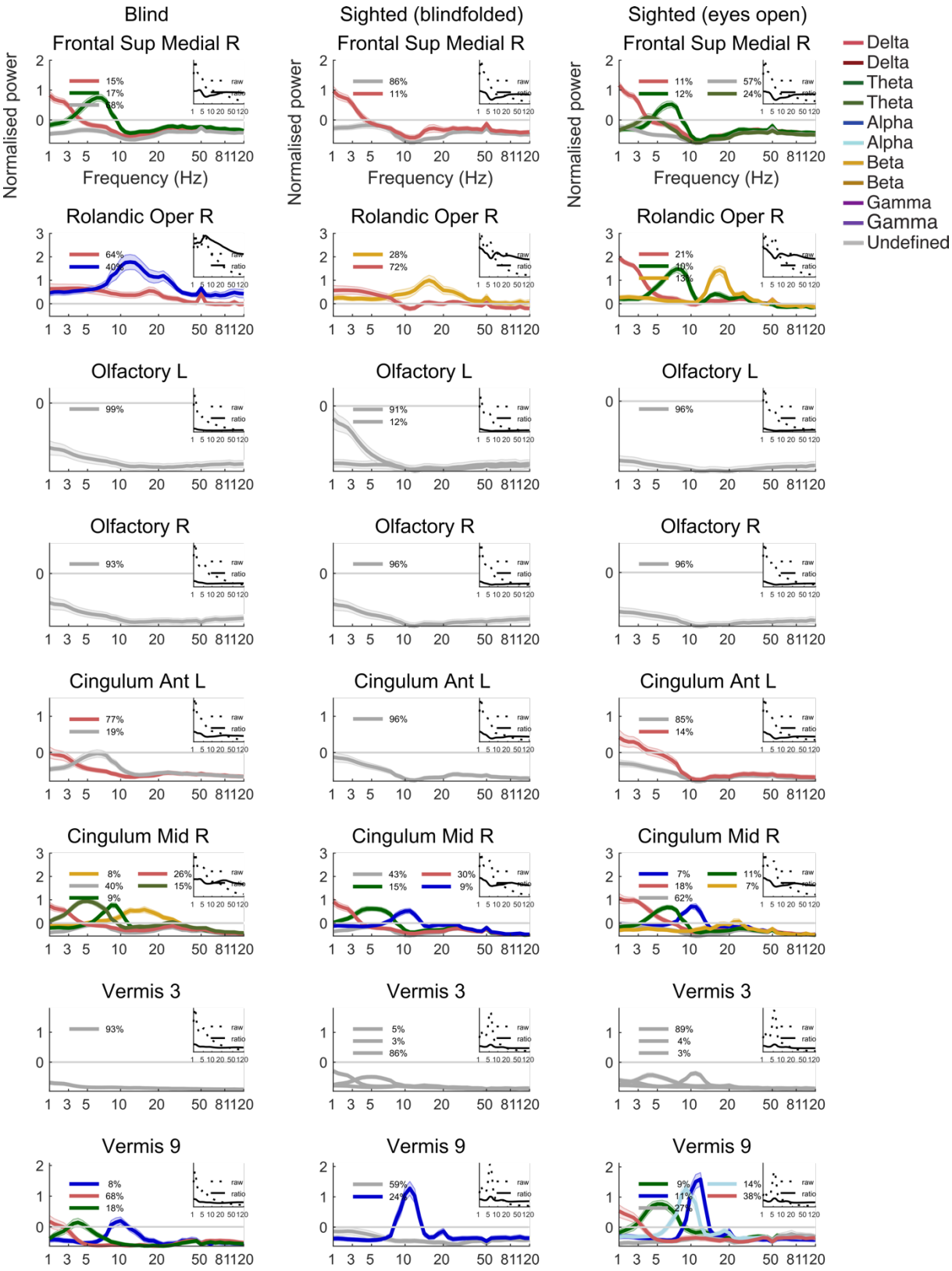
The present study reinforced the existence of robust brain area-specific spectral profiles in the human brain. For the first time, we offer a whole-brain model of spectral profiles in congenitally blind adults. Importantly, we successfully mapped spectral changes due to visual deprivation to specific brain areas — particularly in right visual, auditory, frontal, left anterior temporal, and bilateral cingulate areas. We suggest that power increases in the theta-to-beta frequency bands in

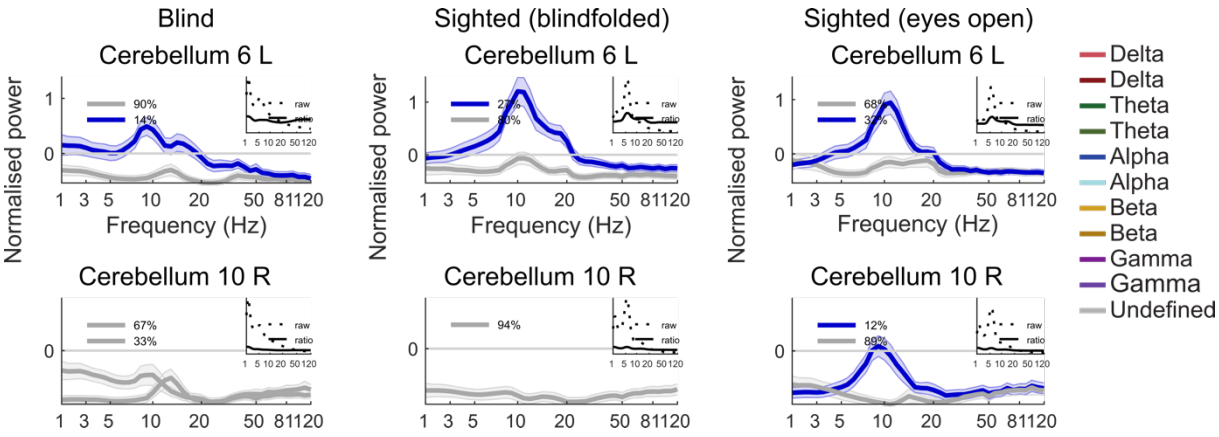
auditory and frontal brain regions may reflect adaptive sensory or higher cognitive processing in blind individuals. Conversely, altered spectral profiles in visual brain regions, characterized by lower alpha and a gamma peak, may indicate a change in the excitation-inhibition balance. While these interpretations remain preliminary, the results pave the way for more focused, task-based studies aiming at elucidating the link between altered spectral profiles and adaptive behavior.

3.7 Supplementary materials

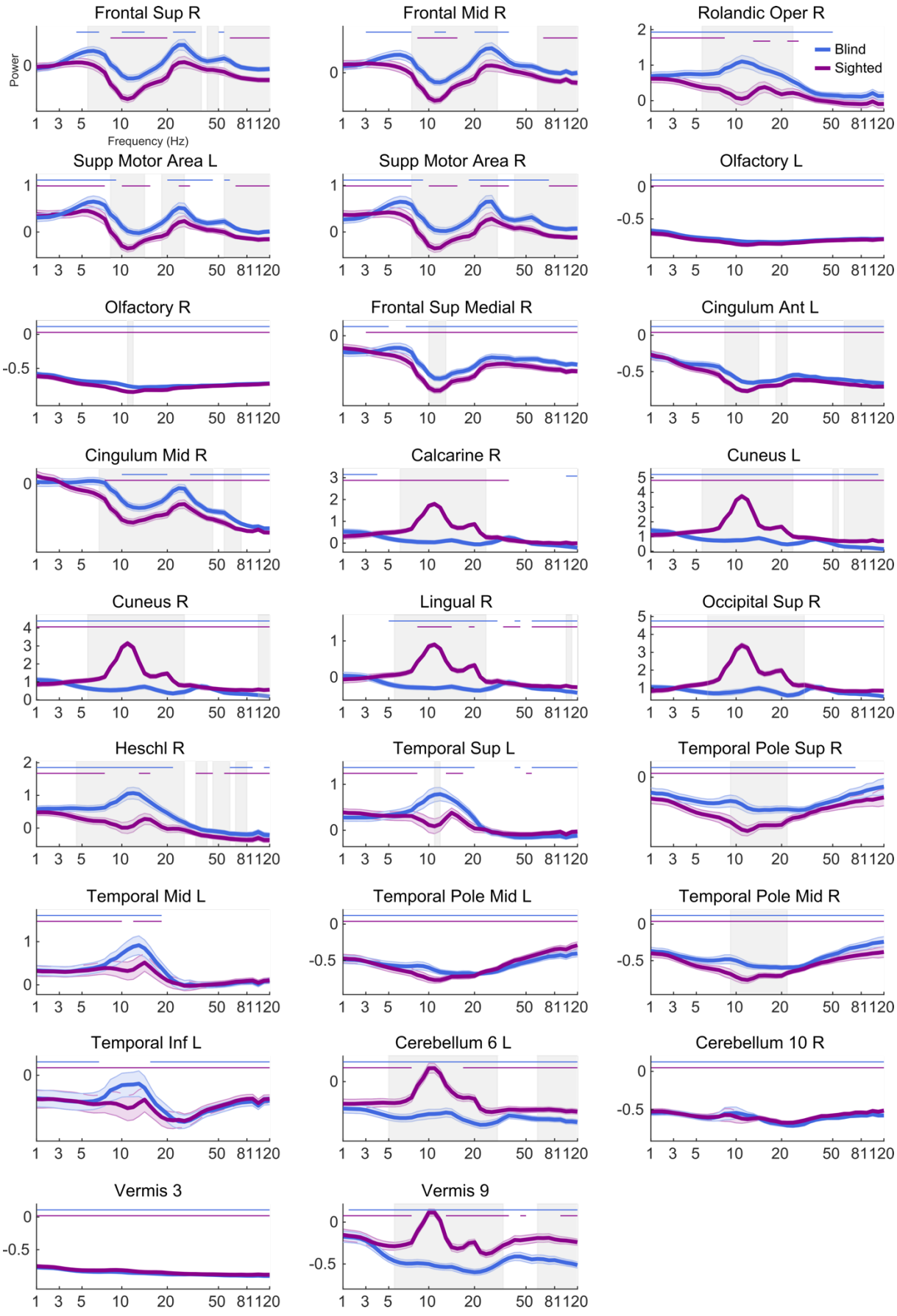



Blind
Sighted (blindfolded)
Sighted (eyes open)
Normalised power
Frequency (Hz)
Temporal Mid L
Temporal Pole Mid L
Temporal Pole Mid R
Temporal Inf L
Frontal Sup R
Frontal Mid R
Supp Motor Area L
Supp Motor Area R
raw
ratio
Delta
Delta
Theta
Theta
Alpha
Alpha
Beta
Beta
Gamma
Gamma
Undefined

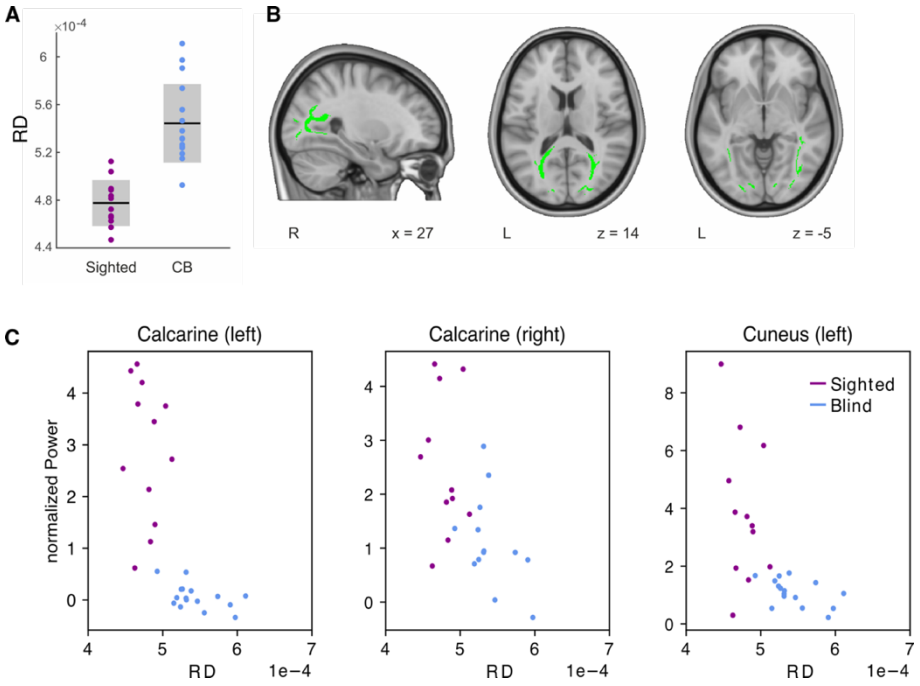

Blind
Sighted (blindfolded)
Sighted (eyes open)
Normalised power
Frequency (Hz)
Frontal Sup Medial R
Rolandic Oper R
Olfactory L
Olfactory R
Cingulum Ant L
Cingulum Mid R
Vermis 3
Vermis 9
raw
ratio
Delta
Delta
Theta
Theta
Alpha
Alpha
Beta
Beta
Gamma
Gamma
Undefined



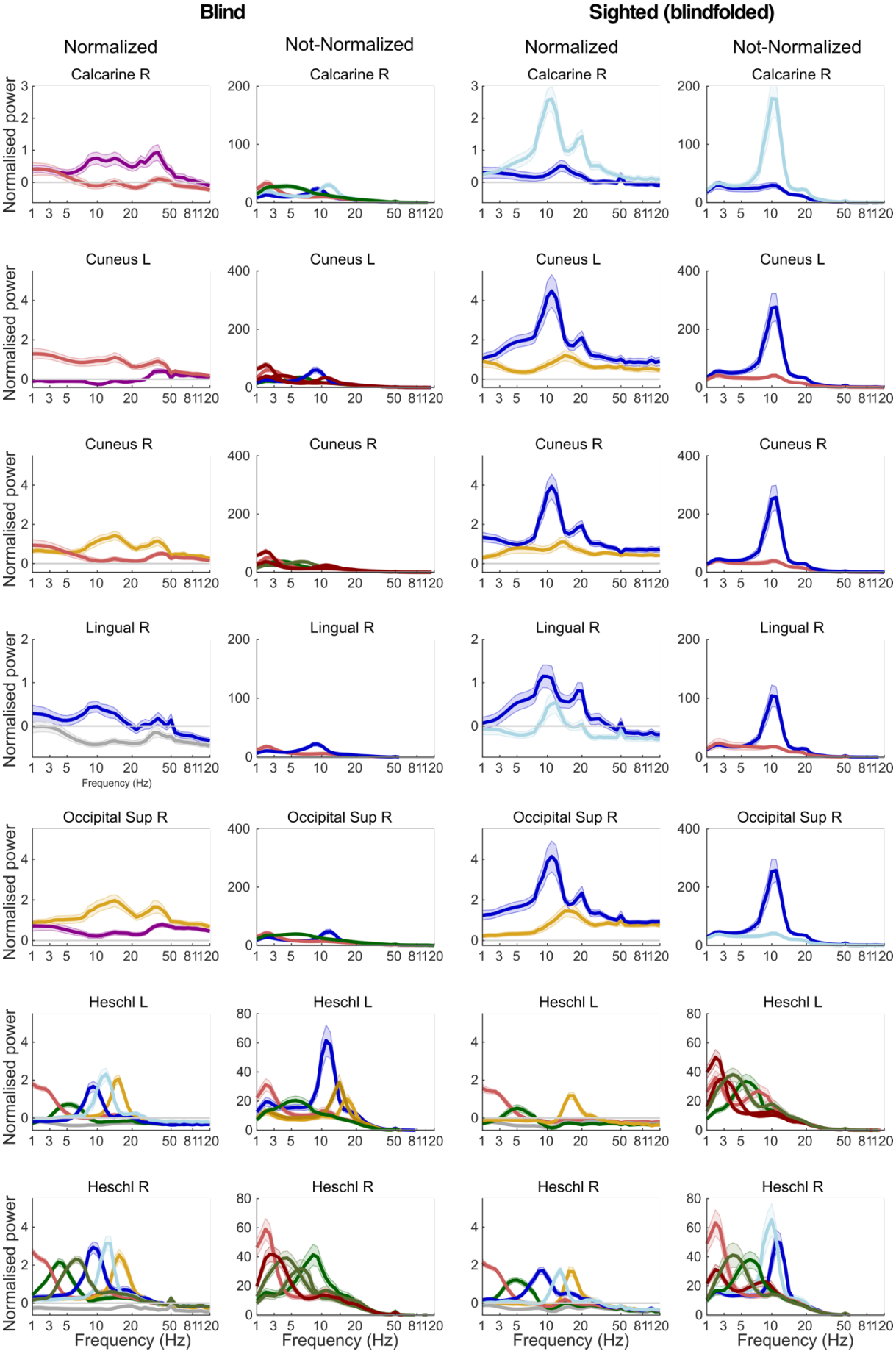
Supplementary Figure 3.1. Spectral profiles of all brain areas with significant cross-group differences. Clustered spectral profiles shown separately for brain areas (rows) and experimental groups (columns). The x-axis illustrates frequency, scaled logarithmically from 1 to 120 Hz; the y-axis illustrates normalized power. Note that some clusters show spikes at 50 Hz and 100 Hz which are due to line noise. Each line corresponds to a cluster forming part of a spectral profile. The legends illustrate cluster durations, i.e., the per- centage of single subject segments contributing to a group cluster, averaged across subjects. Clusters are color-coded according to the peak frequency (red: delta, green: theta, blue: alpha, yellow: beta, grey: no clearly-identifiable peak). Multiple clusters belonging to one frequency range are illustrated by different shades of the same color, i.e. dark and light blue. Shaded error bars represent the standard error of the mean across participants. Only group clusters to which the majority of subjects contributed are shown (i.e., CB: N=18; S-BF: N=17; S-EO: N=16). The insets display the normalized mean power spectra (solid lines) and the unnormalized mean power spectra (dashed lines) (i.e., both without applying the spectral clustering).


Frontal Sup R
Frontal Mid R
Rolandic Oper R
Blind
Sighted
Power
Frequency (Hz)
Supp Motor Area L
Supp Motor Area R
Olfactory L
Olfactory R
Frontal Sup Medial R
Cingulum Ant L
Cingulum Mid R
Calcarine R
Cuneus L
Cuneus R
Lingual R
Occipital Sup R
Heschl R
Temporal Sup L
Temporal Pole Sup R
Temporal Mid L
Temporal Pole Mid L
Temporal Pole Mid R
Temporal Inf L
Cerebellum 6 L
Cerebellum 10 R
Vermis 3
Vermis 9

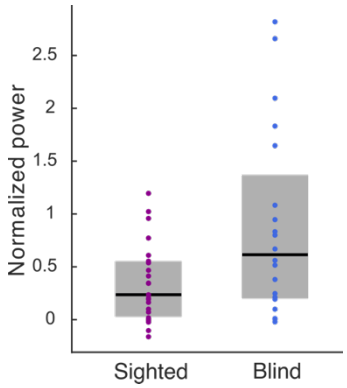
Supplementary Figure 3.2. Post-hoc analysis of spectral differences between sighted and blind participants. Raw normalized spectral profiles (i.e. no clustering) for all brain regions for which cross-group classification was significantly worse compared to classification in the sighted (eyes closed, S-EC). Each panel illustrates spectral power for the S-EC (magenta) and the congenitally blind (CB) group (blue), averaged across subjects of each group. Shaded error bars reflect standard error of the mean across subjects. Frequency areas shaded in grey indicate frequency ranges in which the groups showed significant power differences, as assessed by permutation statistics (Q = 0.05; FDR corrected p-value = .033; p-values < .033). Additionally, we statistically tested spectral power in all depicted ROIs against a baseline of zero at each frequency bin between 1-120 Hz. Due to the ratio normalization, testing power peaks against "zero" here indicates whether at a certain ROI the power is greater compared to the average power across the brain at this frequency. Significant frequencies are illustrated by horizontal lines separately for the groups. The figure demonstrates that in many areas spectral power was significantly different from "zero" across a broad range of frequencies, and crucially for all frequency bands we discuss in the manuscript.



Supplementary Figure 3.3. Microstructural properties in the sighted and blind groups. A. Anatomical connectivity data indicated structural group differences between the congenitally blind and sighted groups. Radial diffusivity (RD) statistical contrast maps were generated using tract-based spatial statistics (TBSS) and display congenitally blind over sighted participants' values. To facilitate the visualization of the pattern of results, box plots (N_{CB} = 16, $N_{Sighted}$ = 12) with the mean (center line) ± *SD* (gray area) RD values of the significant cluster are shown for each group. **B.** Images of the white matter cluster (green) that showed significant group differences (N_{CB} = 16, $N_{Sighted}$ = 12; family-wise error (FWE) corrected, two-sided $p < .05$, threshold- free cluster enhancement) are shown. Neurological convention is used, with MNI coordinates at the bottom of each slice. **C.** Scatterplots visualize relations between region and frequency specific spectral clusters and RD values. More specifically, spectral clusters for which correlations (Spearman correlation, FDR- corrected: Q = 0.05, $p < .016$) between RD and power were significant, are illustrated (cf. Supplementary Table 3.1 for the results of all correlations and a more detailed description of the correlation analysis). Dots represent individual subjects and are color-coded according to the group (blue = congenitally blind, magenta = sighted). L - left; MFG - middle frontal gyrus; MTG - middle temporal gyrus; R - right; SMA - supplementary motor area; STG - superior temporal gyrus.


Blind
Sighted (blindfolded)
Normalized
Not-Normalized
Normalized
Not-Normalized
Calcarine R
Cuneus L
Cuneus R
Lingual R
Occipital Sup R
Heschl L
Heschl R
Normalised power
Frequency (Hz)
1 3 5 10 20 50 81120

Supplementary Figure 3.4. Spectral profiles without normalization (control analysis). Clustered spectral profiles derived from the non-normalized data shown separately for a selection of brain areas (rows) for the congenitally blind (first and second column) and blind- folded sighted participants (third and fourth columns). The first and the third columns represent normalized data, while the second and fourth columns show non-normalized data. The comparison of the spectral profiles of the normalized and the non-normalized spectra displays two observations: First, overall there was congruence between the normalized and non-normalized spectra; Second, the normalized spectra showed a reduced presence of the strong alpha and low-frequency power across the brain, demonstrating that our normalization procedure was successful.



Supplementary Figure 3.5. Higher gamma band power in Calcarine gyrus in the congenitally blind compared to the sighted individuals. In the congenitally blind individuals, the clustering procedure revealed a spectral cluster with a power peak in the low gamma frequency band in right Calcarine gyrus. In contrast, no power peak was apparent in the spectral profile of the sighted individuals in the same frequency band and brain area. When the unclustered spectra were statistically compared (cf., post-hoc analysis in S4), we found no significant group differences within the low gamma band. The comparison of the unclustered spectra is a more conventional approach, which how- ever, is lacking the advantages of the clustering procedure. The clustering procedure allows to detect more nuanced spectral effects by separating distinct spectral clusters. Thus, to further investigate the apparent group difference in gamma power in right Calcarine gyrus, we statistically compared the gamma power as obtained from the clustering procedure. To this end, all single-subject clusters contributing to the group cluster in right Calcarine gyrus were identified. For subjects contributing multiple 1st-level clusters to the group cluster, 1st-level clusters were averaged which resulted in one cluster per subject. Peak power within the low gamma band (33.5-50 Hz) was extracted and a Mann-Whitney-U-test was performed to test for statistical group differences. The analysis showed higher gamma power in the congenitally blind individuals compared to the sighted ($z = -2.082$, $p = .0374$).

Supplementary Table 3.1. Correlation between RD values and spectral power. For the correlation analysis, all brain areas in occipital cortex were included (if power peaks were clearly distinguishable). For each of these brain regions the power peaks of clusters were retrieved in the following way: the single subject clusters (i.e., first-level clusters) that contributed to each cluster of a spectral profile on the group-cluster level were identified and individual power peaks were extracted. Power values were correlated with RD values (Spearman correlation, FDR-corrected: Q = 0.05, corrected *p*-value = .0065, *p*-values < .0065). ROI - region of interest; H - hemisphere; IOG – inferior occipital gyrus; L - left; MOG - middle occipital gyrus; R - right; SOG - superior occipital gyrus.

		Alpha band		Gamma band	
ROI	H	rho	p	rho	p
Calcarine	L	-0.81	< .001*	-0.18	.371
Calcarine	R	-0.55	.005*	0.36	.083
Cuneus	L	-0.67	< .001*	-0.09	.666
Cuneus	R	-0.29	.13	0.24	.222
Lingual	L	0.57	.007	0.40	.070
Lingual	R	0.01	.960	-0.08	.727
SOG	L	-0.51	.022	0.15	.535
SOG	R	-0.19	.333	0.46	.016
MOG	L	0.38	.049	0.43	.026
MOG	R	0.02	.913	0.37	.057
IOG	L	-0.52	.013	0.15	.493
IOG	R	-0.5	.015	0.16	.480

Supplementary Table 3.2. Table of all brain areas (out of 115) with significant classification differences without normalization. Brain areas where classification ranks were significantly different between the congenitally blind and the sighted (blindfolded) are listed. The areas highlighted in bold showed significant group differences in the cross-group classification analyses (i.e., with normalization procedure) and in the control analysis (i.e., without normalization procedure; see Fig4C and Supplementary Fig. 3.4) and will be the focus of interpretation.

Coarse region	Hemisphere	
	Left	Right
Frontal	Superior medial frontal Medial frontal (orb) Gyrus rectus	**Superior medial frontal** Medial frontal (orb) Gyrus rectus Superior frontal (orb) Middle frontal (orb)
Visual	Calcarine gyrus Caudate	**Calcarine gyrus** **Cuneus**
Temporal	Superior temporal pole **Middle temporal pole**	**Heschl's gyrus**
Non-cortical	**Anterior cingulate cortex** Olfactory Cerebellum lobule VI	Anterior cingulate cortex **Middle cingulate cortex**

4 Study 2[2]: Explaining flexible continuous speech comprehension from individual motor rhythms

4.1 Abstract

When speech is too fast, the tracking of the acoustic signal along the auditory pathway deteriorates, leading to suboptimal speech segmentation and decoding of speech information. Thus, speech comprehension is limited by the temporal constraints of the auditory system. Here we ask whether individual differences in auditory-motor coupling strength in part shape these temporal constraints. In two behavioral experiments, we characterize individual differences in the comprehension of naturalistic speech as function of the individual synchronization between the auditory and motor systems and the preferred frequencies of the systems. Obviously, speech comprehension declined at higher speech rates. Importantly, however, both higher auditory-motor synchronization and higher spontaneous speech motor production rates were predictive of better speech-comprehension performance. Furthermore, performance increased with higher working memory capacity (Digit Span) and higher linguistic, model-based sentence predictability – particularly so at higher speech rates and for individuals with high auditory-motor synchronization. The data provide evidence for a model of speech comprehension in which individual flexibility of not only the motor system but also auditory-motor synchronization may play a modulatory role.

[2] This section is highly adopted from a published article (**Lubinus, C.**, Keitel, A., Obleser, J., Poeppel, D., & Rimmele, J. M., 2023, Proc. Roy. Soc. B.). CL and JR designed the project; CL implemented experiment 2 and collected the data; AK contributed the dataset from experiment 1; CL analyzed all data under the supervision of JR and wrote the first draft of the manuscript. All authors revised and approved the final version of the manuscript.

4.2 Introduction

Speech comprehension relies on temporal processing, as speech and other naturalistic signals have a complex temporal structure with information at different timescales (Rosen, 1992). The temporal constraints of the auditory system limit our ability to understand speech at fast rates (Brungart et al., 2007; Nourski et al., 2009). Interestingly, the motor system can under certain conditions provide temporal predictions that aid auditory perception (Morillon et al., 2014; Stokes et al., 2019). Accordingly, current oscillatory models of speech comprehension propose that properties of the auditory but also the motor system affect the quality of auditory processing (Poeppel and Assaneo, 2020; Assaneo et al., 2021). In two behavioral experiments, we investigate how the auditory, the motor system, and their synchronization shape individual flexibility of comprehending fast continuous speech.

Auditory temporal constraints have been observed as preferred rates of auditory speech (Pellegrino et al., 2011; Ding et al., 2017) processing (but also of tones (Teng et al., 2017; Kern et al., 2021), and amplitude modulated sounds (Viemeister, 1979; Drake and Botte, 1993; Teng et al., 2017; Teng and Poeppel, 2020)) and explained in the context of neurocognitive models of speech perception. According to such proposals, humans capitalize on temporal information by dynamically aligning ongoing brain activity in auditory cortex to the temporal patterns inherent to the acoustic speech signal (Ghitza and Greenberg, 2009; Giraud and Poeppel, 2012; Peelle and Davis, 2012; Gross et al., 2013). By hypothesis, endogenous theta brain rhythms in auditory cortex partition the continuous auditory stream into smaller chunks at roughly the syllabic scale by tracking quasi-rhythmic temporal fluctuations in the speech envelope. This chunking mechanism allows for the decoding of segmental phonology – and ultimately linguistic meaning (Giraud et al., 2007; Luo and Poeppel, 2007; Ghitza and Greenberg, 2009; Giraud and Poeppel, 2012). The decoding of the speech signal is accomplished seemingly effortlessly within an optimal range centered in the traditional theta band (Ghitza and Greenberg, 2009), whereas comprehension deteriorates strongly for speech presented beyond ~9 Hz (Brungart et al., 2007; Nourski et al., 2009). While much research has focused on the apparent *stability* of the average acoustic modulation rate at the syllabic scale (Pellegrino et al., 2011; Ding et al., 2017), the *flexibility* in speech comprehension (Greenberg et al., 2003; Pellegrino et al., 2011), that is, what constitutes individual differences in understanding fast speech rates, is poorly understood.

The motor system, and neural auditory-motor coupling in particular, is a plausible candidate to facilitate individual differences in auditory speech processing abilities. Two arguments supporting this notion are the motor systems' modulatory effect on auditory perception (Grahn and Brett, 2007; Schubotz, 2007; Morillon and Baillet, 2017) and its susceptibility to training (Grahn and Rowe, 2009; Cason et al., 2015; Du and Zatorre, 2017). While there is evidence suggesting that the auditory

and speech motor brain areas are intertwined during speech comprehension (Hickok and Poeppel, 2007; Scott et al., 2009; Evans and Davis, 2015; Cheung et al., 2016; Morillon et al., 2019), the extent to which speech motor processing modulates auditory processing is debated (Wu et al., 2014; Stokes et al., 2019; Rogalsky et al., 2022). Specifically, endogenous brain rhythms in both auditory (Luo and Poeppel, 2007; Assaneo and Poeppel, 2018) and motor (Assaneo and Poeppel, 2018; Keitel et al., 2018) cortex have been observed to track the acoustic speech signal, and are characterized by preferred frequencies (Giraud et al., 2007; Keitel and Gross, 2016; Lubinus et al., 2021). In contrast to neural measures of preferred frequencies (Rosanova et al., 2009; Keitel and Gross, 2016; Lubinus et al., 2021), here we used a behavioral estimate termed “preferred” or “spontaneous” rate. Furthermore, neural coupling between auditory and motor brain areas during speech processing (Park et al., 2015; Assaneo and Poeppel, 2018; Keitel et al., 2018; Assaneo et al., 2019) has been hypothesized to provide temporal predictions about upcoming sensory events to the auditory cortex (Morillon et al., 2014; Park et al., 2015; Haegens and Zion Golumbic, 2018; Rimmele et al., 2018). The precision of these predictions may be proportional to the strength of auditory-motor cortex coupling.

Auditory-motor cortex coupling strength varies across the population, as shown by recent work (McPherson et al., 2018; Assaneo et al., 2019, 2021; Kern et al., 2021). Assaneo et al. (2019) developed a behavioral protocol (spontaneous speech synchronization test; SSS-test) which quantifies the strength of auditory-to-motor synchronization during speech production in individuals. The authors reported that auditory-motor synchronization is characterized by a bimodal distribution in the population, classifying individuals into high versus low synchronizers. (The rejection of unimodality has been previously shown with large sample sizes (Assaneo et al., 2019; see also: Rimmele et al., 2022).) Importantly, in addition to superior behavioral synchronization, high synchronizers have stronger structural and functional connectivity between auditory and speech motor cortices (see Assaneo et al., 2019, Figure 3A and B). Thus, the SSS-test provides not only a behavioral measure but also approximates individual differences in neural auditory-motor coupling strength. We propose that the individual variability in auditory-motor synchronization, previously observed to predict differences in word learning (Assaneo et al., 2019), syllable detection (Assaneo et al., 2021), and rate discrimination (Kern et al., 2021), as well as the individual variability in preferred auditory and motor rate, predicts differences in an individuals’ ability to comprehend continuous speech at fast syllabic rates.

The influence of individual auditory-motor coupling strength on behavioral performance has so far been established for behavioral paradigms using rather basic auditory and speech stimuli, e.g. tones or syllables (Assaneo et al., 2019, 2021; Kern et al., 2021). The current study assesses its importance in a more naturalistic context: during the comprehension of continuous speech. This

adds several layers of complexity. First, as speech unfolds over time, processing of continuous, i.e. longer and more complex, speech naturally demands more working memory capacity for maintenance and access to linguistic and context information (Emmorey et al., 2017). Second, rich linguistic context is used to derive linguistic predictions about upcoming words and sentences (Arnal et al., 2011; Lewis and Bastiaansen, 2015; Kuperberg and Jaeger, 2016; Zhang et al., 2021). When linguistic predictability of a sentence is high (Grant and Seitz, 2000), speech comprehension is improved, even in adverse listening situations (Obleser et al., 2007; Obleser and Kotz, 2010). Thus, similar to auditory-motor synchronization, linguistic predictability offers a compensatory mechanism when comprehension is difficult.

In summary, we investigate the role of auditory-motor synchronization with the SSS-test and the role of preferred rhythms of the auditory and motor systems for the individual flexibility of the comprehension of continuous speech. First, based on an established literature (Dupoux and Green, 1997; Ahissar et al., 2001; Brungart et al., 2007; Ghitza and Greenberg, 2009; Pefkou et al., 2017), we expected a decline in comprehension performance at syllabic rates beyond the theta range. Second, as a faciliatory effect of auditory-motor coupling on auditory processing has been observed (Assaneo et al., 2019, 2021; Kern et al., 2021), we hypothesized that individual differences in comprehension performance could be predicted by individual auditory-motor synchronization, with superior speech comprehension for high synchronizers. Such a faciliatory effect might be strongest in demanding listening situations, such as at fast syllabic rates (Stokes et al., 2019; Kern et al., 2021). Third, while the consequences of potential individual variation in the preferred rates of the motor and auditory systems are not clearly understood, based on previous findings (Assaneo and Poeppel, 2018) we expected a systematic relation of both preferred auditory and motor rates with individual speech comprehension performance. Finally, we hypothesized that linguistic predictability and working memory span should positively affect speech comprehension. Similar to auditory-motor synchronization, we expected linguistic predictability to interact with syllabic rate, such that both systems would become stronger predictors for speech comprehension as syllabic rate increases.

4.3 Methods

Two behavioral experiments and a control experiment were conducted: Experiment 1 was performed in the laboratory and investigated the influence of the spontaneous speech motor production rate on speech comprehension performance. In Experiment 2 we aimed to understand the complex interplay of multiple variables during speech comprehension beyond the spontaneous speech motor production rate. To this end, we additionally measured participants' preferred auditory rate, auditory-motor synchronization, and working memory capacity. Experiment 2 and the

control experiment were online studies. All studies were approved by the local ethics committees (Experiment 1: committee of the School of Social Sciences, University of Dundee, UK (No. UoD-SoSS-PSY-UG-2019-88), Experiment 2 and control experiment: procedures were approved by the Ethics Council of the Max Planck Society (2017_12)).

Participants. Participants were English native speakers with normal hearing and no neurological or psychological disorders (Exp 1: N = 34, Exp 2: N = 82, Control: N = 39). Participation was voluntary. For a detailed description of participants, stimuli, exclusion criteria, and tasks please refer to Supplementary Methods, Figures 4.1-2, and Tables 4.1-2.

Design and materials

Speech comprehension task. In two speech comprehension tasks, we measured participants ability to comprehend sentences at various syllabic rates. Sentences were presented at 7 (Exp 1: [8.2, 9.0, 9.8, 11.0, 12.1, 14.0, 16.4]) or 6 (Exp 2: [5.00, 10.69, 12.48, 13.58, 14.38, 15.00]) rates. In Experiment 1, participants performed a classic intelligibility task, i.e. also termed “word identification task” (Beukelman and Yorkston, 1979; Schiavetti et al., 1984; for review: Schiavetti, 1992). On each trial (N = 70), a sentence was presented through headphones and participants verbally repeated the sentence as accurately as possible (Fig. 4.1A). Responses were recorded.

In Experiment 2, speech comprehension was measured by a word-order task. Participants listened to one sentence per trial (N = 240), followed by the presentation of two words from the sentence on screen. Participants indicated via button press which word they heard first (Fig. 4.2A).

Speech production task. In the speech production tasks we estimated participants individual spontaneous speech motor production rate. In Experiment 1, the speech production task was operationalized by participants reading a text excerpt (216 words) from a printout. Participants were instructed to read the text excerpt out loud at a comfortable and natural pace while their speech was recorded (Fig. 4.1B).

In Experiment 2, participants were asked to produce continuous, “natural” speech. To facilitate fluent production, they were prompted by a question/statement belonging to six thematic categories (6 trials; own life, preferences, people, culture/traditions, society/politics, general knowledge, see Supplementary Table 4.2). Each response period lasted 30 seconds and trials were separated by self-paced breaks (Fig. 4.2C). While speaking, participants simultaneously listened to white noise. The white noise was introduced to measure the preferred rate of the motor system, without potential interference from auditory feedback. A second reason was to be consistent with the protocol from the SSS-test (Assaneo et al., 2019; Lizcano-Cortés et al., 2022; also see below). Note that this procedure was not applied in Experiment 1.

Auditory rate task (only Exp 2). To measure participants preferred auditory rate, we implemented a two-interval forced choice (2IFC) task, presenting a reference and a comparison stimulus in random order in each trial. Participants indicated via button press which stimulus they preferred (Fig. 4.2B). Stimuli were presented at syllabic rates from 3.00 to 8.50 syllables/s (3.00, 3.92, 4.83, 5.75, 6.67, 7.58, 8.50). A reference rate, e.g. 3.00 syllables/s, was compared to all syllabic rates, including itself. For each reference/comparison pair the same sentence was presented – that is, the two stimuli in any given trial only differed in their syllabic rate. Additionally, the task included catch trials to measure participant's engagement (see Supplementary Methods for details).

Spontaneous speech synchronization (SSS) test (only Exp 2). We measured participant's auditory-motor synchronization using the SSS-test (for details: Assaneo et al., 2019). In the main task, participants listened to a random syllable train and whispered along for a duration of 80s. They were instructed to synchronize their own syllable production to the stimulus presented through their headphones (Fig. 4.2D). The syllable rate in the auditory stimulus progressively increased in frequency from 4.3 to 4.7 syllables/s in increments of 0.1 syllables/s, every 60 syllables. Participants completed two trials, while the whispering was recorded.

Participants' syllable production was masked by the simultaneously presented auditory syllable train. The masking procedure suppresses auditory feedback, allowing us better to isolate the synchronization of motor production to the auditory input, without interference of auditory feedback (Assaneo et al., 2019).

Digit span test (only Exp 2). Working memory capacity was quantified using the forward and backward (Richardson, 2007) digit span test. As for the backward test data is missing for N = 21 participants, only the forward span is reported. Digit spans were presented auditorily and participants typed in their responses (Olsthoorn et al., 2014).

Control Experiment. We designed a control experiment to test if the correct word order from the word order task of Experiment 2 could be guessed from the target words alone, that is, without understanding the sentence. The task consisted in judging which of two words would be more likely to occur first in a *hypothetical sentence*. On each trial, two words were presented on screen and participants indicated their choice via button press. Importantly, 1) participants did not listen to a full sentence at any time and 2) the target words were taken from the stimulus materials actually presented in Experiment 2.

Analysis

Spontaneous speech motor production rate (Exp 1 + 2). The individual *spontaneous speech motor production rate*, i.e. articulation rate (de Jong and Wempe, 2009), was computed using Praat software (Boersma and Weenik, 2020) by automatically detecting syllable nuclei. The number of

syllable nuclei was divided by the duration of the utterance, disregarding silent pauses. For Experiment 1, the production rate was computed across the entire reading paragraph. For Experiment 2, it was first calculated for each trial (30 s) separately. The motor rate was then averaged across all trials.

Preferred auditory rate (Exp 2). First, participants with low performance in the catch trials of the preferred auditory rate task (below 75% correct) were excluded; amongst the remaining participants (N = 82) catch trial performance was very high (M = 98.48%, SD = 3.71). To compute the preferred auditory rate, a distribution of preferred frequencies was derived from all trials -except catch trials- by aggregating the frequency of each trials' preferred item. Then a gaussian function was fitted to each participants' distribution and two parameters were extracted: the peak as index for the preferred frequency and the full-width-at-half-maximum (FWHM) as index for the specificity of the response (lower FWHM equals stronger preference for one frequency).

Auditory-motor synchronization (Exp 2). From the SSS-test (Assaneo et al., 2019) we derived the participant's auditory-motor synchronization by calculating the phase-locking value (PLV) (Lachaux et al., 1999) between the (cochlea) envelopes of the auditory and the speech signals.

$$\mathrm{PLV} = \frac{1}{\mathrm{T}}\left|\sum_{\mathrm{t}=1}^{\mathrm{T}} \mathrm{e}^{\mathrm{i}(\theta_1(\mathrm{t})-\theta_2(\mathrm{t}))}\right| \quad (1)$$

where T is the total number of time points, t denotes the discretized time, and θ_1 and θ_2 are the phase of the first and the second signals, respectively.

To obtain the cochlear envelope of the syllable train (auditory channels: 180–7,246 Hz), we used the Chimera Software toolbox (Smith et al., 2002). For the recorded speech signal the amplitude envelope was quantified as the absolute value of the Hilbert transform. Both envelopes were downsampled to 100 Hz and bandpass filtered (3.5-5.5 Hz) before their phase was extracted by means of the Hilbert transform. The PLV was first estimated for each trial of the SSS-test (time windows 5s, overlap 2s) and then averaged across runs, resulting in a mean PLV. The distribution of mean PLV values was subjected to a k-means algorithm (MacQueen, 1967) (k = 2) to split participants into a high- and a low-synchronizer group. Speech auditory-motor synchronization (PLV) was treated as bimodal variable based on previous research that rejected unimodality based on larger samples (Assaneo et al., 2019; see also: Rimmele et al., 2022).

Linguistic predictability – Recurrent neural network (Exp 2). Linguistic predictability of all stimulus sentences was measured by deriving single-sentence perplexity from a recurrent neural network (RNN) language model. A language model, such as a recurrent neural network, assigns probabilities to all words in a sequence of words. From the single-word probabilities, we derived one value per sentence, quantifying its predictability (Jurafsky and Martin, 2009; Chien and Ku, 2016).

This so-called perplexity is the most common intrinsic evaluation metric of language models (Mikolov, 2010; Merity et al., 2017; Fernandez and Downey, 2018). It is computed as the inverse of the mean probability of a sentence weighted by sentence length (Jurafsky and Martin, 2009), i.e. lower perplexity values equal higher sentence predictability (see Supplementary Methods for full details on RNN and perplexity).

Mixed effects models. For both experiments we performed mixed effects analyses to quantify how speech comprehension was affected by all variables of interest. Mixed models were computed using the R packages *lme4* (v1.1-29) and *mgcv* (v1.8-39), as set up in Rstudio (version 2022.2.1.461). Mixed-effects, rather than fixed-effects models were chosen to account for idiosyncratic variation within variables, i.e. repeated measures and therefrom resulting interdependencies between data points (Baayen et al., 2008; Barr et al., 2013). Thus, both models included random intercepts for *participant* and *items*.

In Experiment 1, we computed a generalized additive mixed-effects model (GAMM) using the *mgcv:gam* function. For the dependent variable *speech comprehension*, we calculated the percentage of correctly repeated words for each sentence and subject from the speech comprehension task. The number of correct words was counted manually and transformed into a percentage. Then the dependent variable (single-trial data) was modelled as a function of the fixed effects *syllabic rate* and *spontaneous speech motor production rate*. A random slope for *syllabic rate* could not be included because the model failed to converge, thus the model included only random intercepts. Overall, the model explained ~77% of the variance.

In Experiment 2, the dependent variable *speech comprehension* was binary (*correct* vs. *incorrect* word order judgment). Thus, we employed a generalized linear mixed-effects model (GLMM; *lme4:glmer* function) with a binomial logit link function. In terms of fixed effects, the model included all variables of interest: *syllabic rate*, *preferred motor rate*, *preferred auditory rate*, *auditory-motor synchronization*, *working memory, sentence predictability*. Additionally, we introduced several linguistic and other covariates for nuisance control(Sassenhagen and Alday, 2016): *predictability target 1*, *predictability target 2*, *sentence length* (# of words), *target distance* (i.e., distance in words between the target words), *compression/dilation* of audio file. In addition to random intercepts, the model contained a by-participant random slope for *syllabic rate*, allowing the strength of the effect of the rate manipulation on the dependent variable to vary between participants (Baayen et al., 2008; Barr et al., 2013). Continuous predictor variables were z-transformed to facilitate the interpretation and comparison of the strength of the different predictors (Schielzeth, 2010). Thus, the coefficients of all continuous predictors reflect log changes in comprehension for each unit (*SD*) increase in a given predictor. We observed no problems with (multi-)collinearity, all variance

inflation factors were < 1.2 (package car version 3.0-10 (Fox and Weisberg, 2019)). Overall, the model explained ~38% of the variance.

Control experiment. For each trial, we computed how many participants correctly guessed the word order (in percent, "*word order index*"). In a new GLMM analysis, this *word order index* was added as covariate into the model from the main analysis while all other parameters remained the same.

4.4 Results

Spontaneous speech motor production rate predicts speech comprehension

In Experiment 1, we asked the question: to what extent is speech comprehension affected by one's spontaneous speech motor production rate? *Speech comprehension* was measured as the percentage of correctly repeated words in an intelligibility task (2.75% to 93.70% on average across participants). We observed a mean *spontaneous speech motor production rate* of 4.11 syllables per second (*SD* = 0.35, min = 3.35, max = 4.85) across participants (Fig. 4.1C).

As expected, the GAMM revealed a main effect of *syllabic rate*: slower speech stimuli were associated with better speech comprehension (edf = 4.91, $F = 1222.01$, $p < .001$, Fig. 4.1D, see Supplementary Table 4.3). Importantly, we observed that the *spontaneous speech motor production rate* influenced speech comprehension: the higher the individual *spontaneous speech motor production rate*, the better the speech comprehension performance (edf = 1.00, $F = 4.25$, $p = .039$, Fig. 4.1E).

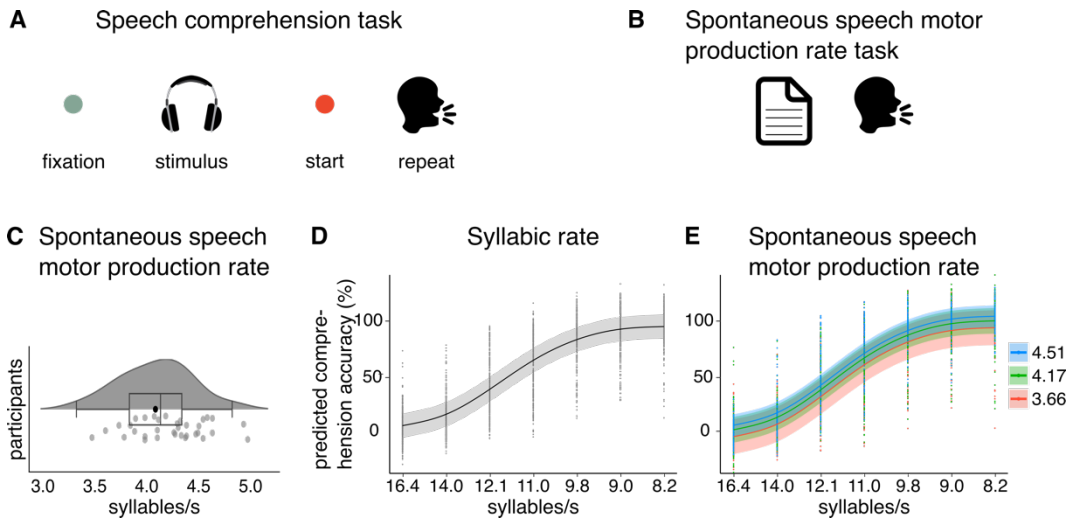


Figure 4.1. Exp. 1 – Speech comprehension is predicted by the speech motor production rate. A. Example trial for the speech comprehension task. Participants fixated on a green fixation dot while presented auditorily with a sentence. On stimulus offset the fixation dot turned red, indicating to commence

recall, i.e., reporting the sentence back. **B.** Spontaneous speech motor production rate task. Participants read a stimulus paragraph from a paper. **C.** Spontaneous speech motor production rate. We observed spontaneous speech motor production rates between 3.35 and 4.85 syllables/s (*M* = 4.11 syllables/s, left). The violin and boxplot show summary statistics and density: the median center line, 25th to 75th percentile hinges, whiskers indicate minimum and maximum within 1.5 × interquartile range. Grey dots represent participants individual speech motor productions rates, averaged across 6 trials. **D.** Main effect of syllabic rate. Plot shows the predicted main effect of syllabic rate from the generalized additive mixed model (GAMM). Black line indicates the predicted effect with 95% confidence interval in grey. Black dots show trial-level speech comprehension performance per subject and rate condition. **E.** Main effect of spontaneous speech motor production rate. Plot shows the predicted main effect of spontaneous speech motor production rate from the GAMM. Colored lines indicate the predicted effect with 95% confidence interval in the corresponding color. Colored dots show trial-level speech comprehension performance per subject and rate condition.

Spontaneous speech motor production rate, auditory-motor synchronization, and working memory predict speech comprehension

First, in line with the first experiment, we observed a mean *spontaneous speech motor production rate* of 4.30 syllables per second across participants (*SD* = 0.45, Min = 3.35, Max = 5.33 syllables per second, Fig. 4.2G). Within-subject variance was low, suggesting that participants' articulation rate was stable across trials. Second, participants showed a *preferred auditory rate* of 5.57 syllables per second (peak: *M* = 5.57, *SD* = 0.86, Min = 4.16, Max = 7.92; FWHM, *M* = 4.89, *SD* = 0.50, Min = 3.23, Max = 5.50; Fig. 4.2F). Third, *auditory-to-motor speech synchronization* was quantified using the SSS-test (Assaneo et al., 2019), classifying participants as HIGH or LOW synchronizers (mean PLV HIGHs = 0.73, *SD* = 0.09, mean PLV LOWs = 0.36, *SD* = 0.09, Fig. 4.2E). Fourth, *working memory* was measured by means of the digit span test (Richardson, 2007) which revealed a mean forward digit score of *M* = 8.46 (*SD* = 2.12, Min = 5.00, Max = 13.00, Fig. 4.2H).

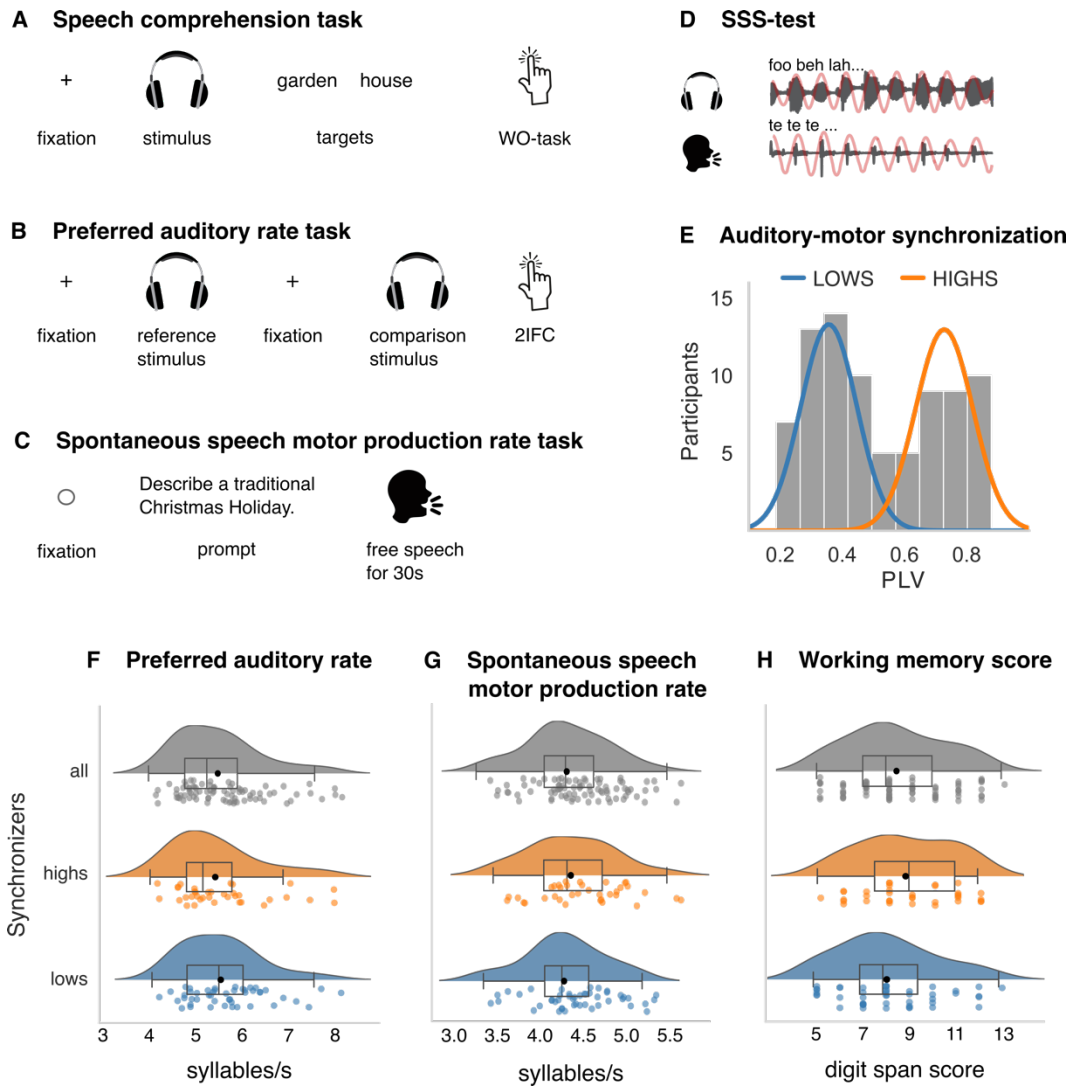


Figure 4.2. Distributions of auditory-motor synchronization, individual auditory and motor rates, and working memory. Panels **A, B,** and **C** visualize example trials for the speech comprehension task, the preferred auditory rate task, and the spontaneous speech motor production task, respectively. **D.** Schematic representation of the SSS-test, used to measure auditory-motor synchronization. Participants whisper a syllable (here /te/). **E.** Histogram of auditory-motor synchronization strength, obtained with the SSS-test. Participants were classified into high and low synchronizers (highs, lows) based on their PLV using k-means clustering. Group affiliation is overlaid by colored lines representing fitted normal distributions. **F.** Participants showed a mean preferred auditory rate of 5.57 syllables/s ($SD = 0.86$), with no differences between high and low synchronizers ($U = 897.5$, $p = .48$). **G.** We observed a spontaneous speech motor production rate between 3.36 and 5.38 syllables/s ($M = 4.32$, $SD = 0.45$) and no group difference between high and low synchronizers ($U = 751.0$, $p = .51$). **H.** Working memory capacity was indicated by a mean digit-span forward score of 8.46 ($SD = 2.12$) and the score did not differ between high and low synchronizers ($U = 666.5$, $p = .14$).

The GLMM revealed that *syllabic rate* significantly influenced participants' comprehension accuracy: for each increase of syllabic rate by one syllable/s, the odds of a correct word order judgment decreased (*odds ratio (OR)* = 0.65, *std. error (SE)* = 0.04, $p < .001$, Fig. 4.3A). This main effect of syllabic rate is consistent with a decline of speech comprehension performance at higher syllabic rates (Brungart et al., 2007). In line with our hypothesis, we observed main effects for *spontaneous*

speech motor production rate and *auditory-motor synchronization*. The higher a participant's *spontaneous speech motor production rate*, the better the performance in the word order task (*OR* = 1.19, *SE* = 0.09, *p* = .014, Fig. 4.3C), replicating our finding from the first experiment. For *auditory-motor synchronization*, being a dichotomous variable (i.e., HIGH vs. LOW)(Assaneo et al., 2019), performance in the word order judgment task was higher for high compared to low synchronizers (*OR* = 1.34, *SE* = 0.20, *p* = .048, Fig. 4.3B). That is, across all trials, high synchronizers were more likely to correctly perform the task. Additionally, the model revealed a positive effect for *working memory score* (*OR* = 1.20, *SE* = 0.09, *p* = .012, Fig. 4.3D). This main effect suggests that better working memory performance enabled participants to better perform on the speech comprehension task. We did not observe a reliable effect of *preferred auditory rate* on speech comprehension (*OR* =1.14, *SE* = 0.08, *p* = .072). In contrast to our hypothesis, we observed no interaction effect of *syllabic rate* and *auditory-motor synchronization* on speech comprehension (*OR* = 0.97, *SE* = 0.07, *p* = .602).

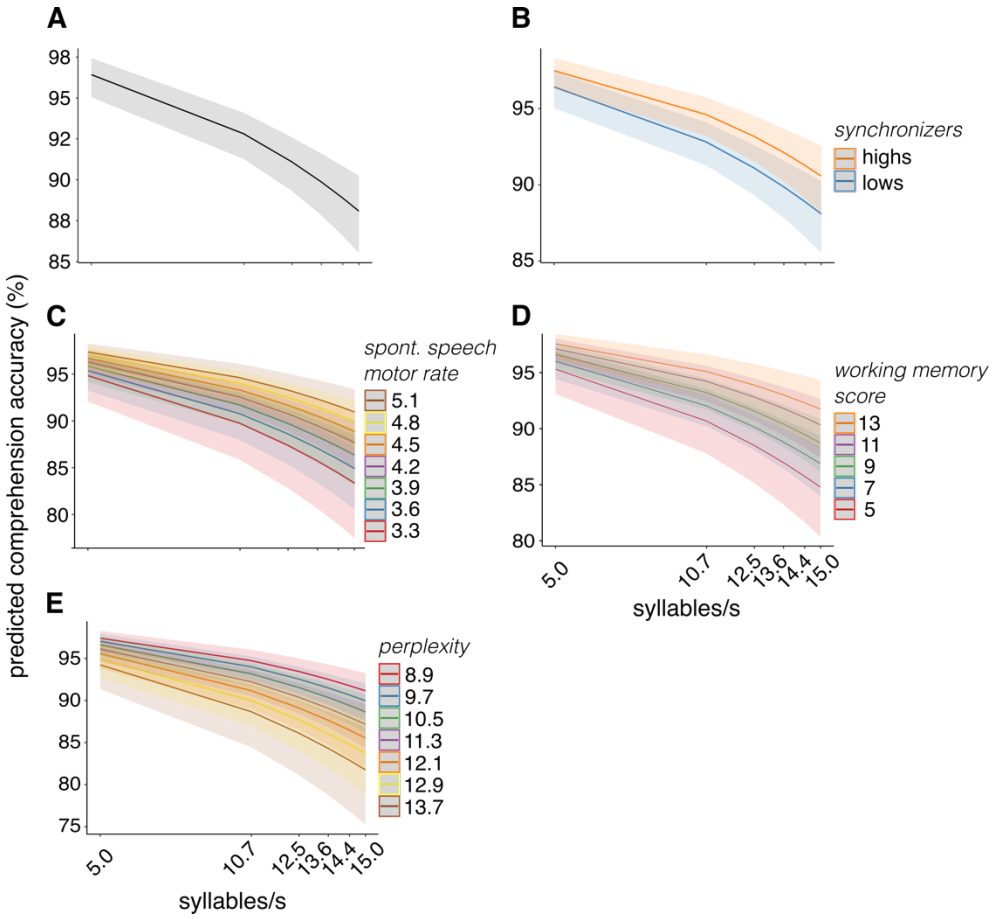


Figure 4.3. Exp. 2 – Speech comprehension is predicted by multiple variables. The generalized linear mixed effects model revealed a negative main effect of syllabic rate (**A**) and positive main effects of auditory-motor synchronization (**B**), spontaneous speech motor production rate (**C**), and working memory score (**D**). For stimulus perplexity we observed a negative main effect (**E**). In all panels, error shades indicate 95% confidence intervals. Note that the predictors are shown as a function of syllabic rate for visualization purposes only.

Contribution of linguistic predictability and further linguistic variables

To account for the effect of linguistic attributes, we expanded the GLMM by adding several (information-theoretic) linguistic variables: *perplexity*, *probability of target words*, *target distance*, and *stimulus length*. Adding these variables (with linguistic variables, AIC: 12675) improved model fit (without linguistic variables, AIC: 12848), as measured by a likelihood ratio test ($\chi2(6) = 184.24$, $p < .001$, see Supplementary Table 4.4).

The full GLMM revealed that *perplexity* had a statistically reliable, negative effect on speech comprehension (*OR* = 0.84, *SE* = 0.04, *p* = .001, Fig. 4.3E) such that sentences with lower perplexity (which is equal to higher sentence predictability) lead to better speech comprehension performance. Additionally, we observed significant negative effects for *probability of target word 1* (*OR* = 0.93, *SE* = 0.03, *p* = .026) and *target word 2* (*OR* = 0.92, *SE*: 0.03, *p* = .021). Contrary to the perplexity effect, this suggests that task performance in the comprehension task was increased for unexpected target words.

Furthermore, the model revealed a positive effect for *target distance* (*OR* = 1.48, *SE*: 0.05, *p* < .001), suggesting that larger distance between targets was associated with better speech comprehension performance. In contrast, suggesting the opposite relation, for *stimulus length* we observed a negative effect (*OR* = 0.61, *SE*: 0.03, *p* < .001), i.e., shorter sentences resulted in higher comprehension performance. Due to the large number of variables introduced for nuisance control, we applied a control for multiple comparisons (i.e. false discovery rate; for full results see Supplementary Table 4.5). All effects remained robust after FDR correction: syllabic rate: *p* < .001; spontaneous speech motor production rate: *p* = .023; preferred auditory rate: *p* = .078; working memory score: *p* = .022; perplexity: *p* = .003; probability target 1: *p* = .034; probability target 2: *p* = .030; compression: *p* < .001; sentence length: *p* < .001; target distance: *p* < .001. Only auditory-motor synchronization changed from a significant effect to a trend (*p* = 0.057) (Note that this was a planned comparison and therefore is discussed).

Finally, we explored interaction effects between *syllabic rate*, *auditory-motor synchronization*, and *perplexity*. Adding the interaction term improved model fit ($\chi^2(3) = 13.84$, $p = .004$ (AIC without interaction term: 12675, AIC with interaction term: 12668)). The model revealed two significant 2-way interaction effects: *syllabic rate* × *perplexity* (*OR* = 0.88, *SE* = 0.05, *p* = .015) and *auditory-motor synchronization* × *perplexity* (*OR* = 0.86, *SE* = 0.04, *p* = .003; see Supplementary Fig. 4.3 and Supplementary Table 4.6). The interaction effect between *syllabic rate* and *perplexity* indicates that particularly comprehension of sentences at fast syllabic rates improves when *perplexity* is low. Furthermore, the *auditory-motor synchronization* × *perplexity* interaction effect suggests that while having better overall speech comprehension, high synchronizers show a stronger effect of *perplexity* compared to low synchronizers, with even better speech comprehension for more predictable

sentences. The *syllabic rate* × *auditory-motor synchronization* effect (*OR* = 0.94, *SE* = 0.07, *p* = .392), as tested before, and the three-way interaction effect of *syllabic rate* × *auditory-motor interaction* × *perplexity* (*OR* = 1.09, *SE* = 0.06, *p* = .106) did not show a statistically reliable effect on speech comprehension.

No confounding word-order effect

In Experiment 2, speech comprehension performance was exceptionally good, even at high syllabic rates. To ensure the high performance was not an artifact of the task or stimuli, we conducted a control experiment. The analysis revealed that *word order index* did not influence speech comprehension in a statistically meaningful way (*OR* = 0.96, *SE* = 0.07, *p* = .219, see Supplementary Table 4.7).

4.5 Discussion

In two behavioral experiments, we show clear effects of *syllabic rate* on the comprehension of continuous speech. This finding is in line with proposals of speech comprehension being temporally constrained such that it is optimal for speech at lower syllabic rates. Crucially, in both protocols we observed that speech comprehension across a wide range of frequencies (5-15 syllables/s) was predicted by participants' *spontaneous speech motor production rate*, with higher rates predicting better speech comprehension. In the second experiment we showed that, beyond the spontaneous rate of the speech motor system, the individual strength of speech auditory-motor synchronization also predicted comprehension. In contrast, the preferred speech perception rate was not related to speech comprehension performance. Together, these findings suggest that while speech comprehension is limited by general processing characteristics of the auditory system, interindividual differences in comprehension flexibility are intertwined with characteristics of the motor system and auditory-motor interactions (Fig. 4.4). Our findings furthermore allow us to generalize the effects of individual differences in the motor system on auditory perception, which have been previously shown for simpler stimuli (Assaneo et al., 2019, 2021; Kern et al., 2021; Orpella et al., 2022), to more natural continuous speech.

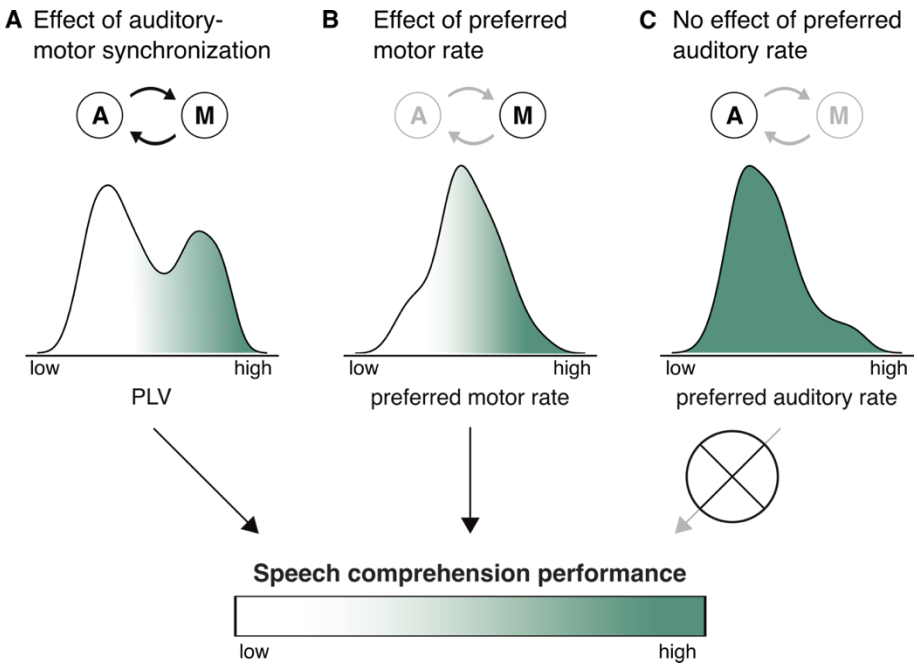


Figure 4.4. Relationship between speech comprehension and auditory-motor synchronization, the preferred motor and auditory rates. All three predictor variables are represented by the corresponding distribution generated from our experimental data. The present data propose that better speech comprehension at demanding rates – and by hypothesis, auditory behavior more generally – is accompanied by a higher preferred rate of the motor system (**B**) as well as stronger auditory-motor synchronization (**A**). In contrast, the preferred rate (**C**) of the auditory system seems not to determine auditory behavior. Circled A and M illustrate the auditory and motor systems. The arrows connecting them express the relevance of synchronization between the systems for the variable in question.

Effects of syllabic rate on speech comprehension

As expected (Dupoux and Green, 1997; Ahissar et al., 2001; Ghitza and Greenberg, 2009; Nourski et al., 2009; Pefkou et al., 2017), we observed that speech comprehension accuracy declined as syllabic rate increased. Although speech comprehension dropped at higher rates in both paradigms, the overall level of comprehension accuracy was much higher in Experiment 2, with accuracy remaining very high (~85%), even for speech as fast as 15 syllables/s. In contrast, in Experiment 1 the increase in syllabic rate resulted in a dramatic drop of comprehension performance. This is in line with our expectations, as the nature of the word-order task is likely to yield overall better performance than the classic intelligibility task. Additionally, our control experiment rules out a potential confound by demonstrating that the high performance in Experiment 2 is not due to simple guessing of the correct word order (see Results section and Supplementary Table 4.7). Interestingly, however, in both experiments performance decreased later than previously observed, that is, beyond rates of 9 syllables/s (Dupoux and Green, 1997; Ghitza, 2014). However, in line with our findings, several other studies, also observed shallower decreases in speech comprehension, with relatively high comprehension at higher syllable rates (~12 syllables/s)(Dupoux and Green, 1997; Brungart et al., 2007; Verschueren et al., 2022; Giroud et al., 2023). We consider

several possible explanations for these discrepancies. One explanation for the different and higher speech-rate decline in comprehension performance is that naturally produced fast speech (with matched degrees of compression across syllabic rates, as used in Experiment 2), in contrast to linearly compressed speech, results in more variance of the speech rate and thus allows for part of the sentences to be understood. However, this explanation does not account for Experiment 1, in which all stimuli were synthesized at the same rate (varying in degrees of compression). Furthermore, the high performance level might be related to different complexity between more naturalistic sentences, providing stronger context information to compensate loss of information, as compared to the words (Ghitza and Greenberg, 2009), digits (Doelling et al., 2014), or simple sentences (Ahissar et al., 2001) used in previous work. Finally, it is notable that while some studies conceptualized the syllabic rate based on the 'theta-syllable' (an information unit defined by cortical function (Ghitza, 2013)), we define syllabic rate as linguistically defined syllables per second, following other studies (Keitel et al., 2018).

Effects of auditory-motor synchronization on speech comprehension

Auditory-motor speech synchronization, a behavioral estimate of auditory-motor cortex coupling strength (Assaneo et al., 2019), had a modulatory -albeit small- effect on speech comprehension. We observed that high compared to low synchronizers exhibited better speech comprehension performance. These results expand on findings which showed superior statistical word learning (Assaneo et al., 2019) or syllable discrimination (Assaneo et al., 2021) for individuals with stronger auditory-motor coupling by showing a similar effect for comprehending more naturalistic, continuous speech. Note that this effect requires further validation as it did not survive control for multiple comparisons (Supplementary Table 4.5). Additionally, we expected an interaction of syllabic rate and auditory-motor synchronization, as reported for rate discrimination in tone sequences (Kern et al., 2021). However, the modulation observed here occurred across all syllabic rates, suggesting that an interaction effect may be masked and compensated for by context and linguistic information in continuous speech comprehension. Alternatively, it is possible -although unlikely- that the interaction of syllabic rate and auditory-motor synchronization was not observed here due to the different frequency resolution at low frequencies. The difference between HIGHs and LOWS in Kern et al. (Kern et al., 2021) manifested between 7.14 and 10.29 Hz. In contrast, in the present experiment, there was no frequency condition between 5 and 10.69 syllables/s.

Faster spontaneous motor rates predict better speech comprehension

Importantly, the spontaneous motor production rate affected speech comprehension, suggesting that individuals with a higher spontaneous motor production rate have increased speech comprehension (at the higher range). We replicated this finding in the second experiment. The finding is

likely to reflect a complex interplay of auditory and motor cortex during speech comprehension wherein not only the coupling strength, but also the preferred rates of the motor cortex affect speech perception. A possible role of the preferred speech motor rate for speech processing has been previously discussed (Assaneo and Poeppel, 2018). Furthermore, our findings are in line with an oscillatory model of speech comprehension (Assaneo et al., 2021). An alternative interpretation of our findings might be that general processes such as vigilance and fatigue are equally reflected in the spontaneous speech motor production rate and the speech comprehension performance. This could be because speech comprehension is tightly intertwined with production, and vigilance effects on production, for example, might similarly affect comprehension. Spontaneous production rates might also be more prone to vigilance effects compared to measures of production performance (e.g. McPherson et al., 2018). It is notable that although the preferred spontaneous motor production rates observed here are close to the rates at which speech comprehension has been reported to decline in earlier studies (Ahissar et al., 2001; Brungart et al., 2007; Ghitza and Greenberg, 2009; Nourski et al., 2009), these rates are further apart in our study. The behavioral protocol does not allow to rule out such an alternative interpretation. However, given that no correlation of a demanding cognitive task (Digit span) with the spontaneous speech motor production rates was observed (see Supplementary Material), we consider this unlikely. Furthermore, for the effects of speech auditory-motor synchronization on syllable discrimination, others have ruled out such an interpretation (Assaneo et al., 2021).

Preferred auditory rate does not influence speech comprehension

Interestingly, the preferred auditory rate (~5.55 syllables/s) had no effect on speech comprehension in our study. A possible explanation is that preferred rates in auditory cortex are less flexible compared to preferred rates in motor cortex and thus less prone to individual difference related improvements of speech comprehension. However, comparing the variances of the distribution of preferred auditory ($s^2 = 0.74$) and motor ($s^2 = 0.20$) rates revealed bigger variance in the auditory rate ($F(1,162) = 22.39$, $p < .001$). Another possibility is that the behavioral estimation of preferred auditory cortex rates was not optimally operationalized. This might also explain the lack of correlation between preferred auditory and spontaneous speech production rates (see Supplementary Material), which we expected to be correlated. Generally, our behavioral protocol only allows for an indirect assessment of preferred neural rates. Nevertheless, behavioral measures have been regarded as proxy for underlying intrinsic brain rhythms (McAuley et al., 2006; Michaelis et al., 2014; Provasi et al., 2014; McPherson et al., 2018). Finally, the rates at which speech comprehension decreases are much higher than the preferred auditory and spontaneous speech motor production rates. While the preferred rates were well within the expected range (Ding et al., 2017;

Poeppel and Assaneo, 2020), the mismatch between maximal comprehension rates and preferred rates was due to the high speech comprehension ability of participants even at high rates.

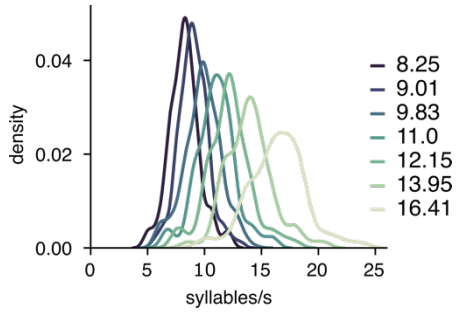
Linguistic predictability affects comprehension

We show that continuous speech comprehension is additionally affected by other higher cognitive and linguistic factors. The relevance of linguistic predictability and working memory capacity have been shown in multiple studies (Obleser et al., 2007; Obleser and Kotz, 2010). In agreement with these studies, such cognitive variables explained a large amount of variance in speech comprehension. Interestingly, our findings suggest that the faciliatory effect of linguistic predictability is particularly effective at fast rates. Second, we tentatively interpret that facilitation due to linguistic predictability may be used more efficiently from individuals with stronger auditory-motor synchronization. A relevant question arising from this is: under what conditions is the impact of the motor system on speech comprehension the strongest? Previous work observed an impact of the motor system on speech comprehension in demanding listening conditions, such as listening to speech in noise (Wu et al., 2014; Stokes et al., 2019). Our data suggests that this view might extend towards conditions of fast speech (which requires more experiments) or might interact with linguistic predictability.

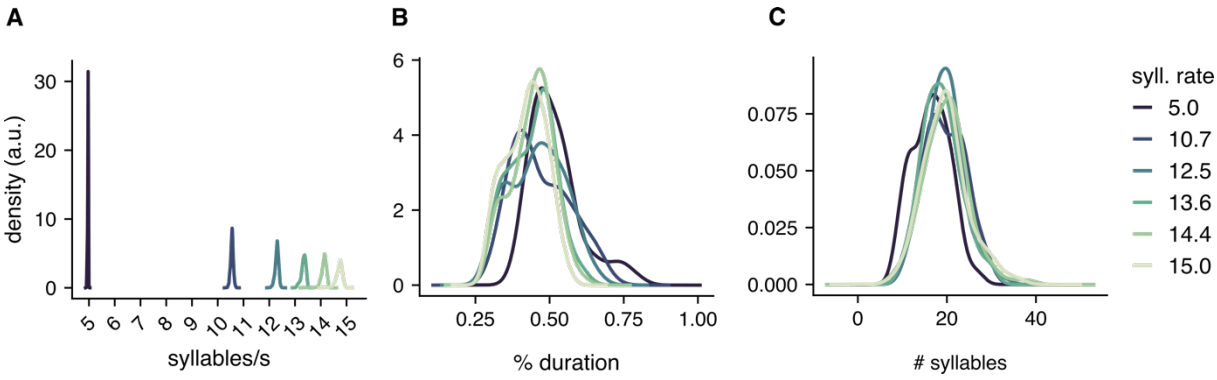
4.6 Conclusions

Speech comprehension is a highly predictive process which is affected by different sources of predictions. Here we show that, while speech comprehension is optimal in a preferred auditory temporal regime, the motor-system possibly provides a source for individual flexibility in continuous speech comprehension. Additionally, we report that the well-known facilitatory effect of linguistic predictability on speech comprehension interacts with individual differences in the motor system. This motivates future assessments of how predictions from these systems interact and under what circumstances the human brain relies more on one over the other.

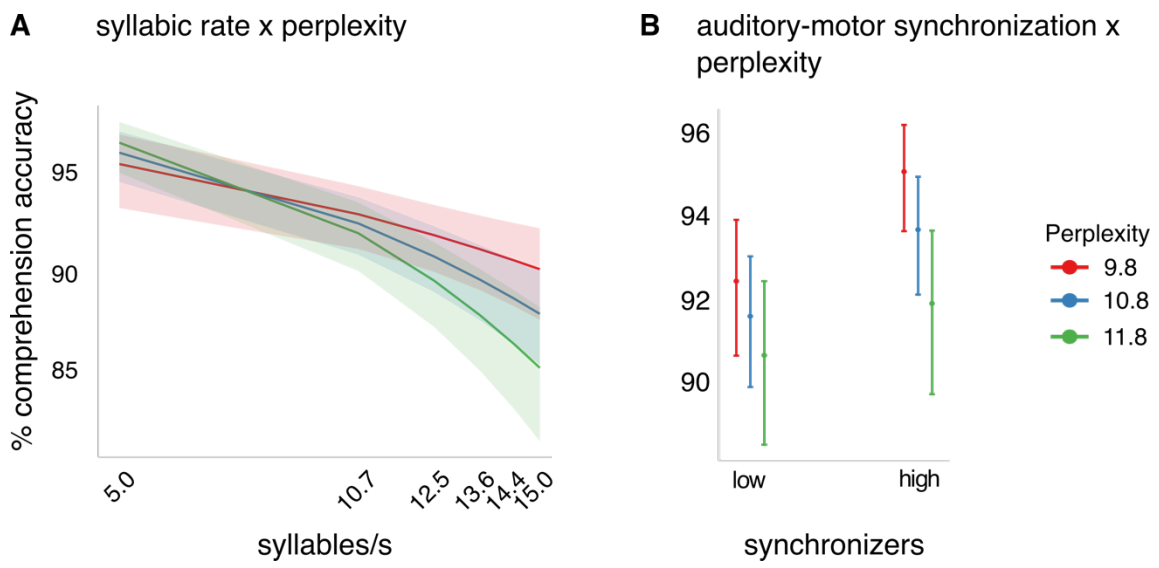
4.7 Supplementary materials



Supplementary Figure 4.1. Exp. 1 – Corresponding syllabic rate and compression rate in speech stimuli. Sentence stimuli were synthesized and then compressed to a percentage (49%, 45%, 41%, 37%, 33%, 29%) of their original duration, yielding sentences of different speech rates. Since we were interested in the effect of syllabic rate on speech comprehension (not compression) and for better comparison with experiment 2, we computed the mean syllabic rate within each compression bin. To this end, we divided the number of syllables by the duration of the compressed stimulus which resulted in a distribution of syllabic rates. The distribution's means were defined as the corresponding syllabic rates. Importantly, the mapping between compression rate and syllabic rate is not unambiguous, as illustrated by the density plots. Color coding reflects the compression rates.



Supplementary Figure 4.2. Exp. 2 – Stimulus parameters for speech comprehension task. A. Syllabic rates. Speech stimuli were manipulated with respect to syllabic rate (5.0, 10.7, 12.5, 13.6, 14.4, 15.0 syllables/s), as visualized by separate distribution for each rate condition. Within rate conditions stimuli showed narrow distributions around the target rate. **B.** Compression rate. To assure comprehension performance was not confounded by systematic differences between compression strength between the syllabic rate conditions, the distribution of compression rates was kept as similar as possible across conditions (see Supplementary Methods 2). **C.** Length of sentences. Across syllabic rates, the length of sentences (i.e. number of syllables) was similar as shown by overlapping distributions, suggesting that the rate conditions should not differ in effects related to sentence length (e.g., working memory load or complexity).



Supplementary Figure 4.3. Exp. 2 – Interaction effects of *syllabic rate*, *auditory-motor synchronization*, and *perplexity*. A. The generalized additive mixed model revealed a significant *syllabic rate* x *perplexity* interaction such that comprehension was best for sentences of highest predictability particularly at high demanding rates. **B.** We observed an interaction effect of *auditory-motor synchronization* x *perplexity,* suggesting a stronger comprehension gain as a function of predictability for high synchronizers, particularly when linguistic predictability is high.

Supplementary Table 4.1. Exp. 2 – Sentence materials: Titles and authors of (audio)books. List of sources from which stimuli for the tasks (speech comprehension and preferred auditory rate tasks) were constructed.

Type	Author	Title	Source
Audiobook	James Weldon Johnson	The Autobiography of An Ex-Colored Man	Lit2Go
Audiobook	Edith Wharton	Ethan Frome	Lit2Go
Audiobook	Frances Hodgson Burnett	A little princess	Librivox
Audiobook	Horace Walpole	Castle of Otranto	Lit2Go
Audiobook	Henry Ossian Flipper	The Colored Cadet at West Point	Lit2Go
PDF	Frances Hodgson Burnett	The Secret Garden	Lit2Go
PDF	Lucy Maud Montgomery	Anne of Green Gables	Lit2Go
PDF	Sinclair Lewis	The Job	Librivox
PDF	Booker T. Washington	Up from Slavery	Lit2Go

Supplementary Table 4.2. Exp. 2 – Prompts for speech production task. Thematic questions used to facilitate natural speech production. Each item is representative of a different thematic category, as introduced by Alexandrou et al.(2016).

Category	Sentence/statement
Own life	What kind of hobbies do you have or have had during your life?
Preferences	What kinds of vacation trips do you like?
People	Describe a known artist, writer, or film director. Why do you find her/him interesting?
Culture/traditions	Describe a traditional Christmas holiday.
Society/politics	What do you know about garbage and recycling policies in your home country?
General knowledge	What do you know about skiing and snowboarding?

Supplementary Table 4.3. Exp. 1 – Predicting single-trial comprehension performance.

Speech comprehension accuracy				
Parametric coefficients				
Predictors	*Estimate*	*std. Error*	*t-value*	*Pr(>|t|)*
Intercept	54.720	2.188	25.01	<0.001
Approximate significance of smooth terms				
	edf	*F-value*	*p-value*	
Compression rate	4.599	1241.436	<0.001	
Speech motor prod. rate	1.000	4.336	0.037	
sub	30.231	17.077	<0.001	
Trial-ID	60.939	7.591	<0.001	
R-sq. (adj) = 76.1		Deviance explained = 77%		
Observations = 2373				

Supplementary Table 4.4. Exp. 2 – Predicting single-trial comprehension performance.

	comprehension_accuracy				
Predictors	*Odds Ratios*	*std. Error*	*CI (95%)*	*z-values*	*p*
(Intercept)	11.27	1.18	9.18 – 13.84	23.14	**<0.001**
syllabic rate	0.65	0.04	0.57 – 0.73	-6.70	**<0.001**
synchronization (HIGH vs LOW)	1.34	0.20	1.00 – 1.79	1.98	**0.048**
speech motor prod. rate	1.19	0.09	1.04 – 1.37	2.45	**0.014**
pref. auditory rate	1.14	0.08	0.99 – 1.31	1.80	0.072
working memory	1.20	0.09	1.04 – 1.39	2.52	**0.012**
perplexity	0.84	0.04	0.76 – 0.94	-3.19	**0.001**
probability target1	0.93	0.03	0.88 – 0.99	-2.23	**0.026**
probability target2	0.92	0.03	0.85 – 0.99	-2.31	**0.021**
compression	1.21	0.06	1.10 – 1.33	3.87	**<0.001**
sentence length	0.61	0.03	0.55 – 0.68	-9.24	**<0.001**
target distance	1.48	0.05	1.37 – 1.59	10.60	**<0.001**
syllabic rate * synchronization	0.97	0.07	0.84 – 1.10	-0.52	0.602
Random Effects					
σ^2	3.29				
$\tau_{00\ file}$	0.74				
$\tau_{00\ sub}$	0.36				
$\tau_{11\ sub.scale(freq)}$	0.02				
$\varrho_{01\ sub}$	0.00				
ICC	0.25				
N_{sub}	82				
N_{file}	495				
Observations	19680				
Marginal R^2 / Conditional R^2	0.139 / 0.357				

Supplementary Table 4.5. Exp. 2 – Predicting single-trial comprehension performance (with FDR-correction)

	comprehension_accuracy				
Predictors	*Odds Ratios*	*std. Error*	*CI (95%)*	*z-values*	*p*
(Intercept)	11.27	1.18	9.18 – 13.84	23.14	**<0.001**
syllabic rate	0.65	0.04	0.57 – 0.73	-6.70	**<0.001**
synchronization (HIGH vs LOW)	1.34	0.20	1.00 – 1.79	1.98	0.057
speech motor prod. rate	1.19	0.09	1.04 – 1.37	2.45	**0.023**
pref. auditory rate	1.14	0.08	0.99 – 1.31	1.80	0.078
working memory	1.20	0.09	1.04 – 1.39	2.52	**0.022**
perplexity	0.84	0.04	0.76 – 0.94	-3.19	**0.003**
probability target1	0.93	0.03	0.88 – 0.99	-2.23	**0.034**
probability target2	0.92	0.03	0.85 – 0.99	-2.31	**0.030**
compression	1.21	0.06	1.10 – 1.33	3.87	**<0.001**
sentence length	0.61	0.03	0.55 – 0.68	-9.24	**<0.001**
target distance	1.48	0.05	1.37 – 1.59	10.60	**<0.001**
syllabic rate * synchronization	0.97	0.07	0.84 – 1.10	-0.52	0.602
Random Effects					
σ^2	3.29				
$\tau_{00\ file}$	0.74				
$\tau_{00\ sub}$	0.36				
$\tau_{11\ sub.scale(freq)}$	0.02				
$\varrho_{01\ sub}$	0.00				
ICC	0.25				
N_{sub}	82				
N_{file}	495				
Observations	19680				
Marginal R^2 / Conditional R^2	0.139 / 0.357				

Supplementary Table 4.6. Exp. 2 – Predicting single-trial comprehension performance including a 3-way interaction term of syllabic rate x synchronization x perplexity. Output of generalized mixed-effects model, showing predictions of single-trial speech comprehension when controlling for word-order effects of the target words.

	comprehension_accuracy				
Predictors	*Odds Ratios*	*std. Error*	*CI (95%)*	*z-values*	*p*
(Intercept)	11.12	1.16	9.06 – 13.65	23.01	**<0.001**
syllabic rate	0.66	0.04	0.58 – 0.75	-6.37	**<0.001**
synchronization (HIGH vs LOW)	1.35	0.20	1.01 – 1.81	2.04	**0.041**
perplexity	0.89	0.05	0.80 – 1.00	-1.99	**0.046**
speech motor prod. rate	1.19	0.09	1.04 – 1.37	2.45	**0.014**
pref. auditory rate	1.14	0.08	0.99 – 1.31	1.80	0.072
working memory	1.20	0.09	1.04 – 1.39	2.52	**0.012**
probability target1	0.93	0.03	0.88 – 0.99	-2.20	**0.028**
probability target2	0.91	0.03	0.85 – 0.98	-2.41	**0.016**
compression	1.22	0.06	1.10 – 1.34	4.00	**<0.001**
sentence length	0.61	0.03	0.55 – 0.68	-9.17	**<0.001**
target distance	1.47	0.05	1.37 – 1.58	10.58	**<0.001**
syllabic rate * synchronization	0.94	0.07	0.82 – 1.08	-0.86	0.392
syllabic rate * perplexity	0.88	0.05	0.80 – 0.98	-2.43	**0.015**
synchronization * perplexity	0.86	0.04	0.78 – 0.95	-2.96	**0.003**
syllabic rate * synchronization * perplexity	1.09	0.06	0.98 – 1.21	1.62	0.106
Random Effects					
σ^2	3.29				
$\tau_{00\ file}$	0.73				
$\tau_{00\ sub}$	0.56				
$\tau_{11\ sub.freq}$	0.00				
$\varrho_{01\ sub}$	-0.60				
ICC	0.28				
N_{sub}	82				
N_{file}	495				
Observations	19680				
Marginal R^2 / Conditional R^2	0.139 / 0.382				

Supplementary Table 4.7. Control experiment – Predicting single-trial comprehension performance including the word order index. Output of generalized mixed-effects model, showing predictions of single-trial speech comprehension when controlling for word-order effects of the target words.

	comprehension_accuracy				
Predictors	*Odds Ratios*	*std. Error*	*CI (95%)*	*z-values*	*p*
(Intercept)	11.25	1.18	9.17 – 13.82	23.13	**<0.001**
syllabic rate	0.65	0.04	0.57 – 0.73	-6.69	**<0.001**
synchronization (HIGH vs LOW)	1.34	0.20	1.01 – 1.80	2.01	**0.045**
speech motor prod. rate	1.19	0.09	1.03 – 1.37	2.42	**0.016**
pref. auditory rate	1.14	0.08	0.99 – 1.31	1.80	0.072
working memory	1.20	0.09	1.04 – 1.39	2.47	**0.013**
perplexity	0.84	0.04	0.76 – 0.94	-3.19	**0.001**
probability target1	0.93	0.03	0.88 – 0.99	-2.19	**0.028**
probability target2	0.92	0.03	0.85 – 0.99	-2.30	**0.022**
compression	1.21	0.06	1.10 – 1.33	3.89	**<0.001**
sentence length	0.61	0.03	0.55 – 0.68	-9.23	**<0.001**
target distance	1.48	0.05	1.37 – 1.59	10.62	**<0.001**
word order	0.96	0.03	0.90 – 1.02	-1.23	0.219
syllabic rate * synchronization	0.97	0.07	0.84 – 1.10	-0.52	0.606
Random Effects					
σ^2	3.29				
$\tau_{00\ file}$	0.74				
$\tau_{00\ sub}$	0.55				
$\tau_{11\ sub.freq}$	0.00				
$\varrho_{01\ sub}$	-0.59				
ICC	0.28				
N_{sub}	82				
N_{file}	495				
Observations	19680				
Marginal R^2 / Conditional R^2	0.134 / 0.379				

Participants

Experiment 1. Participation was voluntary and participants had the chance of winning a £25 voucher by participating in a prize draw. We recruited participants through opportunity sampling at the University of Dundee. All participants were native English speakers (N = 34, female = 18, male = 14, non-binary = 2, age: M = 22.12, SD = 1.87), right-handed and reported normal hearing, as well as no neurological or psychological disorders. The experiment complied with the Declaration of Helsinki and was approved by the ethics committee of the School of Social Sciences, University of Dundee, UK (No. UoD-SoSS-PSY-UG-2019-88). The experimental tasks were presented using Psychtoolbox (version 3.0.16) for MATLAB (version R2017b) on a Windows computer. Participants were equipped with non-wireless DRACO HS-880 headphones (Creative) with an integrated microphone to record all speech signals.

Experiment 2. Participants were native speakers of North-American English, born in the United States or Canada, recruited from the online portal "Prolific" (https://www.prolific.com/). The following criteria were used for Prolific's participants prescreening: normal or corrected-to-normal vision, no hearing issues, no prior or current psychological or neurological diseases, aged 18-45 years. Furthermore, using Prolific's compliance metrics, we only included participants with a minimum approval rate of 90% in previous experiments and a minimum of 50 previous submissions. The final sample included 82 participants (37 Females; age: *M* = 28.6 years; *SD* = 6.3), 36 High and 46 Low synchronizers. Experimental tasks were presented in the web browser using JsPsych (6.1.0) and the experiment was hosted on an in-house Jatos (3.5.5) server. The experiment complied with the Declaration of Helsinki and the procedures were approved by the Ethics Council of the Max Planck Society (2017_12).

Control experiment. A new set of participants (N = 39, 13 per stimulus list, female = 10, age: M = 29.1, STD = 10.1) was recruited from Prolific (same inclusion criteria as in Experiment 2, enriched only by the criterion of not having participated in the previous study). Experimental setup and ethics approval are identical to Experiment 2.

Stimulus selection, recording, and processing

Experiment 1 – Speech comprehension task. During stimulus generation, first, sentences (between 5 to 8 words) were generated using the online tool SKELL (https://skell.sketchengine.co.uk/run.cgi/skell). Second, the sentences were synthesized using Google Cloud's text2speech (male voice 'en-GB-Wavenet-B', https://cloud.google.com/text-to-speech) to generate audio files at a sampling rate of 44,100 Hz and 2dB volume gain. This text2speech algorithm produces human-like speech, generating stimuli consistent in speech rate and loudness. Finally, the synthesized audio files were digitally compressed using the Pitch Synchronous Overlap and Add (PSOLA) algorithm(Moulines and Charpentier, 1990) implemented in Praat (6.0.18)(Boersma and Weenik, 2020). Importantly, speech rate was manipulated by means of compression rate (percentage of stimulus duration), that is, compression rate varied between different syllabic rates. For stimulus definition/generation and the statistical analysis, stimuli were grouped based on compression rate. For visualization and easier comparison with the second experiment, we transformed compression rate into *syllabic rate* (syllables/s). The mapping between compression and syllabic rate is not unambiguous in that any given compression rate contained distinct syllabic rates (for correspondence between both measures see Supplementary Fig. 4.1).

Experiment 1 – Reading excerpt for speech production task. Dennis was different. When he looked in the mirror, he saw an ordinary twelve-year-old boy. But he felt different – his thoughts were full of colour and poetry, though his life could be very boring. The story I am going to tell you begins here, in Dennis's ordinary house on an ordinary street in an ordinary town. His house was nearly exactly the same as all the others in the street. One house had double glazing, another did not. One had a gravel drive, another had crazy paving. One had a Vauxhall Cavalier in the drive, another a Vauxhall Astra. Tiny differences that only really pointed out the sameness of everything. It was all so ordinary, something extraordinary just had to happen. Dennis lived with his dad – who did have a name, but Dennis

just called him Dad, so I will too – and his older brother John, who was fourteen. Dennis found it frustrating that his brother would always be two years older than him, and bigger, and stronger. Dennis's mum had left home a couple of years ago. Before that, Dennis used to creep out of his room and sit at the top of the stairs and listen to his mum and dad shout at each other until one day the shouting stopped. She was gone.

Experiment 2 – Sentence materials for speech comprehension and preferred auditory tasks. When studying the comprehension of accelerated speech, the syllabic rate of stimuli is usually manipulated by digital compression. A consisting finding is that comprehension or intelligibility deteriorates as the syllabic rate of sentences increases(Garvey, 1953; Goldman-Eisler, 1961; Mehler et al., 1993; Ahissar et al., 2001). In these studies, the degree of compression and the syllabic rate of sentences are typically correlated, such that faster sentences are compressed more heavily. Importantly, digital acceleration not only alters speech rate but also the acoustic properties of a signal(Janse, 2004). Considering the relevance of acoustic features for speech intelligibility (e.g. edges(Doelling et al., 2014)), in Experiment 2 we carefully controlled for potential compression artifacts by balancing compression factors across syllabic rates. To this end, we constructed a set of stimuli with a broad range of original speech rates. As speech rate in audiobooks, and natural speech more generally, is robustly centered around 4.5 Hz(Ding et al., 2017; Varnet et al., 2017), part of or stimuli were recorded by speakers speaking as fast as possible while maintaining proper articulation.

All sentence materials were sourced from books and audiobooks that are freely available on LibriVox (https://librivox.org/) or Lit2Go (https://etc.usf.edu/lit2go/) (see Supplementary Table 4.1). The final stimulus set was drawn from five audiobooks and four books for which the recordings were performed by native speakers of North-American English at the Max Planck Institute for Empirical Aesthetics.

Recordings were performed in a sound-attenuated recording booth using MatLab R2017a (Version: 9.2.0.538062) on a Windows 7 Pro (64-bit) and a Neumann U87i studio microphone (https://en-de.neumann.com/u-87-ai) and digitized to a sampling rate of 44 kHz. Speakers spoke them at two speeds: at a normal, natural pace and as fast as possible (while maintaining proper articulation). A total of 788 individual sentences was acquired. In addition to the fast recordings, one batch of stimuli was recorded from a speaker who was instructed to speak as slowly as possible. Processing of all sound files (audio books and recordings) was performed using Praat (6.0.40)(Boersma and Weenik, 2020). Long pauses (> 300ms) were removed to avoid wrong estimates for syllables per second. After compression and expansion of the audio files, all final stimuli were matched for root mean square (RMS) amplitude (69 dB).

Experiment 2 – Stimulus lists for speech comprehension and preferred auditory tasks. From the recordings (N = 10 speakers), we created three stimulus lists for the speech comprehension (N = 240 sentences) and the auditory rate tasks (N = 49 individual sentences) by randomly drawing sentences (without replacement) from the total pool of 788 sentences. The range of original syllabic rates was 1.97-9.71 syllables/s. Within each set of stimulus lists (i.e. speech comprehension and auditory rate tasks) no sentences were repeated. Original speech recordings were time-compressed and time-expanded using the Pitch Synchronous Overlap and Add (PSOLA)(Moulines and Charpentier, 1990) algorithm.

The original sentences were compressed/expanded to the following syllabic rates: speech comprehension task: 5.00, 10.69, 12.48, 13.58, 14.38, 15.00; preferred auditory rate task: 3.00, 3.92, 4.83, 5.75, 6.67, 7.58, 8.50. Syllabic rate conditions were matched for sentence length (number of syllables), compression rate, position of target words, number of speakers (Supplementary Fig. 4.2). Maximal compression and expansion were constrained (compression: factor of 3; expansion: factor of 2). For the slowest condition (5 syllables/s) the variation in speakers is lower because only one speaker could be recorded (and no further recordings were possible due to lockdown).

Experiment 2 – Stimuli for digit span test. The digits 0-9 were synthesized using the Mac OSX text-to-speech application (voice Anna). Using Audacity, each digit was concatenated with silence such that digit and silence amounted to a duration of one second. Digit spans(Ottosson and Grahn, 2005) were created by concatenating the corresponding digits using Audacity so that the digits occurred at a rate of one digit per second.

Experimental tasks

Speech comprehension task. In experiment 1, participants performed an intelligibility task. On each trial (N = 70), a sentence was presented through headphones and participants verbally repeated all perceived words. Responses were recorded. Sentences were presented at various syllabic rates (8.2, 9.0, 9.8, 11.0, 12.1, 14.0, 16.4.) and grouped into blocks according to syllabic rates. While the syllabic rates were presented in the same (descending) manner to all participants, sentences were randomly assigned to each block/syllabic rate.

In experiment 2, participants performed a word-order task in which they listened to one sentence per trial (N = 240), followed by the presentation of two words from the sentence on screen. Participants indicated via button press which word they heard first. The sentences were presented at various syllabic rates: 5.00, 10.69, 12.48, 13.58, 14.38, 15.00. For each syllabic rate 40 sentences were presented, amounting to a total of 240 sentences. Trials were grouped into blocks of 20 sentences, allowing for self-paced breaks between blocks. We randomized sentence order and syllabic rates across participants. The maximal response time was 5000 ms. The inter-trial interval was jittered (uniform distribution between 1000 and 1500 ms). Prior to task begin, participants were familiarized with the task by 1) listening to one stimulus played at all possible frequencies (without task) and 2) performing three practice trials.

Auditory rate task (only Experiment 2). Participants performed a 2IFC task, in which a reference and a comparison stimulus were presented. Within a trial, reference and comparison stimulus were constructed from the same sentence – they only differed with regard to syllabic rate (3.00, 3.92, 4.83, 5.75, 6.67, 7.58, 8.50). Reference and comparison stimulus were separated by an inter-stimulus interval (uniform distribution between 500 and 1000 ms) and trials by a randomized inter-trial interval (uniform distribution between 1000 and 1500 ms). To ensure participants were engaged in the task, we added catch trials (one trial for each reference frequency, i.e., 7 trials). Catch stimuli were manipulated such that one syllable was repeated three times in both the reference and comparison stimuli. Participants were instructed to respond to catch trials by pressing “t”, instead of “x” or “m”, and were informed that poor performance would lead to an exclusion from the experiment. Prior to the main task, participants were familiarized with the task by performing three practice trials for the main task, as well as one practice trial for the catch trials.

Digit span test (only Experiment 2). The digits 0-9 were synthesized using the Mac OSX text-to-speech application (voice Anna). Using Audacity, each digit was concatenated with silence such that digit and silence amounted to a duration of one second. Digit spans (Ottosson and Grahn, 2005) were created by concatenating the corresponding digits using Audacity so that the digits occurred at a rate of one digit per second. The test comprised seven levels (3 to 9 digits) with two items at each level. The procedure started with the shortest digit spans (3 digits) and stopped as soon as participants failed to correctly repeat both spans belonging to one length or when the two longest digit spans (9 digits) were finished.

Procedure. Experiment 1 was conducted in the laboratory at Dundee University. Participants were seated in front of a computer in a quiet experimental room. Prior to experiment start, participants completed a demographic questionnaire. Upon giving informed consent, the speech comprehension task was completed first, followed by the speech production task. Experiment 2 entailed three sessions all of which were conducted online. Overall, the protocol contained 5 tasks (3 main tasks: comprehension task, auditory rate task, speech production task; 2 further tasks: SSS-test, digit span test) and a questionnaire. In each session, one of the three main tasks was completed. The order was randomized across participants. To assure participants wore headphones, a headphone screening test was performed in the be-ginning of each session (55). At the end of the last session, the SSS test, digit span test, and the questionnaire were conducted.

Exclusion criteria (Experiment 2). Only complete datasets (all three sessions completed) were considered for analysis. From participants with complete datasets, participants were excluded if (a) comprehension performance was 2 *SD* below mean performance at the baseline rate (5 syllables/s) in the

speech comprehension task, (b) detection of catch trials was below 75% in the auditory rate task, (c) participants did not whisper but speak normally/loudly during the SSS test, (d) poor audio recordings.

Computing word predictability and perplexity. To account for variation in sentence predictability in the comprehension task, we created a recurrent neural network (RNN) language model which assigns probabilities to sequences (for similar analysis see (ten Oever and Martin, 2021)). RNNs, especially when using long short-term memory (LSTM)(Hochreiter and Schmidhuber, 1997), are well suited to approximate language processing because they can incorporate past input (e.g., words) into the prediction of a current word(Mikolov, 2010), just as done during natural language processing in humans. The model was created to solve the task of predicting each word in the sentence based on the previous word or words. For the training, sentences were taken from a variety of books, to then predict word probabilities for all sentences in the stimulus materials used in our experiment. For this analysis we used Keras(Chollet and others, 2015) with a Tensorflow(Abadi et al., 2016) backend.

The training dataset was curated from 101 freely available classic books and underwent a cleaning procedure (removing punctuation, lower casing all words). We aimed at creating training materials as similar as possible to our stimuli. Therefore, books were selected from the same authors as the stimulus materials where possible, as well as different authors but similar topics and genres. Sentences were selected as training material only if their length matched the stimulus materials (8-25 words). The final data set comprised 252.377 sentences, constructed from 38.195 unique words. Prior to training, the curated data set was split into training and evaluation sets (75 to 25%).

For the RNN analysis, we worked off the scripts provided by ten Oever and Martin(ten Oever and Martin, 2021). Our RNN consisted of 3 layers: a pretrained embedding layer using Google's word2vec embeddings (https://code.google.com/archive/p/word2vec/), an LSTM layer with a tanh activation (300 hidden units) and a dense output layer (softmax activation). The model was trained using the following hyperparameters: batch size = 32, epochs = 100, Adam optimization with a learning rate of 0.001, and regularization to prevent overfitting (dropout of 0.2 for recurrent and output layers and L2 regularization of 0.001). Given the multi-class classification problem (which word out of 38.195 words is most likely to follow), we implemented a sparse categorical cross-entropy loss function.

Upon training, the trained model was used to evaluate and predict sentence predictability of all stimulus sentences (test set, N = 495) used in the comprehension task. From the RNN predictions of the test set, we derived two measures: target word probability (for both target words) and predictability of the whole sentence. To quantifying each sentences' predictability(Jurafsky and Martin, 2009; Chien and Ku, 2016), we derived one value per sentence from the single-word probabilities. This so-called perplexity is the most common intrinsic evaluation metric of language models(Mikolov, 2010; Merity et al., 2017; Fernandez and Downey, 2018). It is computed as the inverse of the mean probability of a sentence weighted by sentence length(Jurafsky and Martin, 2009), i.e. lower perplexity values equal higher sentence predictability.

Control analyses

Correlation of digit span and preferred auditory rate/spontaneous speech motor production rate:

- Digit span and speech motor production rate: rho = 0.064, p = 0.569
- Digit span and preferred auditory rate: rho = 0.040, p = 0.728

Correlation of preferred auditory rate and spontaneous speech motor production rate:

- rho = -0.052, p = .724

5 Study 3: Endogenous auditory and motor cortex brain rhythms predict individual speech tracking

5.1 Abstract

According to oscillatory approaches, slow endogenous brain rhythms in auditory cortex track acoustic amplitude modulations during speech comprehension. Temporal predictions from the motor system potentially enhance this tracking. However, direct evidence for the involvement of endogenous auditory and motor brain rhythms is lacking. Our Magnetoencephalography findings reveal that endogenous peak frequencies (i.e. rates) of individuals‘ resting-state theta rhythm in posterior superior temporal gyrus predict speech tracking during comprehension. For Heschl’s Gyrus we observed less rhythmically constrained processing. Crucially, speech tracking in auditory cortices was predicted by endogenous motor rates only in individuals with high auditory-motor synchronization behavior. Higher rates in the supplementary motor area were related to higher tracking, with opposite effects for inferior frontal gyrus. These findings align with participants’ behavioral data and support oscillatory accounts of auditory-motor interactions during speech perception. Finally, they highlight two distinct processing routes -an auditory-motor and an auditory working memory route- across individuals.

5.2 Introduction

Verbal communication exemplifies action-perception interactions in humans essential for everyday behavior. While the motor system's pivotal role in speech production is undisputed (Hickok and Poeppel, 2007; Tourville and Guenther, 2011; Blumstein and Baum, 2016), ongoing debates persist regarding its functional contributions to speech perception (Hickok et al., 2011). During action, the motor system engages in operations requiring precise timing. Similarly, in auditory (speech) perception, we may recruit these motor systems (Tourville and Guenther, 2011; Hertrich et al., 2016). One hypothesis posits that motor areas, including the supplementary motor area (SMA) and inferior frontal gyrus (IFG), provide temporal predictions about upcoming sensory events (Arnal and Giraud, 2012; Park et al., 2015; Morillon and Baillet, 2017; Haegens and Zion Golumbic, 2018; Rimmele et al., 2018). However, the motor systems' relevance may be circumstantial and more pronounced in effortful listening conditions (Du et al., 2014; Stokes et al., 2019).

Effortful listening conditions can be induced by manipulating the syllabic rate of speech, an ecologically relevant aspect of speech in everyday life. While individuals have shown to adapt to heavily accelerated speech in modern technology, rate variations can pose speech processing challenges under certain circumstances, for instance in aging (Penn et al., 2018) or in children with developmental speech disorders (Guiraud et al., 2018). Beyond studying the motor system's involvement, increasing the syllabic rate allows studying the temporal resolution of the auditory system during speech perception (i.e., faster speech requires more fine-grained timing). Oscillatory speech perception models propose that theta rhythms in auditory cortex synchronize to temporal fluctuations in the speech signal (i.e. the amplitude envelope) to segment it into syllable-sized chunks (Giraud and Poeppel, 2012; Gross et al., 2013). This brain-to-speech alignment is considered optimal around the nominal speech rate (~5 Hz), declining at higher syllabic rates (Ahissar et al., 2001), as speech comprehension decreases (for non-speech: Viemeister, 1979; Brungart et al., 2007; Teng et al., 2017; Lubinus et al., 2023). Endogenous auditory brain rhythms in the theta (and beta) range are proposed to reflect neural oscillations with preferred frequencies that constrain the optimal processing range, and thus timing granularity (Giraud et al., 2007; Giraud and Poeppel, 2012; ten Oever and Sack, 2015; Keitel and Gross, 2016; Kösem et al., 2018; Zoefel et al., 2020; Lubinus et al., 2021). The assumed connection between endogenous rhythms and speech processing, a core aspect of oscillatory theory, however, is rarely investigated (Rimmele and Keitel, 2023). Spectral fingerprinting of resting-state brain activity (Keitel and Gross, 2016) is a promising approach for accessing individual differences in the rates of endogenous rhythms.

Recent work has intriguingly proposed that not only auditory processing but also auditory-motor coupling during listening to syllable sequences has an optimal frequency range (~4.5 Hz, Assaneo and Poeppel, 2018; He et al., 2023). Computational modelling supports an oscillatory account of

auditory-motor coupling during speech perception, assuming oscillators with slightly higher preferred frequencies in the auditory than the motor system (Assaneo and Poeppel, 2018; Assaneo et al., 2021). In dynamical system theory, the *Arnold tongue* phenomenon describes how an oscillator's ability to track a stimulus depends on its preferred rate, the stimulation rate and intensity (Fröhlich & McCormick, 2010). Behaviorally, studies have shown a relation between preferred (spontaneous) motor production rates and the ability to synchronize to sound at different rates in music and sound sequences (McAuley et al., 2006; Kaya et al., 2023; Roman et al., 2023). However, the interplay between preferred rates of the auditory and motor system and their coupling strength remains unknown, and neural evidence is lacking. In a recent behavioral study (Lubinus et al., 2023), we demonstrated superior speech comprehension in individuals with higher auditory-motor synchronization and higher preferred motor rates (with a trend observed for auditory rates that did not survive multiple comparison control). The spontaneous speech synchronization (SSS) test was used to estimate auditory-motor cortex coupling, as its performance correlates with functional and structural neural coupling (Assaneo et al., 2019). To quantify the preferred endogenous motor rates, we used spontaneous rhythmic speech production, similar to the spontaneous tapping measure typically used in the field of dynamic attending (Repp, 2005; McAuley et al., 2006; Repp and Su, 2013; Zamm et al., 2016; Zalta et al., 2020; Kaya and Henry, 2022).

This study aims to better understand the relationship between rate-restricted auditory-motor coupling, endogenous rates, and naturalistic speech processing in the brain. Behaviorally, we expected to replicate our previous work (Lubinus et al., 2023), with higher spontaneous motor production (and preferred auditory) rates and higher auditory-motor synchronization predicting better speech comprehension. Based on these findings, our neurophysiological hypothesis posited that higher endogenous (resting-state) theta rates in auditory and motor cortices, along with stronger auditory-motor coupling, predict enhanced speech tracking during comprehension. Both speech tracking (Ahissar et al., 2001) and auditory-motor coupling (Assaneo and Poeppel, 2018) were expected to be reduced at higher syllabic rates, although we had no strong expectations regarding the tracking because of conflicting results (Pefkou et al., 2017; Verschueren et al., 2022).

5.3 Methods

Our Magnetoencephalography (MEG) and behavioral experiment entailed three separate sessions: a behavioral, an MEG with behavior, and a structural Magnetic Resonance Imaging (MRI) session. The study was approved by the by the local ethics committee of the University Hospital of the Goethe-University Frankfurt (number: 2021-509) in accordance with the Declaration of Helsinki.

Participants. A total of 60 participants (age: M = 26.9, SD = 5.4; 32 female) completed the study. As assessed by self-report, participants had no history of neurological or psychiatric diseases and

had normal hearing, as well as normal (or corrected-to-normal) vision. All participants were native speakers of German and right-handed. Prior to each session, participants provided written informed consent. At the end of the final session, participants received monetary compensation.

Stimuli. In two experimental tasks (speech comprehension task and preferred rate task) participants listened to naturalistic sentences. The sentences, from German books (N_sentences = 306; N_speaker: 8; source: zeno.org) and audiobooks (N_sentences = 138; N_speaker: 3; source: Librivox.org) were sourced from zeno.org and librivox.org, respectively (Supplementary Table 5.1). Sentences from books were recorded by three native speakers of German at the MPIEA. Recordings, conducted in a sound-attenuated booth using MatLab R2017a on a Windows 7 Pro (64-bit) and a Neumann U87i studio microphone, were digitized at a 44 kHz sampling rate. In total, speakers delivered the sentences at their normal, slowest, and maximal speaking rate, while prioritizing proper articulation over speed. All sound files, including audio books, underwent processing using Praat (6.0.40). Long pauses (>300ms) were removed to prevent inaccurate syllable per second estimates.

Three stimulus lists were generated from the sentence materials for both the speech comprehension task (300 sentences) and the auditory rate task (132 sentences). Time compression or expansion was applied to all sentences to create different syllabic rate conditions. Sentences were randomly selected (without replacement) from the total pool of sentences based on two main criteria: the degree of compression required to obtain the desired syllabic rate (compression factor < 3) and the duration of the sentence after compression (> 1s; only applied to comprehension task). No sentence repetitions occurred within each stimulus set across tasks. Time compression or expansion was performed using the Pitch Synchronous Overlap and Add (PSOLA) algorithm in Praat, and all stimuli were standardized to a root mean square (RMS) amplitude at 69 dB (see Fig 5.1).

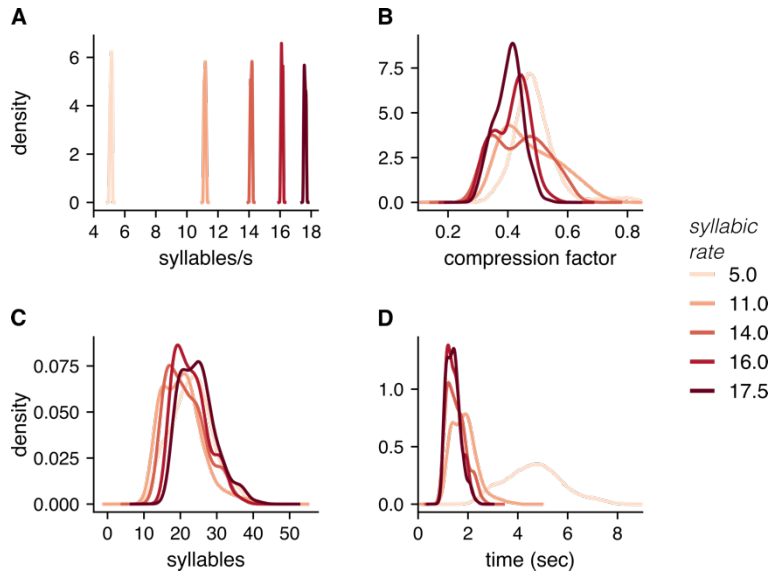


Figure 5.1. Stimulus parameters for speech comprehension task. A. Sentences were presented at five syllabic rates: 5.0, 11.0, 14.0, 16.0, 17.5 syllables/s. **B.** We minimize compression differences between rate conditions, by selecting "raw" sentences (i.e. before compression) with maximum variation in syllabic rates. **C.** Balancing information density across rate conditions was prioritized by selecting sentences with overlapping distributions of sentence length (i.e. number of syllables). **D.** However, this came at the cost of not fully equalizing duration. While the four faster conditions were similar in duration, sentences in the slowest condition (5 syllables/s) were longer.

Experimental tasks

Intelligibility task. To measure speech comprehension, participants performed an intelligibility task (Fig 5.4A). On each trial, participants listened to a sentence via headphones and verbally repeated it as accurately as possible. All responses were recorded, and participants stopped the recording via button press (left or right index finger), initiating the interstimulus interval. Sentences were presented at five syllabic rates (5.00, 11.00, 14.00, 16.00, 17.50 syllables/s) with 60 different sentences each, totaling 300 trials. Trials with different syllabic rates were pseudorandomized within blocks of 30 trials, with self-paced breaks between blocks.

Spontaneous speech synchronization (SSS) test. To assess auditory-motor synchronization, participants performed the spontaneous speech synchronization (SSS) test (Fig 5.2A; for detailed description see the General Introduction, Study 2 and Assaneo et al., 2019). In two trials, participants continuously whispered a syllable (/te) for 80 seconds and aimed to synchronize their own motor output to a stream of syllables. Their whispering was recorded. The auditory stimulus progressively increased in rate from 4.3 to 4.7 syllables/s (increments of 0.1 syllables/s) every 60 syllables. Participants' syllable production was masked by the simultaneously presented auditory syllable train.

Speech production task. The individual spontaneous speech motor production rate was measured by asking participants to freely speak "as they would naturally" (Fig 5.2B). Participants were

prompted with 6 questions/statements to facilitate continuous speech production (6 trials; own life, preferences, people, culture/traditions, society/politics, general knowledge, see Supplementary Table 5.2)(Alexandrou et al., 2016). Participants read the question/statement and initiated the speaking period (30 s) via button-press. White noise was presented via headphones to minimize auditory feedback (see Study 2 for details). Breaks in between trials were self-paced.

Auditory rate task. The preferred auditory rate was assessed using a two-interval forced choice (2IFC) task (Fig 5.2C). On each trial, we presented two versions of the same sentence, randomly ordered, differing only in syllabic rate. After stimulus presentation, participants indicated which of the two stimuli they preferred via button press. Stimuli were presented at twelve syllabic rates from 3.00 to 8.50 syllables/s (in steps of 0.5 syllables/s).

Digit span test. Working memory capacity was quantified using the forward and backward (Richardson, 2007) digit span test. Participants listened to digit spans and typed in their responses after listening (Olsthoorn et al., 2014). The test comprised seven levels, ranging from two/three to nine digits with two items at each level. The procedure started with the shortest digit spans (forward: three digits, backward: two digits) and stopped as soon as participants failed to repeat both spans of the same length or when the two longest digit spans were reached.

Procedure – Behavioral session. Participants were seated in a sound-attenuated experimental booth equipped with a Fujitsu CELSIUS M740B PC. Stimuli were presented using Psychtoolbox (Brainard, 1997) in Matlab (9.7.0.1471314, R2019b) and insert earplugs (ER3C Tubal Insert Earphones; Etymotic Research). Participants' speech and whisper was recorded using a gooseneck microphone (MX418 microflex gooseneck microphone, Shure) and behavioral responses were collected using a standard keyboard. Stimuli were presented at ~70 dB in the preferred auditory rate and digit span tasks. For the spontaneous speech production task and the SSS-test, loudness could be adjusted using a volume control knob on the sound card. Participants were instructed to increase the volume until their own speech or whisper became inaudible, while still being comfortable. The rationale behind this procedure is to isolate motor production by suppressing auditory feedback.

All participants started with the spontaneous speech motor production rate task to avoid priming effects. Next, participants performed either the SSS-test or the preferred auditory rate task, the order was randomized across participants. Finally, all participants finished with the digit span task, followed by the questionnaire (demographics and musicality).

Procedure – MEG session. After application of EOG and ECG channels, participants were seated in the MEG. The stimuli were delivered binaurally using insert earplugs (EARTONE Gold 3A insert earphones; Ulrich Keller Medizin-Technik) and the Matlab (R2017a) software with the Psychtoolbox (3.0.14; Brainard, 1997) extension on a Fujitsu-Technology CELSIUS R940power PC. Participants'

responses were recorded using a button box (Current Designs Package 932). Fiducials were attached to participants' nasion and the preauricular points to continuously measure head position.

First, participants performed the intelligibility task. The task was split into 10 blocks (~5 minutes each). Second, participants completed an auditory localizer task in which they passively listened to a sequence of sounds (pure tones: 0.4 s tone duration; 250 Hz and 1000 Hz, 100 repetitions, jittered intertrial interval 0.5–1.5 s; ~5 min). Third, participants performed a motor localizer task by (~12 min) during which they repeatedly articulated syllables without vocalizing. On each trial, one out of three syllables (/pa/, /ta/, /sa/) was articulated for 3 s and each syllable was repeated in 50 trials. Finally, we recorded resting state activity (~5 min 30 sec). Throughout the experiment, participants were instructed to hold their gaze at a fixation cross (except during instructions). In total, the experiment was of a duration of roughly 180 minutes, consisting of 120 minutes recording time and 60 minutes of preparation time and scanning pauses.

MEG data were acquired at a sampling rate of 1200 Hz using a 275-channel whole-head MEG machine (Omega 2005, CTF Systems Inc.) in a magnetically-shielded room. During scanning, online denoising (higher-order gradiometer balancing) and online low-pass filtering (cut-off: 300 Hz) were applied. Furthermore, participants' head position was measured continuously using fiducials and the Fieldtrip toolbox (Oostenveld et al., 2011, version 20220617), allowing for position adjustment between blocks as well as continuous head movement correction at the analysis stage.

Procedure – MRI session. Individual structural MRI scans (standard 1 mm T1-weighted MPRAGE) were obtained using a 3 Tesla scanner (Siemens Magnetom Trio, Siemens, Erlangen, Germany). Vitamin E capsules were used to mark anatomical landmarks (nasion, left and right preauricular points) to align MRI and MEG data for source reconstruction.

Behavioral Analysis

Speech comprehension. Participant's responses were transcribed manually. To quantify comprehension accuracy, we compared the responses to the original sentences using a sequence matcher algorithm (Python built-in sequence matcher) which quantifies the similarity between two sequences by (order of) item, i.e. letter. The output of the sequence matcher was a percentage for each sentence.

Preferred auditory rate. From the 2IFC task, we derived the preferred frequencies for each trial. This distribution of preferred frequencies across trials was subjected to a curve fitting procedure, fitting a Gaussian function. From the fitted Gaussian, the peak parameter was extracted and defined as individual preferred auditory rate.

Spontaneous speech motor production rate. The spontaneous speech motor production rate was quantified as articulation rate (i.e. number of produced syllables divided by trial duration,

excluding pauses > 300ms). To segment the continuous speech recordings into syllables, syllable nuclei were detected automatically using Praat (de Jong and Wempe, 2009). The spontaneous speech motor production rate was computed for each trial separately and then averaged across trials.

Auditory-motor synchronization. The data from the SSS-test were analyzed according to the protocol of Lizcano-Cortés et al. (2022). First, data quality was ensured: participants whispered (i.e. no vocal cord activation) and audio files were not corrupted (i.e. noise interference). For the analysis, we applied the scripts provided by the authors (Lizcano-Cortés et al., 2022). Auditory-motor synchronization was measured as the phase-locking value (PLV) between the speech envelope of the produced motor signal and the cochlear envelope of the syllable stimulus. For the syllable train, the cochlear envelope was extracted using the Chimera toolbox (auditory channels: 180–7,246 Hz)(Smith et al., 2002). For the produced motor signal, the amplitude envelope was estimated using the Hilbert transform (hilbert.m function in Matlab). Both envelopes were downsampled to 100 Hz and bandpass filtered between 3.5 and 5.5 Hz. From the filtered signal, we extracted the phase (using hilbert.m and phase.m functions in Matlab), computed the PLV for time windows of 5s (overlap 2s) and averaged across time windows. This procedure was completed for both experimental runs separately. To test for consistency between runs, a linear regression was fitted to the data (independent variable: PLV of run 1, dependent variable: PLV of run 2). Participants with PLV pairs outside the 95th confidence interval were excluded from further analysis. For the remaining participants, PLVs from both runs were averaged. The mean PLVs were subjected to a Gaussian mixture model to partition participants into high and low synchronizers.

MEG analysis – speech comprehension task

Speech envelope extraction. For the speech signals presented during the comprehension task in the MEG, we extracted the cochlear envelope using the Chimera toolbox (auditory channels: 180–7,246 Hz)(Smith et al., 2002). Specifically, the spectrograms were computed for multiple frequency bands (auditory channels: 180–7246 Hz) and the absolute values then averaged to obtain the broadband speech envelope. Broadband envelopes were downsampled to 100 Hz to match MEG signals.

Preprocessing. During preprocessing, the data were bandpass filtered (0.5–160 Hz, Butterworth filter; filter order 4) and line-noise was removed (49.5-50.5, 99.5-100.5, 149.5-150.5 Hz, two-pass; filter order 4). We applied a semi-automatic artifact rejection procedure to detect jump, muscle, and threshold artifacts. For jump and muscle artifacts, data were filtered to optimize artifact detection (muscle: 110-140 Hz, jump: median filter) and then z-transformed per sensor and time point. Trials were rejected if they surpassed a priori defined thresholds (jump: z=45, muscle: z=15). For the threshold artifacts, trials were rejected if the range (min-max difference) of activity at any channel

surpassed an a priori defined threshold (threshold = 0.75e−5). The data were downsampled to 500 Hz and epoched (-1000ms to +100ms after trial offset), resulting in epochs of variable length. Using the continuous fiducial measures, trials containing head movements larger than 4mm were identified and rejected. Data from the separate recording blocks were concatenated into one large file and sensors with high z-values ($z > 2$) were rejected. Finally, eye blink, eye movement and heartbeat artifacts were corrected using independent component analysis (infomax algorithm; Makeig et al., 1996). Data were further downsampled to 100 Hz for computational efficiency.

Source localization. Using the three anatomical landmarks (nasion, left, and right preauricular points), individual T1-weighted MRI images were co-registered to the MEG coordinate system in a semiautomatic procedure. T1-weighted MRIs were segmented into three tissues (white matter, grey matter, cerebrospinal fluid) to construct single-shell volume conduction models (headmodel) (Nolte et al., 2008) and warped into MNI space to compute individual grids by inverse warping the template grid (grid resolution: 5mm) onto individual anatomical scans. Using the individual grids and volume conduction models, we computed individual forward models to reconstruct source activity.

For source reconstruction, we used a Linearly Constrained Minimum Variance (LCMV) Beamformer (array-unit LCMV)(Westner et al., 2022). Firstly, we computed the covariance matrix across all trials and created a common filter for all conditions using the individual forward models. The lambda regularization parameter was set to 1% and time-series were extracted for all three dipole orientations per voxel. Secondly, trial data was source-localized by projecting it through (i.e. multiplying it with) the common filter for each condition separately. Computing the common filter across all conditions—instead of separate filters for each condition—ensures that differences between conditions reflect differences in source activity, instead of differences in the spatial filters itself.

Regions of interest. For the speech tracking analysis, regions of interest (ROI, as defined by the Automated Anatomical Labeling (AAL atlas (Rolls et al., 2020)) included primary and non-primary auditory cortex, comprising Heschl's gyrus (HG) and posterior superior temporal gyrus (pSTG), respectively. Both areas are crucial for speech perception, with the pSTG showing increasing responsiveness to complex acoustic speech features compared to HG, as supported by neurocognitive models and empirical studies (Binder, 2000; Hickok and Poeppel, 2004; DeWitt and Rauschecker, 2012; Hamilton et al., 2021). Notably, speech tracking has been consistently observed in both HG and STG across multiple studies (Nourski et al., 2009; Giraud and Poeppel, 2012; Keitel et al., 2017, 2018; Davidesco et al., 2018; Rimmele et al., 2021; Chalas et al., 2023).

For auditory-motor coupling, in addition to HG and pSTG, the ROIs also included two motor areas that are well-established in the speech motor network. Inferior frontal gyrus (IFG), a key component of the core language network (Hickok and Poeppel, 2004) which might contribute to the auditory-

motor coupling during speech perception (Park et al., 2015; Assaneo et al., 2019), exhibits direct structural connections with the auditory cortex (Friederici, 2015; Hertrich et al., 2016) and is involved in various speech processing subroutines such as phonological encoding (Zatorre et al., 1992) or semantic and sentential information integration (Obleser and Kotz, 2010). Importantly, IFG synchronizes to speech envelopes (Keitel et al., 2018; Assaneo et al., 2019) and shows differing degrees of synchronization depending on individual auditory-motor speech coupling (Assaneo et al., 2019). Interestingly, the auditory-motor coupling of IFG has been proposed to reflect neural oscillatory activity (Poeppel and Assaneo, 2020). The supplementary motor area (SMA), although not traditionally considered a language area, plays a central role in speech motor control (Tourville and Guenther, 2011) and, more recently, has been recognized to contribute to higher-level processes during speech processing (Hertrich et al., 2016). Together with the cerebellum and basal ganglia, SMA processes temporal structure (Coull et al., 2011) and sensory temporal information more generally (Macar et al., 2002; Lewis and Miall, 2003b; Schwartze et al., 2012b). Notably, SMA is particularly implicated in adverse listening situations (Adank, 2012), possibly through top-down mechanisms, predicting upcoming sensory events (Kotz et al., 2009; Schwartze et al., 2012b).

We extracted the corresponding parcels for all ROIs (13+14: IFG tri.; 19+20: SMA, 79+80: HG, 81+82: STG) from the source-localized data using the AAL atlas. To obtain a pSTG parcel, we split the original STG parcel into an anterior and posterior part using a median-split. To further reduce the dimensionality of the data, principal components (PC) were computed across voxels within each parcel. To this end, for each trial, all voxel time-series within a parcel were stacked (three time-series, i.e. for each dipole orientation, per voxel) and PCs and their explained variance were extracted. Based on previous work (Chalas et al., 2022), the further analysis only included the first three PCs.

Mutual information – speech tracking. Speech tracking was estimated using information theory (Shannon, 1948). Specifically, we computed gaussian-copula mutual information (GCMI; Ince et al., 2017) to assess statistical dependencies between speech envelopes and source-localized neural activity in auditory and motor ROIs. GCMI capitalizes on the concept of gaussian copulas which is a statistical description of the relationship of two variables regardless of their respective marginal distributions. Mutual information (MI) can be computed as the negative entropy of a statistical copula. This facilitates estimation of MI because 1) no assumptions of the marginal distributions of the variables are needed and 2) the computation is efficient as a Gaussian parametric estimation can be employed. A further advantage of the GCMI approach is that it lends itself naturally to dealing with multivariate data (Ince et al., 2017). This allowed us to use both amplitude and phase information of speech envelopes and source-localized activity for multiple frequency bands.

For the estimation of GCMI, we applied a continuous wavelet transform for 28 frequencies (3.5 to 20 Hz) to the source-localized PCs for each trial using *cwtfilterbank.m* in Matlab. To avoid onset effects, filtered PCs were epoched again from 200ms relative to stimulus onset until trial end. From the complex signals, we first extracted the phase information (real and imaginary parts) and then copula-normalized the parts separately. Finally, the normalized components were concatenated within each condition. This yielded two frequency-by-samples vectors per condition, i.e. source time-series. Importantly, the same procedure (continuous wavelet transform, epoching, extraction of phase information, concatenation) was applied to the speech envelope data with the only difference that each trial only consisted of one time-series per trial (no PCs). Analogous to the source time-series, this resulted in two speech time-series per condition (real and imaginary parts).

In a multivariate analysis, we estimated GCMI between 2 speech time-series and 6 source time-series for each parcel and frequency. To account for stimulus-brain lags, we applied our GCMI estimation at various positive delays (0ms to 300ms, 10ms steps).

Mutual information – Auditory-motor coupling. To examine the hypothesis that audio-motor coupling strength predicts speech tracking (and comprehension), we estimated audio-motor coupling using GCMI. This analysis was identical to the speech tracking analysis above with respect to frequency analysis and GCMI parameters. The important difference lies in the selection of time series: here GCMI was computed between MEG time courses from two different brain areas. Specifically, we computed GCMI between HG (left and right) and IFG pars triangularis (left and right).

We tested whether auditory-motor coupling was higher than chance-level within each condition by comparing the true MI values against the 99th percentile of surrogate data. To create surrogate data, we segmented the concatenated trials of each condition into segments, shuffled the segments for each (N=500) iteration and estimated GCMI between source time-series and shuffled speech time-series. Importantly, segment length differed between conditions and was set at 80% of the cycle length of the stimulation frequency of each condition (i.e. 5 Hz = 160.0 ms, 11 Hz = 72.72 ms, 14 Hz = 57.1 ms, 16 Hz = 50 ms, 17.5 Hz = 45.7 ms). We used the surrogate distribution to create z-transformed MI values, thus subtracting true MI values by the mean surrogate distribution and dividing the result by the surrogate's standard deviation. The normalization was performed within participants for each parcel, condition, and frequency. MI values varied systematically as a function of delay and condition; we extracted MI for each condition at the delay that maximized the MI values.

MEG analysis – resting state

Preprocessing. For the resting state data, preprocessing was similar to the main task but adjusted in a few aspects to match the preprocessing procedure employed in Study 1 (Lubinus et al., 2021).

Specifically, all parameters for artifact rejection were identical to the main task of the current Study. The data were detrended, downsampled to 250 Hz and epoched into 0.8 s trials which resulted in an average of M=380 trials (sd=59.8, min=121, max=412). Sensors were rejected based on the neighborhood-ratio using the *hcp_qc_neighstddratio.m* function (Human Connectome Project, WU-Minn Consortium), which is defined by the sensors noise level relative to the noise level of neighboring sensors (Sensor SD—NeighborSD/NeighborSD). Sensors were rejected if the ratio exceeded a value of 0.5. For trial rejection based on head movements and the independent component analysis, the same procedures as in the main data of the current study were implemented.

Common filter for source projection. We used the same headmodels as in the main analysis. Forward models were obtained using the same procedure as for the main task, but with a different grid (resolution = 10mm; see Study 1 and (Keitel and Gross, 2016)). For source projection, we used an LCMV Beamformer. First, we computed the covariance matrix across all trials for each individual. Second, the common filter was generated using the previously computed forward models. The lambda regularization parameter was set to 7% and the optimal dipole orientation for each voxel was extracted (single value decomposition). The beamformer coefficients were obtained in preparation for the later source projection of Fourier spectra.

Spectral analysis and source projection. Complex Fourier spectra were computed in sensor space for individual epochs (0.8s) at 1-120 Hz (logarithmic frequency resolution) and each participant. Each epoch was zero-padded to a duration of 2s before applying a multitaper spectral analysis (DPSS multitaper, 2 tapers, +/- 2 Hz spectral smoothing). The complex spectra were projected into source space by multiplication with the previously obtained LCMV coefficients, summing across tapers for each trial. Source-localized spectra were ratio-normalized (the spectral power for each voxel and segment was divided by the average power across all voxels and segments) and subtracted by 1, yielding values above/below zero. To reduce dimensionality, we parcellated the data into 115 parcels using the AAL atlas by averaging power spectra across voxels within each parcel.

Clustering spectral data. For each participant and parcel the source-localized single-trial Fourier spectra were subjected to the k-means algorithm. The k-means algorithm treated each trial's spectrum as a 42-dimensional space and grouped all trials into *k* mutually exclusive spectra according to spectral (dis)similarities. These spectral clusters represent the spectral modes the brain area engages in over time. We set *k* to 10 (see: Keitel and Gross, 2016), used the Cosine distance metric, and initiated the k-means algorithm 10 times (max. 100 iterations). To quantify the proportion of trial-spectra –as determined by the k-means algorithm– belonging to each cluster, k-means clusters were subjected to a Gaussian Mixture (GM) Model algorithm.

Prior to the group-level analysis, we determined the optimal number of clusters across participants for each parcel. We evaluated the Silhouette criterion (1000 iterations) for solutions between 1-15 clusters and selected the number of clusters characterized by the highest Silhouette value. To obtain group-level spectra for each parcel, single-subject clusters were stacked and then fed into the k-means algorithm. The results were again subjected to the GM Model algorithm.

Selection of individual source-localized theta peaks. We extracted individual peak frequencies from the theta clusters for HG, pSTG, SMA, and IFG (see above for motivation of ROIs). Spectral clusters were defined as theta cluster if their peak frequency was within the range of 4 to 8 Hz. To this end, first, we extracted single-subject spectral clusters that contributed to the theta group-level cluster in question. From the single-subject clusters, we computed the peak frequency (henceforth *endogenous theta frequency of area X*). If multiple clusters from one participant contributed to the group cluster, the cluster with the highest peak amplitude was selected. On the group-level, HG was characterized by two theta clusters. We extracted the individual clusters contributing to the lower group cluster because it was contributed to by N=56 participants, compared to N=44 for the higher cluster.

Mixed model analysis

To assess the hypothesized links between speech comprehension and speech tracking and the corresponding predictor variables (see Hypotheses), we computed Generalized Linear Mixed Models (GLMM). Models were computed using R (version 4.1.3, 2022-03-10) set up in Rstudio (version 2022.2.1.461). For the GLMMs, we used the R package *lme4* (version 1.1-35.1). Tables and effects are visualized using *sjPlot* (version 2.8.15) and *ggplot2* (version 3.4.4). For all models, continuous predictor variables were z-transformed.

First, for the behavioral GLMM, trial-wise speech comprehension performance (% words correct) was regressed against syllabic rate, preferred auditory rate, spontaneous speech motor production rate, PLV, working memory score, compression factor, sentence length (in syllables), and stimulus order. We were particularly interested in the interaction effect of PLV * preferred auditory rate * spontaneous speech motor production rate. Thus, in addition to main effects, the model contained a three-way interaction. As random effects, we introduced a by-participant random slope for syllabic rate and a random intercept for trial-ID. Overall, the model explained 47.8% of the variance. To reduce the complexity, we computed the model separately for high and low synchronizers, which allowed us to reduce the three-way interaction to a two-way interaction (no PLV term). These high and low synchronizer models explained 47.1% and 48.1% of the variance, respectively.

```
Speech comprehension  ~  poly(syllabic rate, 2) +
            preferred auditory rate * motor production rate * PLV +
            working memory + compression + stimulus length +
            stimulus order +
            (0+freq | sub-ID) + (1 | trial-ID)
```

Second, we designed a GLMM that computed the relationship between speech tracking and the neural predictor variables. All neural variables (speech tracking, endogenous rates, auditory-motor coupling) were averaged across hemispheres. We ran two separate models, one focused on HG ("HG model") and one on pSTG ("pSTG model"). In the HG model, speech tracking in HG was regressed against syllabic rate, auditory-motor coupling (IFG-to-HG and SMA-to-HG) and the endogenous theta frequencies of HG, IFG, and SMA. The model included two three-way interactions between auditory-motor coupling and the corresponding endogenous theta frequencies. The pSTG model was identical except that all HG variables were exchanged for pSTG variables (endogenous rate and coupling). Overall, the HG model explained 24.8% of the variance, the pSTG model 35.6%. Analog to the behavioral GLMM, we additionally computed the models separately for high and low synchronizers.

```
HG Speech tracking  ~  poly(syllabic rate, 2) +
            HG frequency * SMA frequency * SMA-HG coupling +
            HG frequency * IFG frequency * IFG-HG coupling +
            (0+freq | sub-ID)
```

Third, to quantify the link between speech comprehension and speech tracking, we computed a GLMM predicting speech comprehension by speech tracking and syllabic rate. A by-participant random slope for syllabic rate was introduced.

```
Speech comprehension  ~  poly(syllabic rate, 2) + speech tracking +
            (0+freq | sub-ID)
```

5.4 Results

Replication: Speech comprehension performance is predicted by behavioral auditory and motor parameters

In the behavioral session, participants completed behavioral tasks assessing their spontaneous speech motor production rate (henceforth *motor production rate*), preferred auditory rate, and audio-motor synchronization (Fig. 5.2A-C). Participants exhibited an average motor production rate of M = 4.24 syllables/s (SD = 0.51, Fig. 5.2E). Their preferred auditory rate was about one syllable/s faster than the motor rate and averaged at M = 5.61 syllables/s (SD = 0.78, Fig. 5.2F). Consistent

with previous work (Assaneo et al., 2019, 2021; Lubinus et al., 2023), auditory-motor synchronization, measured using the SSS-test, was bimodally distributed (Fig. 5.2D). In this sample, N=28 participants were classified as high synchronizers and N=20 as low synchronizers (N=9 participants were not clearly classified, i.e. probability of being a HIGH was ~50%, and thus excluded from further analyses).

Speech comprehension was measured using an intelligibility task (Fig. 5.4A,C). Trial-based comprehension accuracy (i.e. % words correct) was regressed against several variables (see Fig. 5.3A,B and Supplementary Table 5.3). As expected, we observed a main effect of syllabic rate with decreased comprehension for higher syllabic rates (linear: β = -70.81, SE = 1.78, $p < .001$; quadratic: β = -24.36, SE = 0.98, $p < .001$). Replicating our previous findings (Lubinus et al., 2023), stronger auditory-motor synchronization (β = 0.03, SE = 0.ß2, p = .023) and faster motor production rates (β = 0.03, SE = 0.02, p = .044, not FDR-corrected) predicted better speech comprehension. In contrast to our previous findings, the motor production rate effect did not survive the control for multiple comparisons (p = .056). In addition, faster preferred auditory rates predicted better speech comprehension (β = 0.09, SE = 0.02, $p < .001$). These effects were refined by two-way interaction effects of PLV*motor production rate (β = 0.07, SE = 0.02, $p < .001$) and motor production rate*preferred auditory rate (β = -0.06, SE = 0.02, p = .001), and a three-way interaction of motor production rate*preferred auditory rate*PLV (β = -0.06, SE = 0.02, $p < .001$). The positive effect of the motor production rate was particularly observed in high synchronizers (PLV*motor production rate), whereas the positive effect of the preferred auditory rate was stronger in individuals with low motor production rates. The 3-way interaction suggests that the interplay of auditory and motor rates is particularly evident in individuals with stronger auditory-motor synchronization. Several control variables —better working memory (β = 0.07, SE = 0.01, $p < .001$), shorter sentences (β = -0.11, SE = 0.02, $p < .001$), and occurrence late in the experiment (β = 0.05, SE = 0.00, $p < .001$)— facilitated speech comprehension.

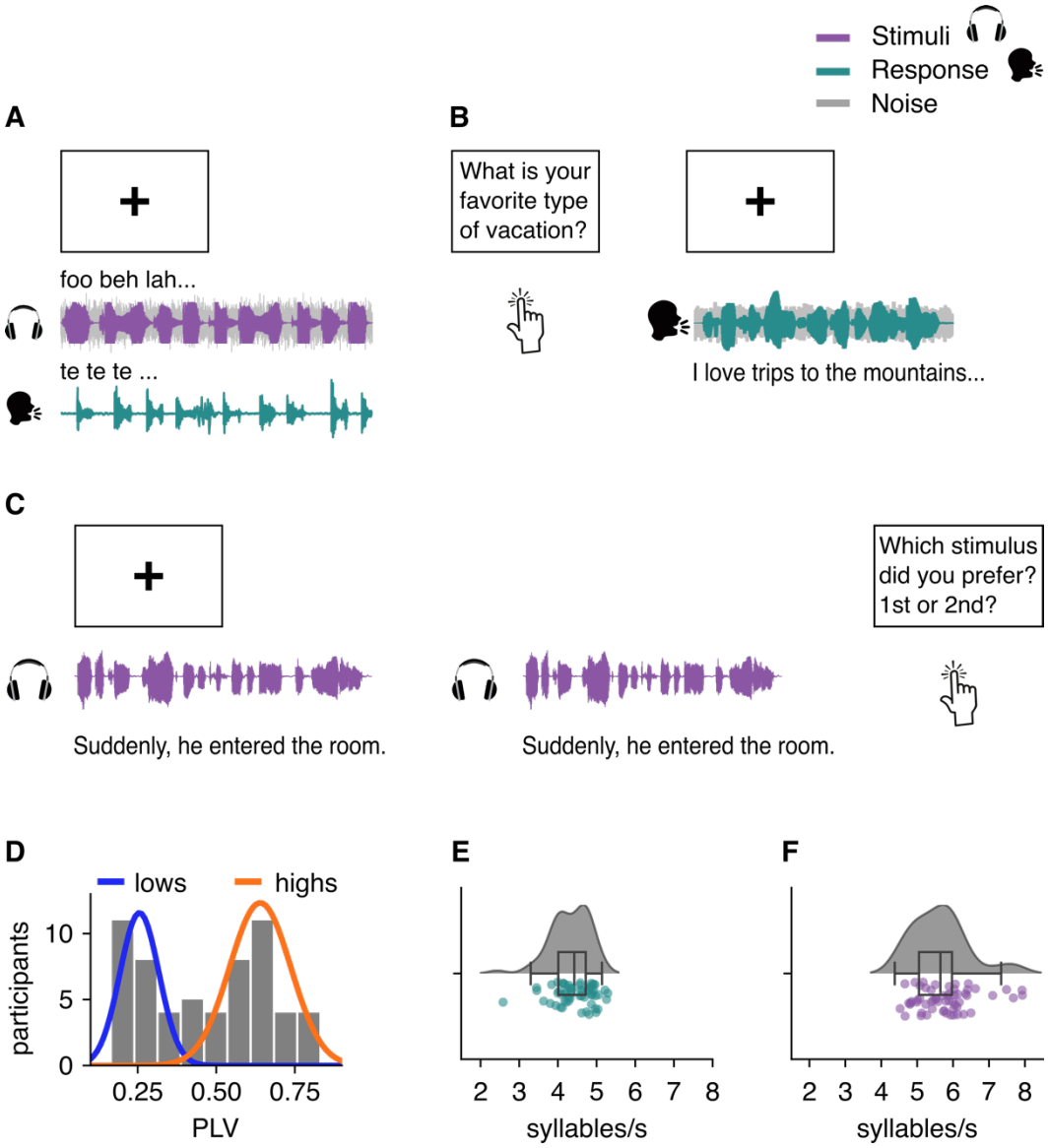


Figure 5.2. Experimental paradigms and descriptive data from the behavioral session. A. We measured auditory-motor synchronization using the (explicit) SSS-test (Lizcano-Cortés et al., 2022), wherein participants whisper the syllable /te while listening to a syllable train. They are instructed to synchronize their motor output to the auditory input. **B.** To measure the motor rate, we conducted a speech motor production task. Upon being prompted with a question/statement, participants spoke freely for 30s. **C.** In the preferred auditory rate task, participants indicated which speech rates they liked better. To this end, two versions of a sentence, differing in their syllabic rates, were presented at each trial and participants chose the sentences (i.e. syllabic rate) they preferred. **D.** Histogram illustrates audio-motor synchronization (measured as PLV). Colored lines represent fitted normal distribution, obtained by a Gaussian mixture model. **E, F.** Density and dot plots visualize the motor production rate (**E**) and preferred auditory rate (**F**) of participants.

Because it has been hypothesized that high and low synchronizers behave fundamentally different (Assaneo et al., 2021)and to reduce the complexity of the model, we computed separate models for the two groups (see Fig. 5.3C and Supplementary Table 5.4). Both groups showed a main effect of syllabic rate (LOW: (linear: β = -45.67, SE = 2.02, $p < .001$; quadratic: β = -15.82, SE = 0.92, $p < .001$); HIGH: (linear: β = -53.61, SE = 1.69, $p < .001$; quadratic: β = -19.73, SE = 0.98, $p < .001$))

and a main effect of preferred auditory rate (LOW: (β = 0.08, SE = 0.02, p < .001), HIGH: (β = 0.05, SE = 0.02, p = .006). For low synchronizers, we further observed a main effect of working memory (β = 0.13, SE = 0.02, p < .001). In high synchronizers, speech perception was additionally predicted by the motor production rate (β = 0.08, SE = 0.02, p < .001) and a motor production rate*preferred auditory rate interaction effect (β = -0.04, SE = 0.02, p = .009). While we acknowledge that the statistical power is lower in the model of low synchronizers due to less participants (N=20 vs. N=27), these results suggest—together with the three-way interaction of the main model—that motor production rate effects and the interaction of the auditory and motor rates are mainly present in individuals with high auditory-motor synchronization.

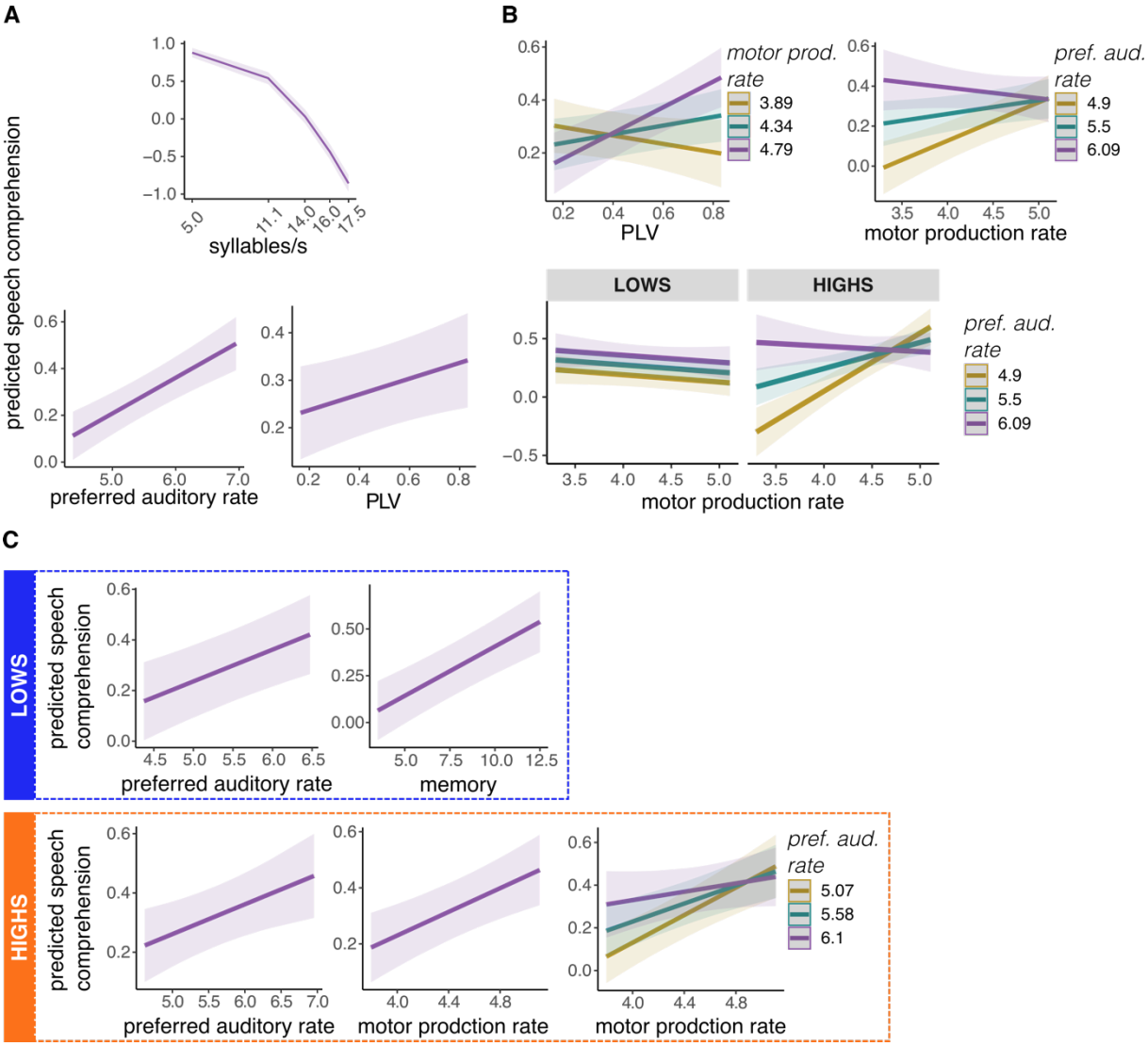


Figure 5.3. Different effects of auditory and motor parameters on speech comprehension for high vs. low synchronizers. A. The full GLMM revealed a negative main effect of syllabic rate and positive main effects of the preferred auditory rate and PLV. **B.** We further observed interaction effects between preferred auditory rate and PLV, preferred auditory rate and motor production rate, and a three-way interaction of preferred auditory rate, motor production rate and PLV. **C.** Results from GLMMs computed separately for low (blue) and high (orange) synchronizers. In low synchronizers, speech comprehension

was predicted by the preferred auditory rate. In contrast, for high synchronizers preferred auditory rate, motor production rate, and their interaction predicted speech comprehension.

Auditory cortex tracking and auditory-motor coupling across syllabic rates

We analyzed whether speech tracking was predicted by auditory-motor coupling and the endogenous theta frequencies of auditory (HG, pSTG) and motor (IFG, SMA) brain areas. To quantify speech tracking (see Fig. 5.4B for pipeline), we computed Gaussian-Copula Mutual Information (GCMI) between neural activity in auditory brain areas (HG and pSTG) and the speech signal's amplitude envelope. Group-level results from this analysis are shown in Fig. 5.4D. MI was normalized using surrogate data, yielding in z-transformed MI values (see Supplementary Fig. 5.1). Speech tracking was significant at the syllabic rate in all conditions in both hemispheres, surpassing the 99th percentile threshold. Speech tracking occurred exclusively at the syllabic rate of sentences and at no other of the tested frequencies. The normalized MI spectrum, however, revealed an offset in the 5 syllables/s condition, relative to the faster conditions. Therefore, we treated the normalized data cautiously and computed the subsequent analyses on the non-normalized MI values instead (Fig. 5.4 D; Note that our findings are consistent across normalized and non-normalized MI analyses).

Next, we quantified auditory-motor coupling by computing MI between speech motor areas (IFG and SMA) and auditory areas (HG and pSTG; see Fig. 5.5). Using the optimal delay per condition (5 syllables/s = 200ms, 11 syllables/s = 110ms, 14 syllables/s: 90ms, 16 and 17.5 syllables/s: 70ms), we extracted normalized MI spectra, i.e. z-transformed using the surrogate data, for all conditions and ROI pairs (see Supplementary Fig. 5.2 for non-normalized spectra). Auditory-motor coupling peaked around the syllabic rate of the sentences that participants listened to. That is, in the 5 syllables/s condition, auditory-motor coupling peaked around 5 Hz, in the 11 syllables/s condition around 11 Hz, etc. This pattern was more evident for HG-coupling than pSTG-coupling. Additionally, the faster conditions (i.e. > 5 syllables/s) displayed a second peak in the theta frequency (~5-10 Hz). Overall, coupling appeared stronger between HG and motor areas than pSTG and motor areas.

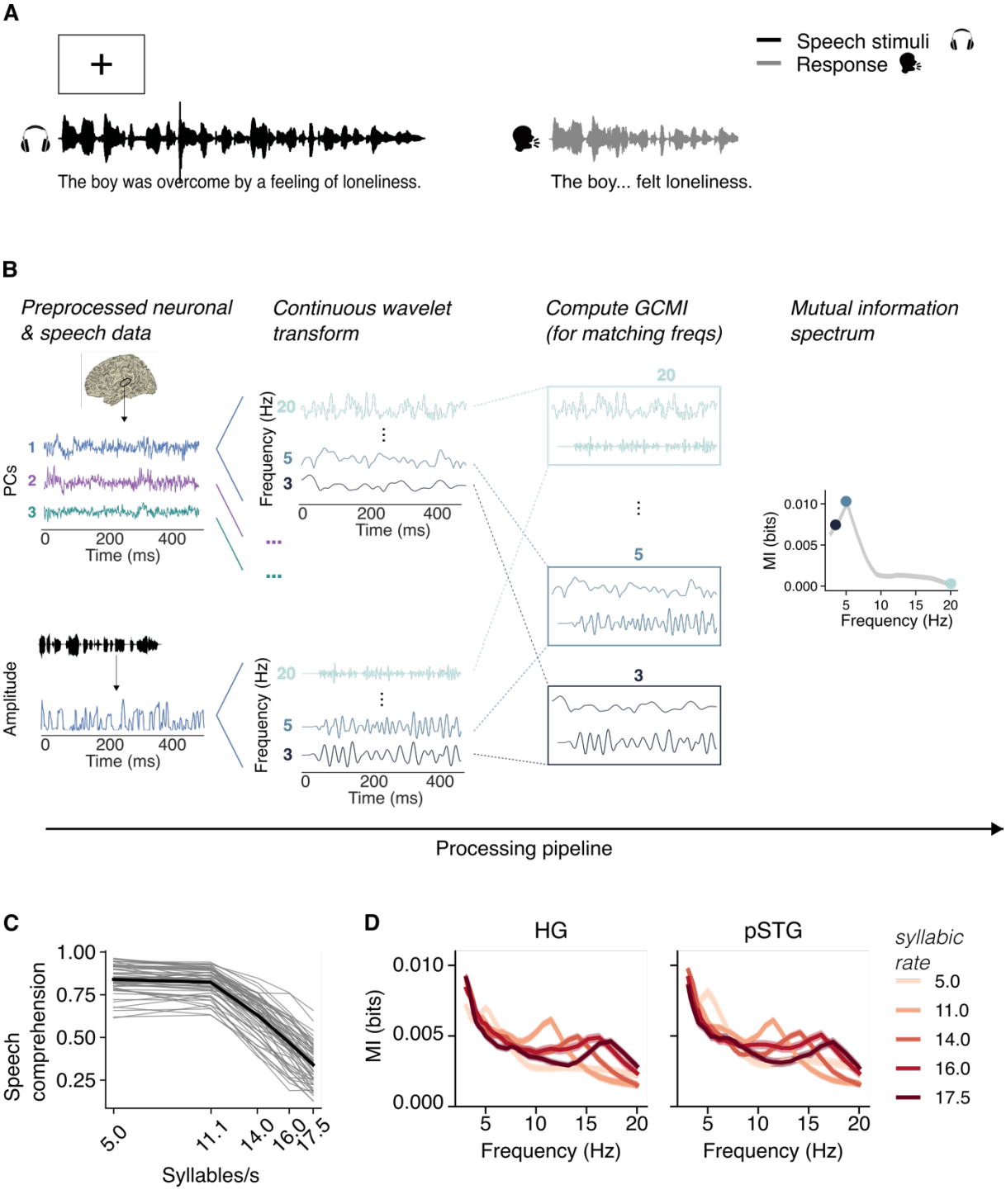


Figure 5.4. HG and pSTG track speech at all syllabic rates. A. During MEG, participants performed an intelligibility task. **B.** Illustration of Gaussian-Copula Mutual Information (GCMI) processing pipeline (reads from left to right). **C.** Line graph displays the behavioral results of intelligibility task: speech was comprehended well at 5 and 11 syllables/s, however, comprehension deteriorated for the faster syllabic rates. Grey lines illustrate single participant comprehension, thick black line represents the average over participants. **D.** Line graph illustrates MI spectra in HG (left) and pSTG (right). For both panels, lines are color-coded according to the syllabic rate of sentence stimuli (see legend) and shaded error bars represent standard error of the mean across participants. HG = Heschl's gyrus, pSTG = posterior superior temporal gyrus; IFG = interior frontal gyrus; SMA = supplementary motor area.

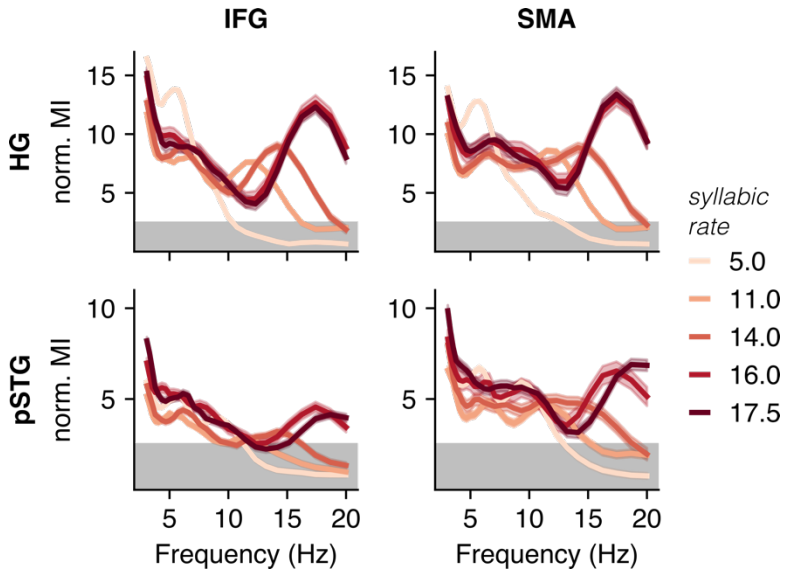


Figure 5.5. Auditory-motor coupling during speech listening. Line graphs illustrate normalized auditory-motor coupling (MI) between different auditory and motor brain areas as a function of frequency: IFG-HG, SMA-HG (top), IFG-pSTG , SMA-pSTG (bottom). Lines are color-coded according to the syllabic rate of sentence stimuli and shaded error bars represent standard error of the mean across participants. HG = Heschl's gyrus, pSTG = posterior superior temporal gyrus; IFG = interior frontal gyrus; SMA = supplementary motor area.

Individual rate of endogenous theta brain rhythms in auditory and motor cortex

We hypothesized that the individual eigenfrequency of the endogenous theta rhythms in auditory and speech motor areas (as assessed in the resting state data) predict speech tracking. To test this hypothesis, we identified spectral profiles in HG, pSTG, IFG, and SMA using a clustering approach. From the spectral profiles we obtained the participants' endogenous theta frequency for each brain region (Fig. 5.6A,C). In all regions, we observed single-subject spectral clusters with clear spectral peaks in the theta range (Fig. 5.6B). Although not all subjects exhibited a theta cluster in all ROIs, the great majority of subjects did (HG-L: N = 55; HG-R: N = 56, pSTG-L: N = 56, pSTG-R: N = 55, IFG-L: N = 56 , IFG-R: N = 56, SMA-L: N = 57, SMA-R: N = 56).

At the group-level, we identified characteristic profiles comprising multiple spectral clusters for HG, pSTG, and SMA (Fig. 5.6D). In HG and pSTG, spectral clusters peaked in the nominal delta, theta, alpha and beta frequency bands, while SMA exhibited clusters peaks in theta (both hemispheres) and beta (only left hemisphere) frequency bands. For IFG, the spectral profiles appeared less characteristic. Although IFG's spectral profile appeared less distinct, individual clusters were extracted due to our motor hypothesis targeting IFG and clear peaks observed in the theta range.

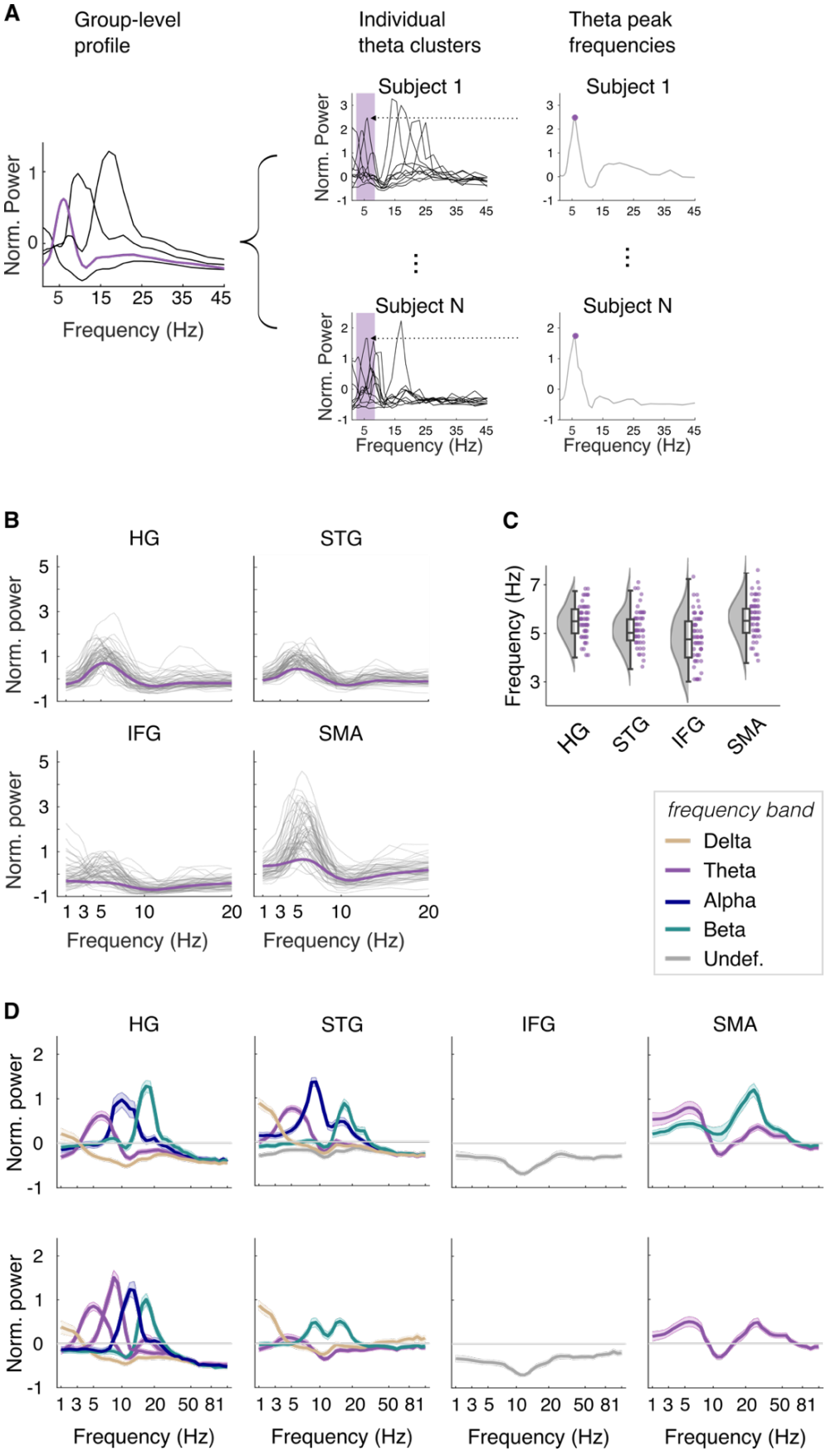

A
Group-level profile
Individual theta clusters
Theta peak frequencies
Norm. Power
Frequency (Hz)
Subject 1
Subject N
B
HG
STG
IFG
SMA
Norm. power
Frequency (Hz)
C
Frequency (Hz)
frequency band
Delta
Theta
Alpha
Beta
Undef.
D
HG
STG
IFG
SMA
Norm. power
Frequency (Hz)

Figure 5.6. Spectral profiles of HG, pSTG, IFG, and SMA. **A.** Illustration of single-subjects theta cluster extraction. Group-level spectral profiles contain multiple spectral clusters, each represented by a line, express region-specific spectral power relative to the power across the whole brain. For further analysis only the cluster peaking in the theta range (i.e. magenta line) was considered. We reconstructed which single-subject clusters (multiple clusters possible) contributed to the group-level theta cluster and selected the individual theta cluster with the highest amplitude. Finally, the frequency at the cluster peak was extracted as individual theta frequency. **B.** Line graphs illustrate individual theta clusters (grey lines) corresponding to the group-level clusters (thick magenta line), averaged across hemispheres. Note that for IFG the group cluster is also colored in magenta (not grey as done above in panel E) to better distinguish group and individual clusters. **C.** Distribution- and boxplots visualize the peak frequencies extracted from the individual theta clusters. **D.** Group-level spectral profiles in the left (top) and right (bottom) hemisphere. Clusters are color-coded according to their peak frequency (legend on the right). Shaded error bars reflect standard error of the mean across participants. HG = Heschl's gyrus; pSTG = posterior superior temporal gyrus; IFG = interior frontal gyrus; SMA = supplementary motor area.

Speech tracking in HG and pSTG differentially affected by auditory-motor parameters

Finally, we combined all neural variables to assess whether speech tracking was affected by the endogenous theta frequency of auditory (HG, pSTG) and motor (IFG, SMA) brain areas, as well as auditory-motor coupling. To this end, we computed two GLMMs, a "HG model" and a "pSTG model".

The HG model (Fig. 5.7A, Supplementary Table 5.5) revealed a negative main effect of syllabic rate (linear: β = - 4.73, SE = 1.00, $p < .001$), reflecting a decrease in speech tracking in HG with increasing syllabic rate. For the endogenous theta frequency of SMA, we observed a positive main effect (β = 0.29, SE = 0.08, p = .006), such that speech tracking was higher in individuals with higher endogenous theta frequencies in SMA. We further observed two two-way interactions: SMA-coupling*endogenous theta frequency in SMA (β = 0.15, SE = 0.06, p = .046) and IFG-coupling*endogenous theta frequency in IFG (β = -0.17, SE = 0.07, p = .034). Both effects showed that the relation of auditory-motor coupling and the endogenous motor frequency did not affect speech-tracking for individuals with low auditory-motor coupling. However, when auditory-motor coupling was stronger, differences in the endogenous frequency of the motor system differentially predicted speech tracking strength. For SMA, speech tracking and endogenous theta frequency in SMA were positively related. In contrast, a negative trend was observed between endogenous theta frequency in IFG and speech tracking. Finally, there was a three-way interaction of theta frequency in HG*theta frequency in IFG*IFG-coupling (β = 0.21, SE = 0.07, p = .021), suggesting that IFG coupling and endogenous theta frequency in IFG interact more strongly in individuals with lower endogenous frequencies in HG.

The pSTG model (Fig. 5.7B, Supplementary Table 5.7) revealed a negative main effect of syllabic rate (linear: β = -6.19, SE = 0.85, $p < .001$), suggesting that speech tracking decreased for higher syllabic rates. Importantly, in contrast to HG, the pSTG model showed a main effect of the endogenous theta frequency (β = -0.33, SE = 0.07, $p < .001$), with increased speech tracking for

individuals with lower endogenous theta frequencies. Like for HG, for the pSTG model main effects of the endogenous theta frequencies in SMA (positive) and IFG (negative) were observed but did not survive the correction for multiple comparisons.

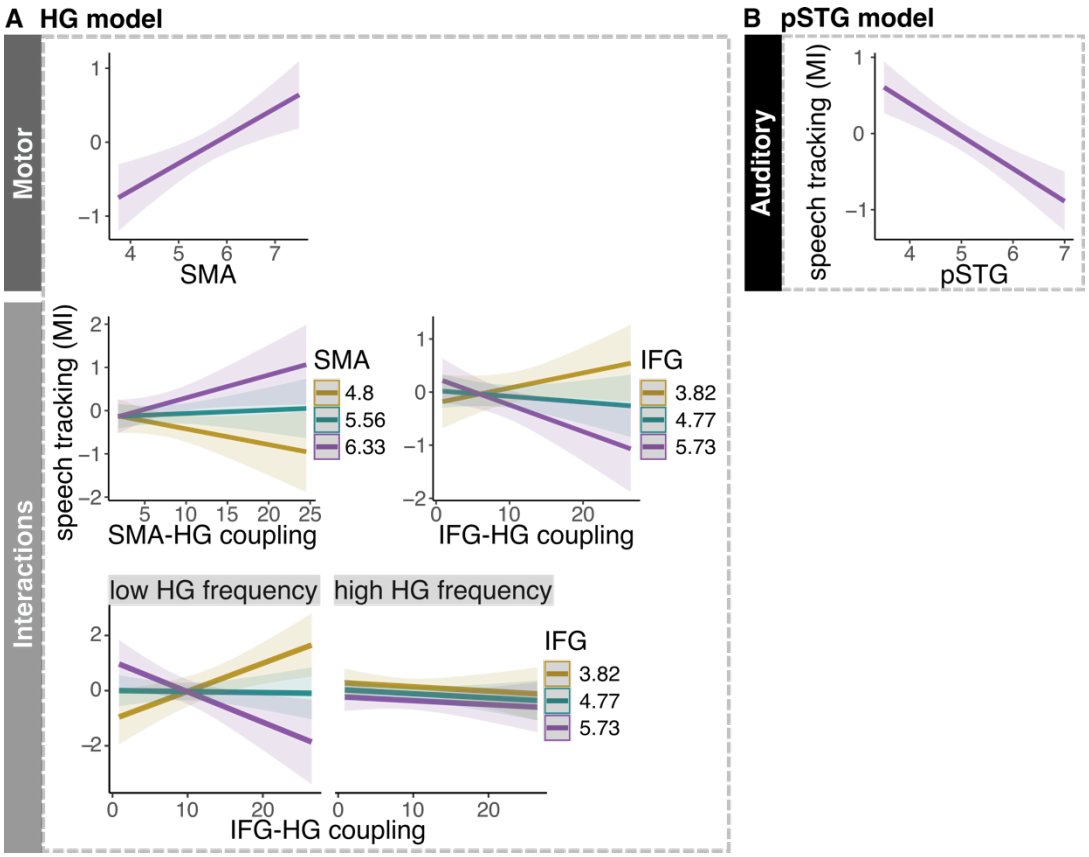


Figure 5.7. Speech tracking in HG and pSTG is predicted by auditory and motor variables. A. The HG model revealed a main effect of SMA frequency on speech tracking in HG. Furthermore, there were interaction effects for SMA-HG coupling and IFG-HG coupling and the respective endogenous motor frequencies. A three-way interaction of HG frequency, IFG frequency, and IFG-HG coupling, suggests stronger interaction of IFG-HG coupling and IFG in individuals with low HG theta frequencies. **B.** In the pSTG model, we observed a negative main effect of the endogenous rate of pSTG on speech tracking (in pSTG). In all panels, error shades indicate 95% confidence intervals. HG = Heschl's gyrus, pSTG = posterior superior temporal gyrus; IFG = interior frontal gyrus; SMA = supplementary motor area.

Speech tracking is related to motor parameters only in high synchronizers

Based on our theoretical assumptions and the behavioral findings, we also computed separate GLMMs for high and low synchronizers (see Supplementary Tables 5.6 and 5.8). For low synchronizers, the HG model revealed an interaction effect of the endogenous theta frequencies of IFG and HG (β = -0.51, SE = 0.17, p = .029) that suggests speech tracking was best when IFG and HG theta frequencies were mismatching (i.e. high IFG and low HG frequency, Fig. 5.8A). As this model produced a singular fit, we conducted a model with simplified random structure (random slope for

participant), which yielded the same effect (β = -0.45, SE = 0.20, p = 0.024). However, it did not withstand FDR control. In the pSTG model, we observed a negative main effect of endogenous theta frequency in pSTG (β = -0.72, SE = 0.15, p < .001; Fig. 5.8B).

For the high synchronizers, the HG model (Fig. 5.8A) showed a positive main effect of endogenous theta frequency in SMA (β = 0.43, SE = 0.13, p = .010) and a negative main effect for endogenous theta frequency in IFG (β = -0.48, SE = 0.16, p = .019). Furthermore, there was an interaction effect of the endogenous frequencies of HG and SMA (β = 0.37, SE = 0.14, p = .041), suggesting that speech tracking was highest in high synchronizers with high endogenous auditory and motor frequencies. A three-way interaction of theta frequency in HG*theta frequency in SMA*SMA-HG coupling (β = 0.25, SE = 0.10, p = .041) further specified the two-way interaction by suggesting the increased tracking for matched high auditory and motor theta frequencies was driven more strongly by individuals with high SMA-HG coupling.

The pSTG model (Fig. 5.8B) for the high synchronizers showed a main effect of the endogenous theta frequency of pSTG (β = -0.38, SE = 0.09, p < .001), as observed in the low synchronizers and the full model. Furthermore, similar to the HG model, the pSTG model revealed a positive main effect of SMA theta frequency (β = 0.33, SE = 0.09, p = .003) and a negative main effect of IFG theta frequency (β = -0.28, SE = 0.09, p = .007). In addition, we observed an interaction effect of the endogenous theta frequencies in SMA and pSTG (β = -0.39, SE = 0.11, p = .002). For individuals with low endogenous theta frequency in SMA, speech tracking was unaffected by the theta frequency in pSTG. However, for individuals with higher theta frequency in SMA, speech tracking varied with theta frequency in pSTG such that tracking was highest in individuals with lower pSTG theta frequencies.

In summary, the comparison between HG and pSTG models reveals that the endogenous theta frequency of pSTG had a direct association with speech tracking, while the endogenous theta frequency of HG only played a role by interacting with other variables (such as endogenous motor frequencies and coupling). When comparing low and high synchronizers, a remarkably consistent pattern emerges suggesting that high synchronizers exhibit stronger rhythmic motor system engagement during tracking (i.e. effects of the endogenous motor frequency or SMA-HG coupling).

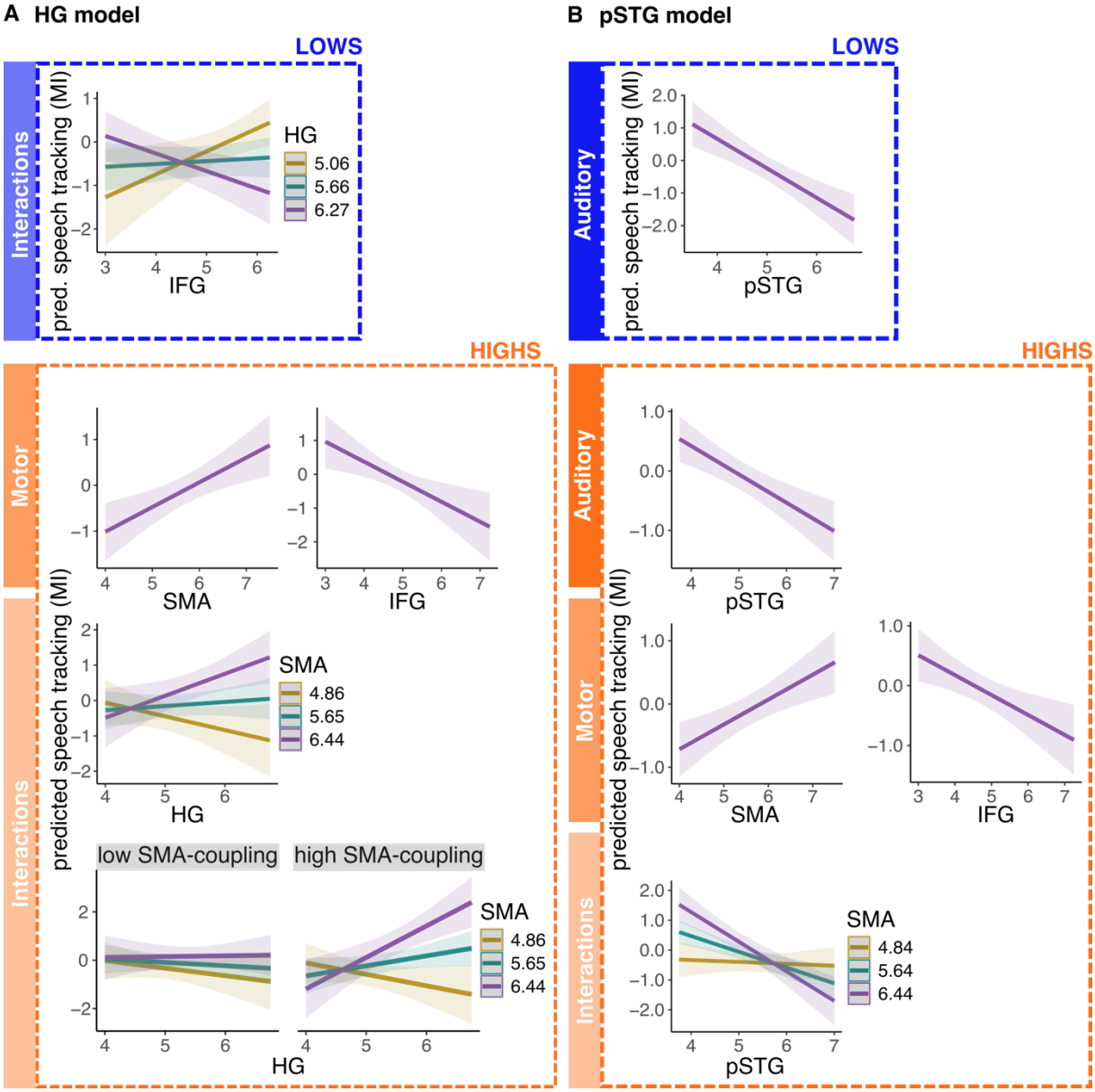


Figure 5.8. Speech tracking in HG and pSTG is predicted by different variables in low vs. high synchronizers. A. In low synchronizers (blue), the HG model revealed an interaction effect of the endogenous frequencies of HG and IFG. In high synchronizers (orange), we observed main effects of the endogenous SMA (positive) and IFG (negative) frequencies. Additionally, speech tracking was predicted by a HG frequency * SMA frequency interaction effect, and a three-way interaction of HG frequency, SMA frequency, and SMA-HG coupling, suggesting a stronger interaction of SMA and HG frequencies in individuals with higher SMA-HG coupling. **B.** For the pSTG model, we observed a negative main effect of pSTG frequency in both low and high synchronizers. In the high synchronizers, several additional effects were observed: main effects of the endogenous SMA (positive) and IFG (negative) frequencies as well as an interaction effect of pSTG and SMA frequencies. In all panels, error shades indicate 95% confidence intervals. HG = Heschl's gyrus, pSTG = posterior superior temporal gyrus; IFG = interior frontal gyrus; SMA = supplementary motor area.

Similar patterns in behavioral and neural measures

Both behavioral and neural analyses exhibited similar patterns, with effects of the motor system and auditory-motor coupling observed predominantly in individuals displaying stronger behavioral auditory-motor synchronization (high synchronizers). Despite similar patterns of results for behavioral and neural measures, no correlations were observed. We computed partial correlations between speech comprehension and speech tracking, to account for the correlation of syllabic rate and both comprehension and tracking. The partial correlations revealed no significant effects (HG: rho = -0.03, p = .679, pSTG: rho = -0.05, p = .382). Regarding preferred rates, we hypothesized that the behavioral measures of the auditory and motor rates and auditory-motor synchronization reflect behavioral readouts of the underlying -supposedly- oscillatory properties of the corresponding brain systems. However, correlation analyses showed no significant relations between the corresponding neural and behavioral rates (all ps > .05), even when exploring high and low synchronizers separately.

5.5 Discussion

Our findings show that individual's rates of endogenous (resting-state) auditory and motor brain rhythms and auditory-motor coupling predict speech tracking in auditory cortex during continuous listening. This supports oscillatory accounts and refines models of auditory speech tracking and auditory-motor interactions, revealing that distinct processes may be recruited for different parts of the population. Specifically, we find that across the population individual rates of endogenous theta brain rhythms of auditory association cortex (pSTG) predicted speech tracking in this area. Furthermore, individual rates of endogenous theta brain rhythms of speech motor cortices (SMA, IFG) predicted speech tracking in auditory association areas—but only in individuals with behaviorally quantified high auditory-motor synchronization. A slightly different picture emerged for primary auditory cortex (HG). No main effect of the endogenous HG rhythm was observed, neither across the population, nor separately in low or high synchronizers. Instead, in high synchronizers, only individuals with high endogenous SMA rates showed an effect of the endogenous HG rate on speech tracking. Similarly, as for auditory association cortex, endogenous rates of theta brain rhythms of speech motor cortices (SMA, IFG) predicted speech tracking in HG in high synchronizers. Additionally, primary auditory cortex exhibited effects of auditory-motor coupling. Primary auditory cortex thus seems more receptive to external stimulation rates and less rhythmically constrained than pSTG. Together with the behavioral findings, our data suggest that during speech comprehension, individuals with higher-auditory motor synchronization use motor top-down predictions through coupled auditory-motor oscillators. In contrast, low auditory-motor synchronizers rely on auditory oscillatory populations and additionally more heavily recruit working memory processes.

Speech comprehension and tracking decline at higher syllabic rates

As expected, we found that comprehension decreased at higher syllabic rates (above 11 syllables/sec). Although the decline was observed at slightly higher rates than previously reported in studies using simpler stimuli (e.g. words or short sentences; Ahissar et al., 2001; Ghitza and Greenberg, 2009; Doelling et al., 2014), recent studies have observed similar effects (Verschueren et al., 2022; Giroud et al., 2023; Lubinus et al., 2023). A slower decline across rates, may be accounted for by enhanced comprehension due to linguistic predictability particularly at faster syllabic rates (Lubinus et al., 2023). Despite the observed decline, auditory cortex successfully tracked the amplitude envelope even at non-intelligible rates, and unlike Ahissar et al. (2001) we observed no correlation of comprehension and tracking (after controlling for confounding correlations). Although our findings are in line with previous studies pointing towards a more complex link between comprehension and tracking (Howard and Poeppel, 2010; Pefkou et al., 2017; Verschueren et al., 2022; Kösem et al., 2023), we acknowledge potential limitations due to a lack of trial-wise neural data analysis reducing statistical power.

The rate of endogenous brain rhythms in pSTG predicts speech tracking

Endogenous theta rhythms in auditory cortex are believed to phase-align to the acoustic envelope of incoming speech (Lakatos et al., 2005; Luo and Poeppel, 2007; Giraud and Poeppel, 2012; Gross et al., 2013; Peelle et al., 2013). Our study, for the first time, establishes a direct relationship between an individual's rate of the endogenous theta rhythm (measured during rest) and speech tracking in auditory association cortex (pSTG) during comprehension. We found that lower endogenous theta frequencies in pSTG were consistently associated with stronger speech tracking (in both high and low synchronizers), suggesting recruitment of an oscillatory population during speech tracking (Giraud et al., 2007; Giraud and Poeppel, 2012; Keitel and Gross, 2016; Rimmele and Keitel, 2023). In contrast, the endogenous theta frequency of primary auditory cortex (HG) was not predictive of its speech tracking (Note that there was a positive predictive effect for a subgroup of high synchronizers with high endogenous SMA theta rates). We speculate that due to a stronger preference for speech signals in pSTG (Poeppel, 2003; Hamilton et al., 2021), the role of its endogenous rhythm may be more relevant to speech tracking. Intracranial results demonstrated that pSTG preferentially represents "syllable-level temporal structure" (Hamilton et al., 2021), whereas HG reflects less complex acoustic features (Hullett et al., 2016; Hamilton et al., 2021). Furthermore, oscillatory neural populations in pSTG may be involved in speech segmentation (Poeppel, 2003; Giraud and Poeppel, 2012), with possibly weaker oscillatory properties in HG.

Endogenous motor cortex rhythms predict speech tracking only in high synchronizers

Endogenous speech motor cortex theta rhythms reliably predicted speech tracking in high, but not in low synchronizers. The endogenous theta frequencies of SMA and IFG exhibited contrasting effects on speech tracking, with higher SMA rates relating to higher and higher IFG rates to lower speech tracking. This is consistent across tracking in auditory association and primary auditory cortex (pSTG and HG; Fig. 5.8). Given SMA's role in temporal processing (Macar et al., 2002; Lewis and Miall, 2003b; Schwartze et al., 2012b; Hertrich et al., 2016) and IFG' involvement in temporal sequencing and speech motor processing (Hertrich et al., 2016), these findings indicate that high synchronizers may rely more heavily on temporal motor predictions for speech processing, compared to low synchronizers. This interpretation is further supported by our behavioral observation of increased comprehension performance in high synchronizers. The endogenous theta frequency in IFG showed a negative effect, indicating enhanced tracking with lower endogenous rates. Given the proposed direct connection between IFG and pSTG (Hertrich et al., 2016), observing the same direction of effects may imply a synergistic effect of (the endogenous frequencies in) these two areas on tracking. In contrast, the positive effect of the endogenous theta frequency in SMA on tracking paralleled our behavioral findings and showed the expected direction of the effect. It is worth noting that this positive effect of SMA endogenous rates matched the direction of HG endogenous rate effects (Fig. 5.8; which however was only observed in a small subpopulation). The discrepancies between the SMA and IFG effects may reflect differences in network connections, as well as whether the areas are connected in an excitatory or inhibitory manner (Roux et al., 2022). Further research is warranted to elucidate the distinct relation of endogenous IFG and SMA rhythms to speech tracking. The finding of selective motor effects for high synchronizers is in line with previous work (Assaneo et al., 2019, 2021; Kern et al., 2021; Rimmele et al., 2022; Lubinus et al., 2023; Barchet et al., 2024). Particularly, Assaneo, Rimmele et al. (2021) proposed, based on behavioral data and a neural computational model, that oscillatory auditory-motor coupling was engaged only in high synchronizers. Here, in low synchronizers, we only detected an interaction effect between the endogenous theta peak frequencies in IFG and HG. High synchronizers may not only rely more strongly on the recruitment of the motor system but also more extensively engage additional areas like (parts of) the SMA, reflecting other aspects of temporal processing or alternative processing routes (Hertrich et al., 2016).

Auditory-motor coupling interacts to affect speech tracking in HG

We observed similar patterns for auditory-motor coupling between different motor (IFG, SMA) and auditory (HG, pSTG) areas (Fig. 5.5), with significant auditory-motor coupling at all syllabic rates. Coupling, however, decreased from 5 Hz to 11 Hz and increased again in the faster conditions. The findings align with reports of a "sweet-spot" for auditory-motor coupling around 4.5 Hz

(Assaneo and Poeppel, 2018). Furthermore, such a sensitivity for certain frequencies in the theta and beta ranges has been previously shown for the auditory cortex (Teng and Poeppel, 2020) and for the motor cortices (as reflected in their endogenous rhythms; Keitel and Gross, 2016; Lubinus et al., 2021). A different oscillatory regime in the beta range may dominate when the auditory input surpasses the theta regime. Alternatively, the increased coupling in the beta range could be driven mainly by the endogenous beta rhythms of the motor areas, instead of reflecting genuine auditory-motor coupling. However, in our data, we do not observe a beta cluster in the spectral profiles of IFG (Fig. 6D) and, thus, it is unlikely that the endogenous brain rhythms could explain this effect. Speech tracking was predicted by HG-SMA-coupling only in high synchronizers and in a three-way interaction with the endogenous rates (Fig. 5.8). High coupling was related to higher tracking only for high SMA and HG endogenous rates. Our results are limited by the focus on phase-phase coupling at the syllabic stimulation rate. Further research is required to understand oscillatory processing routes within a spectrally and spatially complex auditory-motor network.

Behavioral findings parallel the neurophysiological results

Replicating our behavioral study (Lubinus et al., 2023), we show that higher speech comprehension was related to higher individual spontaneous motor production rates, auditory-motor synchronization, and preferred auditory rates (in our previous study, the later was a trend that did not survive multiple-comparison control). A 3-way interaction between these variables suggests a complex interplay of the preferred rates and auditory-motor synchronization. Interestingly, the behavioral findings remarkably paralleled our neural results. In low synchronizers, comprehension was predicted by the preferred auditory rate and interestingly by working memory scores. In contrast, in high synchronizers, comprehension was predicted by the preferred auditory rate, and crucially the motor production rate and the interaction of auditory and motor rates.

Despite similar effects in behavior and the brain, we observed no correlations between the corresponding measures. Speech comprehension is intricate and shaped by numerous interacting variables. While the neural dynamics underlying speech tracking are crucial, they likely represent just one aspect of the complex computations—including linguistic and situational predictions, as well as working memory—that facilitate speech comprehension.

Limitations

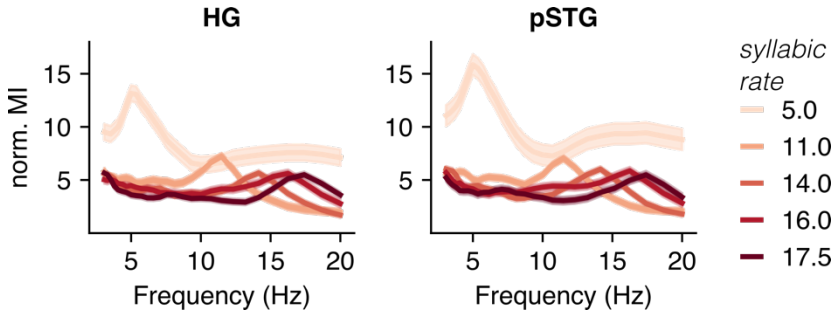
Speech comprehension is multifaceted, engaging various cognitive processes and multiple brain areas in a vast language network. In this study, we focused on specific nodes within this network (Hickok and Poeppel, 2004; Tourville and Guenther, 2011) to quantify the relationship between individual endogenous theta brain rhythms and speech processing, aiming to elucidate a model of oscillatory auditory-motor coupling. Furthermore, the asymmetric sampling in time hypothesis

(Poeppel, 2003) postulates distinct time preferences for left and right pSTG (Giraud et al., 2007). Given the strong correlations between the left and right brain areas observed in the current work, we averaged data across hemispheres to reduce dimensionality. Future research could investigate more detailed hypotheses concerning the complex interplay of hemispheric lateralization in endogenous brain rhythms.

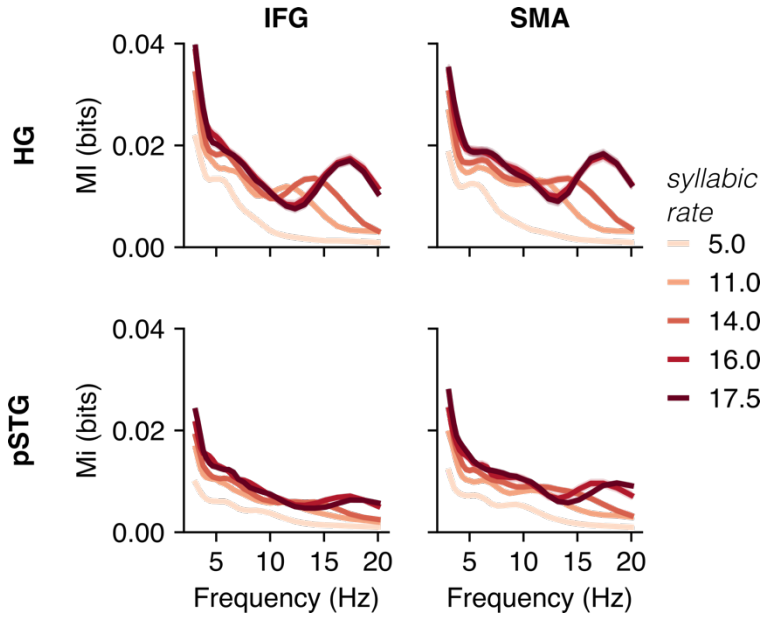
5.6 Conclusion

In a large sample of participants, we demonstrate that endogenous theta rhythms of pSTG predict speech tracking. Interestingly, endogenous theta rhythms of speech motor areas (SMA, IFG) were predictive of speech tracking only in those individuals showing high auditory-motor synchronization. Our findings support an oscillatory model of auditory-motor interactions during speech comprehension. While some individuals recruit the motor system (SMA, IFG), likely providing temporal predictions to enhance comprehension, others predominantly rely on auditory processing, possibly taxing working memory processes, more strongly. Furthermore, our findings suggest differences of primary auditory and association areas in their oscillatory characteristics relevant for speech tracking and processing routes.

5.7 Supplementary materials



Supplementary Figure 5.1. Normalized MI spectra for speech tracking in auditory cortex. Line graph illustrates normalized MI spectra in Heschl's gyrus (HG, left) and posterior superior temporal gyrus (pSTG, right). For both panels, lines are color-coded according to the syllabic rate of sentence stimuli and shaded error bars represent standard error of the mean across participants.



Supplementary Figure 5.2. Raw MI spectra for auditory-motor coupling. Line plots illustrate auditory-motor coupling (MI) between different auditory and motor brain areas as a function of frequency: IFG-HG, SMA-HG (top), IFG-pSTG , SMA-pSTG (bottom). Lines are color-coded according to the syllabic rate of sentence stimuli and shaded error bars represent standard error of the mean across participants. HG = Heschl's gyrus, pSTG = posterior superior temporal gyrus; IFG = interior frontal gyrus; SMA = supplementary motor area.

Supplementary Table 5.1. Sentence materials: Titles and authors of (audio)books. List of sources from which stimuli for the tasks (speech comprehension and preferred auditory rate tasks) were constructed.

Type	Author	Title	Source
PDF	Hedwig Dohm	Sibilla Dalmar	Zeno
PDF	Hedwig Dohm	Schicksale einer Seele	Zeno
PDF	Hedwig Dohm	Christa Ruland	Zeno
PDF	Dora Duncker	Großstadt	Zeno
PDF	Bertha von Suttner	Martha's Kinder	Zeno
PDF	Bertha von Suttner	Die Waffen nieder	Zeno
PDF	Stefan Zweig	Schachnovelle	Zeno
PDF	Ida Boy-Ed	Fanny Förster	Zeno
Audiobook	Frances Hodgson Burnett	Der kleine Lord	Librivox
Audiobook	Ludwig Thoma	Tante Frieda	Librivox
Audiobook	Ludwig Thoma	Lausbubengeschichten	Librivox

Supplementary Table 5.2. Prompts for speech production task. Thematic questions used to facilitate natural speech production. Each item is representative of a different thematic category, as introduced by Alexandrou et al.(2016). Sentences were translated to German.

Category	Sentence/statement
Own life	Welche Hobbies haben Sie oder hatten Sie in Ihrem bisherigen Leben? What kind of hobbies do you have or have had during your life?
Preferences	Welche Art von Urlaubstrips mögen Sie am liebsten? What kinds of vacation trips do you like?
People	Beschreiben Sie bitte einen Künstler, Schriftsteller oder Filmdirektor. Warum finden Sie ihn/sie interessant? Describe a known artist, writer, or film director. Why do you find her/him interesting?
Culture/traditions	Beschreiben Sie bitte ein traditionelles Weihnachtsfest. Describe a traditional Christmas holiday.
Society/politics	Was wissen Sie über Müll- und Recyclingregelungen in Ihrem Heimatland? What do you know about garbage and recycling policies in your home country?
General knowledge	Was wissen Sie über Skifahren und Snowboarden? What do you know about skiing and snowboarding?

Supplementary Table 5.3. Predicting single-trial speech comprehension performance.

Predicted speech comprehension

Predictors	beta	SE	t	p
syllabic rate	-70.81	1.78	-39.76	<0.001
syllabic rate^2	-24.36	0.98	-24.82	<0.001
PLV	0.03	0.02	2.28	0.032
motor rate	0.03	0.02	2.02	0.056
auditory rate	0.09	0.02	5.94	<0.001
working memory	0.07	0.01	4.94	<0.001
compression	-0.00	0.01	-0.19	0.850
num. syllables	-0.11	0.02	-4.99	<0.001
stimulus order	0.05	0.00	10.21	<0.001
PLV * motor rate	0.07	0.02	4.16	<0.001
PLV * auditory rate	0.01	0.02	0.50	0.708
motor rate * auditory rate	-0.06	0.02	-3.29	0.002
PLV * motor rate * auditory rate	-0.06	0.02	-3.58	0.001

Random effects

Group	Name	Variance	STD
Trial-ID	(Intercept)	0.203231	0.45081
Subjects	Syllabic rate	0.000377	0.01942
Residual		0.2944557	0.54264
N (subs)		47	
Observations		14100	
Marginal R^2		0.478	

Supplementary Table 5.4. Predicting single-trial speech comprehension performance, separately for high and low synchronizers.

Predicted speech comprehension

	High synchronizers				Low synchronizers			
Predictors	beta	SE	t	p	beta	SE	t	p
syllabic rate	-53.61	1.69	-31.76	<0.001	-45.67	2.02	-22.58	<0.001
syllabic rate^2	-19.73	0.98	-20.15	<0.001	-15.82	0.92	-17.14	<0.001
motor rate	0.08	0.02	3.73	<0.001	-0.02	0.02	-0.97	0.473
auditory rate	0.05	0.02	2.73	0.010	0.08	0.02	3.93	<0.001
working memory	-0.02	0.02	-0.76	0.506	0.13	0.02	6.08	<0.001
compression	0.01	0.01	0.50	0.621	-0.01	0.01	-0.67	0.630
num. syllables	-0.12	0.02	-4.82	<0.001	-0.10	0.02	-4.50	<0.001
stimulus order	0.03	0.01	4.74	<0.001	0.07	0.01	9.70	<0.001
motor rate * auditory rate	-0.04	0.02	-2.60	0.013	0.01	0.02	0.37	0.765

Random effects

		High synchronizers		Low synchronizers	
Group	Name	Variance	STD	Variance	STD
Trial-ID	(Intercept)	0.213540	0.46210	0.189647	0.43548
Subjects	Syllabic rate	0.003219	0.01794	0.000547	0.02339
Residual		0.300949	0.54859	0.288296	0.52693
N (subs)		27		20	
Observations		8100		6000	
Marginal R^2		0.471		0.481	

Supplementary Table 5.5. Predicting speech tracking in HG.

Predicted speech tracking in HG

Predictors	beta	SE	t	p
syllabic rate	-4.73	1.00	-4.74	<0.001
syllabic rate^2	-0.18	0.95	-0.19	0.848
HG	-0.03	0.08	-0.40	0.793
SMA-HG coupling	0.03	0.06	0.44	0.793
SMA	0.29	0.08	3.44	0.006
IFG-HG coupling	-0.05	0.07	-0.70	0.704
IFG	-0.14	0.09	-1.57	0.268
HG × SMA-HG coupling	0.03	0.07	0.39	0.793
HG × SMA	0.15	0.08	1.82	0.188
SMA x SMA-HG coupling	0.15	0.06	2.46	0.046
HG × IFG-HG coupling	-0.02	0.07	-0.33	0.794
HG × IFG	-0.13	0.10	-1.26	0.417
IFG x IFG-HG coupling	-0.17	0.07	-2.65	0.034
(HG × SMA) × SMA-HG coupling	0.06	0.07	0.94	0.556
(HG × IFG) × IFG-HG coupling	0.21	0.07	2.98	0.017

Random effects

Group	Name	Variance	STD
Subjects	Syllabic rate	0.001271	0.03566
Residual		0.659555	0.81213
N (subs)		54	
Observations		270	
Marginal R^2		0.248	

Supplementary Table 5.6. Predicting speech tracking in HG, separately for high and low synchronizers.

Predicted speech tracking in HG

	High synchronizers				Low synchronizers			
Predictors	beta	SE	t	p	beta	SE	t	p
syllabic rate	-3.04	0.92	-3.32	0.010	-3.77	1.10	-3.43	0.015
syllabic rate^2	-0.57	0.83	-0.69	0.608	1.27	1.05	1.21	0.469
HG	0.09	0.12	0.70	0.608	-0.10	0.13	-0.79	0.632
SMA	0.43	0.13	3.29	0.010	-0.01	0.10	-0.08	0.969
SMA-HG coupling	0.00	0.08	0.02	0.987	-0.18	0.13	-1.37	0.469
IFG	-0.48	0.16	-2.97	0.019	0.07	0.12	0.56	0.765
IFG-HG coupling	-0.10	0.10	-0.96	0.490	0.14	0.12	1.20	0.469
HG × SMA	0.37	0.14	2.53	0.041	-0.10	0.10	-0.98	0.524
HG × SMA-HG coupling	0.19	0.09	2.09	0.089	-0.15	0.12	-1.22	0.469
SMA × SMA-HG coupling	0.14	0.09	1.60	0.199	-0.00	0.11	-0.04	0.969
HG × IFG	0.00	0.14	0.02	0.987	-0.51	0.17	-2.99	0.029
HG × IFG-HG coupling	-0.01	0.09	-0.11	0.987	0.02	0.14	0.16	0.969
IFG × IFG-HG coupling	-0.23	0.11	-2.17	0.086	-0.10	0.10	-1.05	0.524
(HG × SMA) × SMA-HG coupling	0.25	0.10	2.56	0.041	-0.02	0.10	-0.15	0.969
(HG × IFG) × IFG-HG coupling	0.14	0.08	1.73	0.171	0.26	0.15	1.69	0.382

Random effects

		High synchronizers		Low synchronizers	
Group	Name	Variance	STD	Variance	STD
Subjects	Syllabic rate	0.001891	0.04348	0.000	0.000
Residual		0.492249	0.70160	0.8133	0.9018
N (subs)		28		20	
Observations		140		100	
Marginal R^2		0.503		0.276	

Supplementary Table 5.7. Predicting speech tracking in pSTG.

Predicted speech tracking in pSTG

Predictors	beta	SE	t	p
syllabic rate	-6.19	0.85	-7.32	<0.001
syllabic rate^2	0.68	0.88	0.78	0.719
pSTG	-0.33	0.07	-4.75	<0.001
SMA-pSTG coupling	0.05	0.06	0.77	0.719
SMA	0.17	0.07	2.27	0.120
IFG-pSTG coupling	0.06	0.06	0.96	0.719
IFG	-0.16	0.07	-2.18	0.120
pSTG × SMA-pSTG coupling	-0.00	0.06	-0.06	0.952
pSTG × SMA	-0.16	0.08	-1.97	0.160
SMA-pSTG coupling × SMA	0.02	0.06	0.31	0.910
pSTG × IFG-pSTG coupling	-0.03	0.06	-0.47	0.851
pSTG × IFG	-0.02	0.08	-0.21	0.910
IFG-pSTG coupling × IFG	-0.04	0.06	-0.68	0.719
(pSTG × SMA) × SMA-pSTG coupling	-0.05	0.07	-0.74	0.719
(pSTG × IFG) × IFG-pSTG coupling	-0.08	0.06	-1.18	0.641

Random effects

Group	Name	Variance	STD
Subjects	Syllabic rate	0.0005466	0.02338
Residual		0.6000252	0.77461
N (subs)		50	
Observations		250	
Marginal R^2		0.356	

Supplementary Table 5.8. Predicting speech tracking in pSTG, separately for high and low synchronizers.

Predicted speech tracking in pSTG

	High synchronizers				Low synchronizers			
Predictors	beta	SE	t	p	beta	SE	t	p
syllabic rate	-4.61	0.77	-5.98	<0.001	-4.60	0.79	-5.82	<0.001
syllabic rate^2	-0.02	0.80	-0.02	0.981	1.79	0.87	2.07	0.223
pSTG	-0.42	0.08	-4.99	<0.001	-0.37	0.10	-3.68	0.003
SMA	0.30	0.08	3.62	0.002	0.10	0.10	1.01	0.842
SMA-pSTG coupling	0.11	0.08	1.42	0.236	0.00	0.10	0.02	0.981
IFG	-0.26	0.08	-3.27	0.004	0.14	0.10	1.44	0.614
SMA-pSTG coupling	0.06	0.07	0.86	0.445	-0.02	0.11	-0.18	0.977
pSTG × SMA	-0.37	0.09	-4.09	<0.001	0.06	0.14	0.45	0.977
pSTG × SMA-pSTG coupling	-0.00	0.07	-0.03	0.981	0.05	0.10	0.49	0.977
SMA × couplingSMA	0.11	0.08	1.28	0.269	-0.00	0.10	-0.03	0.981
pSTG × IFG	-0.25	0.10	-2.52	0.030	-0.13	0.13	-1.06	0.842
pSTG × couplingIFG	-0.24	0.09	-2.65	0.024	-0.02	0.10	-0.25	0.977
IFG × couplingIFG	-0.11	0.08	-1.41	0.236	0.02	0.11	0.23	0.977
(pSTG × SMA) × SMA-pSTG coupling	-0.16	0.10	-1.59	0.205	0.03	0.11	0.27	0.977
(pSTG × IFG) × IFG-pSTG coupling	-0.21	0.12	-1.81	0.145	0.05	0.10	0.48	0.977

Random effects

		High synchronizers		Low synchronizers	
Group	Name	Variance	STD	Variance	STD
Subjects	Syllabic rate	0.00019	0.0137	0.000	0.000
Residual		0.52049	0.72145	0.6032	0.7767
N (subs)		27		19	
Observations		135		95	
Marginal R^2		0.463		0.393	

6 General discussion

In three experimental studies, this thesis investigated the temporal constraints of speech perception and their relationship with the endogenous dynamics of both the auditory and motor systems, as well as their interaction. Each study approached this overarching objective from a distinct perspective (1) leveraging long-term neural reorganization of endogenous brain rhythms in a population of congenitally blind individuals known to show behavioral adaptation of temporal constraints, (2) investigating individual differences of behavioral and (3) neural indices of auditory and motor endogenous rhythms and coupling, and their relation on speech processing. While detailed discussions of the results are provided in the respective sections (chapters 3-5), the following chapter intends to offer a comprehensive and integrated perspective on all studies. First, I will revisit the concept of spectral plasticity, discussing the possibility of changes in brain rhythms and how studying such adaptations can advance our understanding of the brain more broadly. Next, I will review the influence of the motor system on speech comprehension and tracking, reflecting how top-down predictions, including those from the motor system, may be integrated during speech comprehension. Lastly, I will focus on individual differences, emphasizing the importance of accounting for individual variation to deepen our comprehension of how the brain processes speech.

6.1 Summary of experimental results

Visual deprivation is associated with behavioral adaptations, including superior performance in auditory perception tasks, such as an advanced ability for fast speech comprehension. Study 1 explored the hypothesis that these established behavioral adaptations may be accompanied by spectral adaptations of endogenous brain rhythms, arising from neural reorganization. Using resting-state MEG data and spectral fingerprinting, we identified spectral characteristics of individual brain areas across cortex in both normally sighted and congenitally blind individuals. Replicating Keitel and Gross (2016), classification analysis in the sighted illustrated that region-specific spectral profiles effectively distinguished between different brain areas. Extending these findings, similar classification accuracies were observed in congenitally blind individuals. Cross-group classification revealed both similarities and differences in spectral signatures. Specifically, reduced classification performance in sensory and right frontal brain regions suggests spectral reorganization in congenital blindness. In contrast, various brain regions, including left frontal and bilateral parietal cortices, showed consistent spectral signatures across groups. Notably, congenitally blind individuals displayed a second, faster theta cluster in the spectral profile of right Heschl's gyrus (HG), while only one cluster was observed in the normally sighted group. Although we did not correlate this neural finding with performance, it aligns with previously reported enhancements in fast speech perception

abilities among congenitally blind individuals. This alignment supports the proposed relationship between endogenous auditory theta brain rhythms and the capacity for speech tracking.

In a behavioral design, Study 2 investigated if speech comprehension can be predicted by behavioral estimates of endogenous auditory and motor brain rhythms and auditory-motor coupling. In two experiments and multiple sessions, we measured individuals' preferred perceptual speech rate, spontaneous speech motor production rate, auditory-motor synchronization, and their ability to comprehend speech as a function of syllabic rate. The results demonstrated that individual speech motor production rates and auditory-motor synchronization contribute to predicting a person's speech comprehension abilities. Specifically, individuals with higher spontaneous speech motor production rates and stronger auditory-motor synchronization exhibited better speech comprehension across a range of syllabic rates. Moreover, this study highlights the complexity of naturalistic speech comprehension, revealing that comprehension increased with the predictability of sentences particularly evident at fast speech rates. Notably, individuals with stronger auditory-motor synchronization derived greater advantages from sentence predictability in speech processing.

Study 3 investigated the link between the neural endogenous rates of the auditory and motor cortex, the coupling strength between these areas, and speech tracking in auditory cortex. Given the proposed functional role of speech tracking, this study further delineates the computational level at which the endogenous auditory and motor rhythms may contribute to speech comprehension, specifically focusing on syllabic segmentation. Brain-to-speech synchronization, i.e. speech tracking, was estimated at multiple syllabic rates using mutual information. The study examined how individual tracking was influenced by the endogenous theta peak frequencies of the auditory (HG and pSTG) and motor (IFG and SMA) brain areas, measured during resting state, as well as auditory-motor coupling, measured during the intelligibility task. Consistent with the behavioral findings, speech tracking in pSTG was predicted by the endogenous theta peak frequency of pSTG. Furthermore, peak frequencies of the endogenous motor rhythms and auditory-motor coupling affected speech tracking, but this effect was observed only in individuals with strong behavioral auditory-motor synchronization.

6.2 Revisiting spectral plasticity – adapted brain rhythms

Little is known about the plasticity of spectral signatures when brains undergo reorganization. Study 1 of the current thesis provides the first whole-brain characterization of local spectral profiles in congenital blindness. The results replicate the well-established observation of a reduced or absent visual alpha rhythm following visual deprivation (Adrian and Matthews, 1934; Noebels et al., 1978; Kriegseis et al., 2006; Hawellek et al., 2013; Schubert et al., 2015). This finding is hypothesized to reflect atrophy of the cortico-thalamic and cortical pathways, the proposed source of posterior alpha

generators (Lopes da Silva, 1991; Lőrincz et al., 2009). In the current sample, spectral profiles in right HG and left STG were distinct between sighted and congenitally blind individuals. While blindfolded, congenitally blind individuals displayed two theta clusters in left HG along with higher power, whereas the sighted control group showed only one theta peak in this brain region. This parallels the finding of adaptive speech comprehension in congenital blindness, particularly the enhanced abilities to understand accelerated speech (Hertrich et al., 2009, 2013; Dietrich et al., 2013). While the current work lacks a behavioral task to substantiate such claims, follow-up studies could directly test whether the faster endogenous theta rhythms in HG and pSTG in the congenitally blind predict their comprehension of fast speech. Spectral adaptations observed in the context of brain reorganization and behavioral adaptations offer a potent avenue for enhancing our understanding of the functional role of brain rhythms. On the one hand, examining how brain rhythms adapt in congenital blindness can elegantly evaluate their functional relevance for perceptual and cognitive processes. Brain rhythms believed to support a certain cognitive process (i.e. speech segmentation) would be expected to be altered, if the cognitive process is known to be adapted in a given population. On the other hand, characterizing spectral changes in congenital blindness may also contribute to better understanding brain reorganization and adaptive behavior, potentially yielding important implications for clinical applications.

Functional relevance of endogenous resting-state brain rhythms

One question that may arise concerns the value of using resting-state data in researching cognitive phenomena. Despite Hans Berger's pioneering observations in the human EEG, emphasizing the pervasive nature of intrinsic brain activity (Berger, 1929), the majority of (early) cognitive neuroscience research primarily concentrated on reactive or task-related brain activity. Researchers meticulously controlled experimental stimuli and settings to enhance the signal to noise ratio by averaging over single trials and eliminating intrinsic activity (Deco et al., 2011; Raichle, 2015). Resulting from this emphasis on evoked activity, endogenous activity was even considered noise (Pinneo, 1966).

Interestingly, functional neuroimaging studies revealed that brain activity during resting state is not random but rather characterized by structured large-scale functional connectivity patterns, arising from temporally correlated activation of constituent brain regions. Various different networks have been delineated since, with the default mode network emerging as particularly prominent during resting behavior (Greicius et al., 2003; Fox and Raichle, 2007; Smith et al., 2009; Raichle, 2015). Initially, some speculated that low-frequency resting-state networks might originate from non-neural (vascular, cardiac, or respiratory) sources that are also picked-up by the fMRI signal (Mayhew et al., 1996; Birn et al., 2006; Chang et al., 2009). However, similar networks have been detected in resting-state recordings employing EEG and MEG, modalities that measure the

electrophysiological activity of neurons more directly. These spectro-temporal signatures are predominantly observed within the frequency range of 10 to 30 Hz, reflecting correlations of spectral power over time (Mantini et al., 2007; de Pasquale et al., 2010; Brookes et al., 2011; Hillebrand et al., 2012; Hipp et al., 2012). Notably, resting-state networks have been associated with functional networks supporting cognitive and perceptual brain processes (Greicius et al., 2003; for similar results in monkeys, see Vincent et al., 2007; Smith et al., 2009), with variability in evoked responses being partially explained by individual intrinsic activity at rest (for similar results in cats, see Arieli et al., 1996; Fox et al., 2006). Together, resting-state networks are posited to reflect the functional neural architecture that subserves cognition (Singer, 2013).

In addition to characterizing large-scale spectral connectivity, individual brain areas' spectral resting-state signatures, i.e. spectral profiles, have been examined. Some studies have identified the most dominant frequency within brain regions, revealing a systematic distribution of dominant frequencies across cortex (Groppe et al., 2013; Mahjoory et al., 2020). Other studies, including studies 1 and 3 of this thesis, provided a more comprehensive understanding of the spectral landscape within individual brain regions by illustrating the combination of various spectral signatures (Keitel and Gross, 2016; Komorowski et al., 2018; Lubinus et al., 2021). This approach revealed that brain regions engage in different spectral modes over time, with individual clusters reflecting dominant spectral signatures at different timepoints. This more localized perspective is relevant because while many cognitive processes involve distributed networks, subprocesses likely occur locally within specific brain areas, necessitating what has been described as "global integration of local integrators" (Park and Friston, 2013).

Do spectral profiles measure endogenous brain rhythms?

Another question is whether spectral profiles extracted from resting-state data provide an effective means of quantifying endogenous rhythms. Neural oscillations are commonly defined as narrow spectral peaks in the power spectrum, measured in the absence of an external stimulation (Donoghue et al., 2020). In Studies 1 and 3, such spectral peaks were quantified in a pipeline including traditional spectral analyses, clustering algorithms, and normalization by the average power spectrum across the brain, mitigating the 1/f dynamics inherent to power spectra. At the group level distinctly defined power peaks were observed in many brain areas (Study 1), including auditory and speech-motor regions (Study 1 and 3). This ability to successfully characterize group-level profiles suggests considerable spectral similarity amongst individuals. Notably, clear power peaks in the theta frequency band (Study 3) were discernible already at the individual level, with features of the spectral clusters varying across subjects. This variability was expressed as heterogenous peak frequencies, amplitude and temporal persistence of each cluster.

The findings from studies 1 and 3 indicate the presence of endogenous theta rhythms within auditory cortex. Brain signals are often assumed to reflect these endogenous brain rhythms, even when recorded during perceptual or cognitive tasks, where distinguishing between evoked and endogenous activity is challenging (Zoefel et al., 2018b). Since the endogenous brain rhythms were extracted from resting-state activity in the current thesis, the spectral clusters likely represent endogenous activity. However, given the global nature of MEG recordings, reflecting superimposed activity of many thousands of neurons, it remains difficult to ascertain whether such spectral peaks signify “true” neural oscillations. Nonetheless, the current findings align with animal research (Lakatos et al., 2005) and intracranial studies in patients (Morillon et al., 2012; Halgren et al., 2018). While local field potentials also measure activity of many neurons, they operate at considerably smaller scales than EEG/MEG and can arguably reflect true “oscillations-in-process" (Van Bree et al., 2024). For instance, Morillon et al. (2012) recorded resting-state activity in auditory cortex of a patient using depth electrodes and observed that it was dominated by 5-6 Hz and 7.5 Hz theta activity in the right and left hemisphere, respectively. Similarly, Halgren et al. (2018) identified endogenous delta and theta activity in temporal, frontal, and parietal cortical areas, with local generators predominantly located in superficial cortical layers. This endogenous slow frequency activity was observed not only during resting behavior but also during sleep.

Endogenous rhythms shed light on oscillatory mechanisms in speech processing

Establishing the functional contribution of endogenous rhythms in task-related processing is pivotal for understanding whether perceptual and cognitive processes engage oscillatory mechanisms. Neural oscillations are expected to reflect active phenomena, meaning fluctuations of neural excitability are present in the absence of an external stimulation or task-related processing (Giraud and Poeppel, 2012; Van Bree et al., 2022). So far, evidence supporting this link is limited and mostly indirect. In the context of speech processing, Keitel and Gross (2016) demonstrated increased theta power amplitude from resting-state activity to speech tracking during passive speech listening. In contrast, alpha power decreased. Additionally, Becker and Hervais-Adelman (2023) found that theta power during normal speech comprehension correlated with enhanced word-in-noise recognition. Study 3 expands upon these findings, providing evidence for an oscillatory speech processing mechanism. Specifically, the endogenous theta peak frequency in pSTG, measured during resting state, predicted speech tracking in the same area. In other words, individual temporal characteristics of auditory association cortex were related to its ability to synchronize with incoming speech signals. While the direction of this effect differed from our initial hypotheses, these findings support the relevance of the endogenous theta brain rhythm in pSTG in speech tracking.

In Study 3, extracting single-subject theta clusters posed the challenge of cluster selection due to the prevalence of multiple theta clusters in many individuals. While it is conceivable that brain areas

exhibit multiple rhythms with similar eigenfrequencies, the observation of multiple clusters may also result from the clustering method employed. Here, we opted to select the cluster with the highest amplitude, based on the rationale that lower amplitude clusters may reflect more noise. Nevertheless, alternative selection criteria are viable. For instance, assuming that each individual theta cluster reflects genuine oscillatory activity in the theta frequency range, averaging these clusters—perhaps weighted by the duration in they dominate the brain regions' spectral signature—could be considered. Alternatively, the cluster most closely aligned with an individual's behavioral performance could be selected (Gunasekaran et al., 2023). However, particularly the latter option requires caution to avoid cherry-picking. An important question for future work will be to discern whether multiple spectral peaks within a frequency band reflect physiology or result from the clustering methods. Analyses akin to the one outlined here (or by Gunasekaran et al. (2023)) provide a valuable starting point for gaining insights into the variability of spectral signatures across individuals and over different timepoints within a brain region. Nonetheless, these findings emphasize the importance of establishing standardized peak selection criteria and considering potential refinements to current analysis pipelines. Connecting individual rhythms to behavioral performance may be essential to understand the functional relevance of endogenous rhythms in speech perception, as well as cognition more broadly.

6.3 Integration of top-down predictions in speech segmentation

Studies 2 and 3 of this thesis revealed the contribution of both behavioral and neural estimates of the preferred rate of the motor system and auditory-motor synchronization in predicting speech comprehension and speech tracking (as indicator of speech segmentation), respectively. These findings align with and expand upon prior research (Assaneo et al., 2021), supporting an oscillatory model of speech segmentation that engages capacities of the motor system alongside the auditory cortex. Previous research (Haegens and Zion Golumbic, 2018; Rimmele et al., 2018; Meyer et al., 2020) emphasizes the relevance of understanding how top-down predictions from the motor system (likely related to temporal predictions) and linguistic processes are integrated and interact with an oscillatory speech segmentation process.

Study 3 suggests that the motor system influences speech segmentation through an oscillatory mechanism, as endogenous peak frequency effects of SMA and IFG (triangularis) were observed to affect speech tracking in auditory cortex (HG, pSTG). Top-down predictions may affect speech tracking through neural gain control, as recently suggested for linguistic predictions (ten Oever and Martin, 2021), or through top-down phase-resetting of delta and theta rhythms via precentral gyrus and a fronto-temporal network, respectively, during listening to continuous speech (Park et al., 2015). Surprisingly, we found no main effects of phase-phase auditory-motor coupling (i.e., HG-

IFG, HG-SMA, pSTG-IFG, pSTG-SMA) on tracking. Instead, phase-phase auditory-motor coupling showed interaction effects with the endogenous motor peak frequencies, suggesting stronger influence of the endogenous frequency in individuals with stronger auditory-motor coupling. This suggests a more intricate auditory-motor network rather than direct phase-resetting between these brain areas (pSTG, IFG, SMA) during speech comprehension. Furthermore, the peak frequency of endogenous theta brain rhythms in SMA was related positively to coupling, i.e. stronger coupling was accompanied by faster peak frequencies, while the opposite trend was observed for the peak frequency of IFG, i.e. stronger coupling was accompanied by lower IFG frequencies. The primary objective of Study 3 was to investigate whether these speech-motor areas engage in speech segmentation through an oscillatory mechanism by showing effects of the peak frequency of endogenous theta brain rhythms. Consequently, the study did not prioritize exploring the interactions of a more complex auditory-motor network. This limitation highlights the need for future research. A potential explanation for the absence of direct coupling effects, consistent with previous research (Tourville and Guenther, 2011; Hertrich et al., 2016; Zuk et al., 2022), is that the impact of IFG and SMA is indirect via precentral gyrus or (pre)motor cortex. Another possibility is that phase-resetting of theta rhythms in auditory cortex is mediated through delta or beta rhythms in motor cortices (Arnal and Giraud, 2012; Park et al., 2015; Morillon and Baillet, 2017; Morillon et al., 2019).

The engagement of the motor system during auditory and speech perception is proposed to reflect predictive timing (Arnal and Giraud, 2012). This phenomenon entails heightened sensitivity of sensory brain areas, facilitating the processing of forthcoming sensory events by minimizing uncertainty about their temporal onset. As such, the top-down influence of predictive timing aids in distinguishing relevant and irrelevant information within a speech signal by contributing to the alignment of high excitability phases with informative moments in the signal (Zoefel and VanRullen, 2015). In the current study, we investigated involvement of the motor system during continuous speech comprehension, which likely reflects both top-down predictive and bottom-up feedforward processes.

Contextual and linguistic predictions facilitate speech processing

Predictive timing represents just one facet of top-down predictions theorized to shape speech comprehension. It is widely acknowledged that the comprehension of speech and language is also influenced by high-level predictions, which encompass abstract conceptual or linguistic elements such as semantics, syntax, or grammatical form. This aligns with the results from Study 2, where speech comprehension was predicted by a measure of sentence-level predictability, indicating better comprehension for more predictable sentences.

A consistent finding is that listeners and readers predict upcoming information and words during language comprehension and that these contextual predictions facilitate word recognition (Miller

and Isard, 1963; Tulving and Gold, 1963; Fischler and Bloom, 1979; Grosjean, 1980; Forster, 1981; Marslen-Wilson et al., 1988; Smith and Levy, 2013). Such predictions are proposed to be probabilistic (Kuperberg and Jaeger, 2016; Heilbron et al., 2022), constrained by an internal language model informed by a language's phonotactic, semantic, and syntactic rules (Spivey-Knowlton et al., 1993; Hagoort et al., 2004 p.200; Lee and Federmeier, 2009). A prominent hypothesis posits that speech recognition is facilitated because plausible candidates for upcoming units are pre-activated based on contextual cues (Federmeier and Kutas, 1999; DeLong et al., 2005; Federmeier, 2007; Kutas and Federmeier, 2011). The mechanism behind this facilitation, whether it operates in a bottom-up manner with top-down predictions selecting the appropriate candidate, or is predominantly a top-down process, remains a topic of ongoing debate (Gwilliams et al., 2023).

Extensive research has established that the brain evaluates contextual predictions upon the reception of new sensory input. A prominent EEG marker in this context is the N400, which has been shown to differentiate between predictable and unpredictable words. Violations of contextual expectations or less expected words are typically associated with higher N400 amplitudes (Kutas and Hillyard, 1980, 1984; Friederici et al., 1993; Hagoort et al., 2004; Frank et al., 2015; Kuperberg and Jaeger, 2016; Broderick et al., 2018; Nieuwland et al., 2020). While the N400 is commonly interpreted as indicator of predictive processing, the relationship is only indirect, as the component reflects late brain responses following the presentation of a predicted word (Van Petten and Luka, 2012; Kuperberg and Jaeger, 2016; Molinaro et al., 2016). However, recent studies were able to show predictive effects more directly, revealing that high-level predictions can influence low-level phonetic processing and representation (Broderick et al., 2019; Donhauser and Baillet, 2020; Gwilliams et al., 2022), aligning with the assumptions of the predictive coding framework (Rao and Ballard, 1999; Friston, 2005). While most of the research demonstrating higher-level predictions has primarily focused on ambiguous or noisy speech settings, these recent studies underscored that contextual prediction facilitate speech processing also under optimal and more naturalistic (continuous speech) listening conditions (Broderick et al., 2019; Donhauser and Baillet, 2020; Gwilliams et al., 2022).

To estimate the influence of predictability on sentence-level comprehension, many studies contrast high and low predictable words as measured by the cloze probability of the sentence final word (i.e. target)(Taylor, 1953; Obleser and Kotz, 2010; Nieuwland et al., 2020). In such a paradigm, sentences are typically designed for this special purpose, systematically varying along this stimulus dimension. In Study 2, we examined prediction effects in more naturalistic speech, presenting sentences that naturally varied in their predictability without introducing a manipulation. We observed facilitated comprehension for more predictable sentences, suggesting that also subtle differences in predictability differentiate comprehension, which is in line with recent work (Broderick et al., 2019;

Donhauser and Baillet, 2020; Gwilliams et al., 2022). Furthermore, the present work extends previous findings quantifying the effect of sentence-level predictability, integrating the predictability of all words instead of focusing on the predictability of the final word, on comprehension.

Towards more plausible models of speech tracking, including modulatory top-down effects

While the influence of higher-level linguistic processes on speech comprehension is widely recognized, the neural mechanisms underpinning this influence, particularly concerning early neural processes like speech segmentation, remain unclear. Oscillatory models of speech segmentation have emphasized the alignment of endogenous brain rhythms with incoming speech signals, irrespective of top-down effects (Giraud and Poeppel, 2012). However, evidence suggests that higher-level processes such as speech intelligibility (Peelle and Davis, 2012; Gross et al., 2013; Peelle et al., 2013; Park et al., 2015; Rimmele et al., 2015; Steinmetzger and Rosen, 2017), attention (Zion Golumbic et al., 2013 p.2; Rimmele et al., 2015), perceptual relevance (Keitel et al., 2018), or motor production-related temporal predictions (Park et al., 2015) can modulate speech tracking. Thus, a comprehensive model of speech segmentation is likely to encompass both oscillatory speech tracking in auditory cortex and the influence of top-down factors, including temporal and linguistic predictions, as previously suggested (Haegens and Zion Golumbic, 2018; Rimmele et al., 2018; Meyer et al., 2020). Recent accounts have begun to bridge this gap by extending oscillatory models to incorporate temporal or linguistic predictions. These accounts propose mechanisms through which rigid oscillations may adapt to accommodate more complex predictions (beyond isochronous patterns), thereby facilitating speech segmentation as evidenced by increased speech-to-brain synchronization.

One line of research advances an oscillatory model wherein temporal predictions originating from the speech production motor cortex influence speech perception. Specifically, behavioral auditory-motor synchronization determines performance in statistical learning, syllable and rate discrimination (Assaneo et al., 2019, 2021; Kern et al., 2021), with stronger synchronization being associated with enhanced performance. Assaneo, Rimmele et al. (2021) substantiated their behavioral results with a mathematical model, reinforcing the notion of an oscillatory mechanism governing auditory-motor integration in speech perception. This model represented auditory and motor cortex as a system of coupled phase oscillators with bidirectional coupling of varying coupling strengths. The model accurately simulated the performance in the syllable discrimination task for both high and low synchronizers only when accounting for their distinct coupling strength measured by the SSS-test.

Consistent with these findings, we provide neurophysiological evidence corroborating the involvement of oscillatory auditory and motor units in facilitating speech segmentation. Specifically, Study 3 elucidated the relationship between the auditory cortex (pSTG) endogenous theta brain rhythm

(peak frequency) and speech tracking in pSTG. Additionally, we revealed that speech tracking (in pSTG and HG) was predicted by the endogenous theta brain rhythms of speech-motor cortex (peak frequencies of IFG, SMA), but only in high synchronizers. Extending the previous work, we measured the preferred rates of individual auditory and motor cortex brain rhythms (peak frequencies) as predictors of the speech tracking rather than relying on approximations across individuals based on previous literature. Assaneo, Rimmele et al. (2021) underscored the relationship between auditory-motor synchronization and perceptual sensitivity in syllable discrimination (a task which they assumed involves syllable segmentation). In line with this, our results indicate that phase-phase auditory-motor coupling, in interaction with the preferred rates of endogenous auditory and motor brain rhythms, contributed to predicting speech tracking in individuals with higher auditory-motor synchronization. While Assaneo, Rimmele et al. (2021) focused on the impact of auditory-motor coupling in the context of syllable discrimination, we extend their findings by demonstrating coupling effects on speech segmentation during continuous speech comprehension. Furthermore, while the model of Assaneo et al. (2021) proposes bidirectional coupling between auditory and motor cortices, the current results cannot clarify the directionality of coupling.

In an attempt to elucidate the influence of linguistic predictions on speech segmentation, some approaches combined artificial neural networks, approximating internal language models, with oscillatory computational models or neurophysiological data. Donhauser and Baillet (2020) were able to demonstrate that speech tracking can be modulated by the internal language model of a listener. Specifically, the authors estimated contextual information in continuous narratives by quantifying the predictive distribution of upcoming phonemes, i.e. uncertainty, and the unexpectedness of the actually presented phoneme, i.e. surprisal. This work revealed that theta speech tracking, measured as a temporal response function, was modulated by uncertainty, with a tracking gain for more uncertain events. These results suggest an interplay of tracking of the acoustic stimulus and semantic expectations during speech processing that automatically upregulates sensory processing resources in uncertain contexts. Ten Oever and Martin (2021) specified that linguistic information may mitigate the quasi-rhythmicity in speech (syllabic or word rate) by interacting with excitability phase modulations of an oscillatory process through gain control. The authors combined an oscillatory computational model with predictions of upcoming words from a recurrent neural network. They demonstrated that an oscillatory auditory model can explain adaption to quasi-rhythmicity in speech if gain control through word predictions is taken into account. By showing that gain modulation through word predictions was inherently tied to temporal information, these results suggest that higher-level and temporal information are intertwined.

In addition to the main effects, Study 2 revealed an interaction effect of linguistic predictability and auditory-motor synchronization. This interaction effect suggests that individuals with high auditory-

motor synchronization (possibly related to enhanced temporal processing) more effectively used predictions from the linguistic context. As such, Study 2 provides tentative support for the facilitation of speech processing through the combination of (sentence) predictability and motor-based predictions at the level of speech comprehension. Generally, stronger auditory-motor synchronization is proposed to relate to enhanced temporal predictions which likely facilitates speech processing. Accordingly, the interaction effect observed in Study 2 may suggest that high synchronizers do not only utilize linguistic and temporal predictions in parallel but in a synergistic manner. Such an interaction of production-based and linguistic predictions is in line with theoretical proposals (Pickering and Gambi, 2018). While our behavioral data cannot directly speak to an oscillatory speech segmentation process, we cautiously interpret it as such given the parallel findings of Study 2 and the speech tracking results in Study 3. This interpretation aligns with initial evidence suggesting that one's individual prediction tendency predicts speech tracking strength (Schubert et al., 2023). It is possible that temporal and semantic (or other linguistic) predictions present as different but interacting neural mechanisms. Considering the proposal by ten Oever and Martin (2021), one could speculate that during continuous speech comprehension semantic predictions enhance processing via gain modulation, even when the theta phase is suboptimal at stimulus arrival. Additionally, temporal predictions likely enhance speech processing via phase-resetting slow rhythms in auditory cortex, thereby aligning processing capacities to acoustic events (Arnal and Giraud, 2012; Park et al., 2015; Morillon and Baillet, 2017; Morillon et al., 2019). Nevertheless, we acknowledge that the interpretations in this paragraph are highly speculative and thus to be taken with caution. Study 2 did not test directly whether high and low synchronizers differ in their temporal prediction ability but such different abilities are assumed based on the difference in auditory-motor synchronization behavior (which has been typically related to temporal predictions, Repp, 2005; Repp and Su, 2013), as well as previously shown differences in task processing (Assaneo et al., 2021; Kern et al., 2021; Barchet et al., 2024). Moreover, even if temporal predictability capabilities differ according to this reasoning, it cannot be ruled out that other factors are contributing to the interaction effect with linguistic predictability. Finally, to answer whether motor-based temporal predictions and linguistic predictions (at various levels) facilitate comprehension by enhancing segmentation, further neurophysiological research is required.

6.4 Considering individual differences in auditory-motor coupling and speech processing

This thesis demonstrates that features of the motor system (i.e. endogenous rates) and auditory-motor synchronization influence speech processing. The results underscore the critical importance of accounting for individual differences in endogenous motor rates, auditory-motor coupling, and speech tracking. Notably, we observe that the degree to which the motor system contributes to comprehension varies amongst individuals, with enhanced engagement in those exhibiting

stronger auditory-motor synchronization (as measured by our behavioral spontaneous speech perception production synchronization test). This aligns with previous research (Assaneo et al., 2019, 2021; Kern et al., 2021; Barchet et al., 2024). Studies 2 and 3 expand on these findings in various ways: We demonstrate that individual differences in auditory-motor synchronization predict continuous speech comprehension, complementing previous work showing similar effects during syllable sequence processing (Assaneo et al., 2019). Additionally, in individuals with high vs. low auditory-motor synchronization, the auditory and motor systems contributed distinctly to speech processing. Specifically, in individuals with high synchronization, both the spontaneous speech motor production rate and the preferred auditory rate predicted speech comprehension, whereas, in those with low synchronization, only the preferred auditory rate had predictive value. A similar pattern emerged on the neural level: in individuals with high synchronization, the endogenous theta brain rhythms in pSTG, IFG, and SMA predicted speech tracking in pSTG, while such an effect was observed only for the theta brain rhythms of pSTG in those with low synchronization. Previous behavioral (Assaneo et al., 2021; Kern et al., 2021; Barchet et al., 2024) and neural (Assaneo et al., 2019) research posited an enhanced recruitment of the motor system during speech processing in individuals with high auditory-motor synchronization, yet direct evidence was lacking. Together, individuals with stronger auditory-motor synchronization behavior not only demonstrated stronger interactions between their auditory and motor systems but also more robustly engaged their motor system during speech processing. Finally, the finding that individual preferred auditory and motor rates, measured both behaviorally and neurally, predict speech segmentation, provides evidence supporting a neural oscillatory mechanism of speech processing, as proposed by behavioral and computational modelling results (Assaneo et al., 2021). Previous work suggests that certain populations, such as kids, older adults, and individuals with specific language impairment, may exhibit reduced recruitment of the production system for prediction (in that case semantic prediction), potentially leading to less efficient speech perception (Pickering and Gambi, 2018). Our findings extend this framework by suggesting that individual differences in the amount of motor cortex recruitment furthermore exist in healthy young adults, affecting speech processing.

Different strategies for successful speech comprehension

Given the differential recruitment of the auditory and motor systems, our results suggest that individuals with high versus low auditory-motor synchronization employ different strategies during speech comprehension. Study 3 suggests that low synchronizers may heavily rely on their auditory system, possibly leading to closer monitoring of the auditory signal and a greater emphasis on bottom-up processing. This would be in line with observations in individuals with non-fluent variant primary progressive aphasia, where atrophied speech-motor cortex but intact structure and function of temporal cortex resulted in increased activation in parts of right pSTG during an audio-visual

speech perception task (Cope et al., 2023). Conversely, high synchronizers, alongside auditory processing, displayed effects of the spontaneous speech motor production rate and the endogenous rhythms in IFG and SMA, at the behavioral and neural levels, respectively. Despite modest effect sizes, these motor effects during speech processing were consistent across experiments (Study 2, experiment 1 and 2 and Study 3) and modalities (behavior and MEG), with behavioral and tracking advantages for the population displaying stronger motor recruitment.

The two groups may further diverge in their usage of non-linguistic cognitive capacities (e.g. working memory) and reliance on non-temporal predictions during speech processing. High synchronizers excel in auditory tasks requiring precise temporal processing, likely due to enhanced motor-based temporal prediction abilities, which may interact synergistically with linguistic predictions in speech (for discussion, see 6.3). In contrast, low synchronizers demonstrate poorer auditory behavior across various auditory tasks related to reduced motor-based predictions, including continuous speech comprehension. Nevertheless, the latter group proficiently comprehends speech in everyday contexts, suggesting the presence of alternative compensatory mechanisms. Study 3 indicates that working memory may serve as one such compensatory mechanism, as it predicted speech comprehension in low but not high synchronizers. Working memory has been shown to enhance perceptual acuity (Anvari et al., 2002; Strait et al., 2011; Kraus et al., 2012; Politimou et al., 2019; Kim et al., 2023) and to contribute to rhythm discrimination independently of auditory-motor synchronization (Kim et al., 2023). Importantly, the relative contribution of rhythmic auditory-motor abilities and working memory in rhythmic music and language perception supposedly shifts during development, although the direction is not yet clear. While Kim et al. (2023) observed a shift from auditory-motor synchronization to working memory from younger to older children, Anvari et al. (2002) suggested older children rely more on temporal parameters, due to increased temporal precision. Study 3 aligns with the latter notion, as speech comprehension was predicted more strongly by working memory capacity in those with lower motor-based temporal predictions (i.e. low synchronizers). This suggests that working memory may serve a compensatory role rather than interact synergistically with motor-based temporal predictions. Such a compensatory function could be particularly beneficial in demanding working memory situations, as observed in our study with long sentences and without situational context. Future research could explore whether these differential effects of working memory extend to simpler contexts. However, it is possible that differences between high and low synchronizers are more pronounced in challenging listening conditions and thus may not be evident in overly simple tasks.

IFG and SMA contribute to speech segmentation, but in opposite directions

The observation that both IFG and SMA contribute to speech tracking is not surprising, given their established roles and connectivity in speech and language processing. IFG, part of the core

language network, is implicated in various linguistic tasks (Hickok and Poeppel, 2007; Friederici and Gierhan, 2013; Hagoort, 2014). In contrast, SMA, likely a node in the language association network, has recently been acknowledged to support speech perception alongside its well-established role in speech production (Hertrich et al., 2016). These regions are directly connected via the intra-lobar frontal aslant tract (Dick et al., 2014; Hertrich et al., 2016; Dragoy et al., 2020). This network is associated with timing and sequencing aspects during spontaneous speech fluency, with different emphasis on temporal scales for subparts of both IFG and SMA (Hertrich et al., 2016). Specifically, IFG (BA45) and pre-SMA support supra-second processes like metrical timing and sequential processing. Conversely, IFG (BA44) and SMA-proper are associated with sub-second temporal processes (Hertrich et al., 2016; Dragoy et al., 2020; Zuk et al., 2022).

In Study 3, the theta peak frequency of SMA positively predicted speech tracking, with higher frequencies associated with stronger tracking. This parallels the behavioral findings of studies 2 and 3, wherein individuals with faster speech production rates displayed better speech comprehension. This supports our hypothesis that higher endogenous theta frequencies facilitate segmentation through enhanced temporal resolution. The motor system's assumed role in an oscillatory model of speech processing involves providing temporal predictions (Schroeder et al., 2010; Arnal and Giraud, 2012; Morillon and Baillet, 2017), making the SMA a plausible candidate given its role in sensory and sensorimotor processing of temporal structure (Schwartze et al., 2012a). Moreover, the SMA supports sequence processing (Schwartze et al., 2012a) and is implicated in naturalistic, sentential speech rather than word-level processing (Adank, 2012), highlighting its role in procedural processing. Some have proposed that such procedural processing may be linked to top-down predictive processing (Hertrich et al., 2016). Its connectivity with cerebellar-thalamic and basal ganglia-thalamic circuits, central to temporally structure human behavior (Schwartze et al., 2012a), further support SMA's suitability for temporal predictions. Furthermore, SMA likely supports speech intelligibility, especially during challenging listening conditions. Its role in both speech production and perception support the notion of shared networks during these processes (Adank, 2012). Note that while the temporal processing aspects relate particularly to pre-SMA, the current findings cannot distinguish between pre-SMA and SMA-proper.

IFG displayed the opposite effect with higher endogenous frequencies predicting lower speech tracking. This observation contradicted our hypothesis. However, the same direction was observed for the endogenous frequency effect of pSTG, with lower pSTG theta frequencies predicting higher speech tracking. One possible explanation for the similar effects could be the tight connection between IFG and pSTG or "auditory/phonetic cortex" (Hertrich et al., 2016). While SMA is connected with auditory/phonetic cortex via premotor cortex, the temporo-parietal interface, and IFG, there are direct connections from IFG to auditory/phonetic cortex.

A plausible explanation for these contrasting negative and positive effects of IFG and SMA, respectively, could be a differentiation between inhibitory and excitatory functions. Supporting this notion, IFG (along with precentral gyrus) has been proposed to govern the inhibitory control of speech, evidenced by increased high-gamma activity during abrupt, voluntary stops in speech production (Zhao et al., 2023). Furthermore, left IFG supposedly reconciles perceptual predictions during speech perception and collaborates with precentral gyrus in creating prediction errors in case of mismatching predictions and sensory input (Cope et al., 2023), potentially entailing inhibition of unlikely units. Finally, higher hippocampal theta frequencies have been linked to stronger excitation during working memory processes (Roux et al., 2022). However, contrary to a faciliatory effect of SMA, some findings propose that both IFG and pre-SMA are part of the "frontal inhibitory control network" (Sharp et al., 2010; Swann et al., 2012; Aron et al., 2014). The current work cannot disentangle the potential disparity in inhibitory and excitatory functions of IFG and SMA, however, it presents an interesting avenue for future investigations. One possible approach could involve a more fine grained parcellation of both IFG and SMA, as recent intracranial work underscores that different inhibitory and excitatory may coexist within a single region (Zhao et al., 2023).

6.5 Limitations

This work specifically focuses on the "preferred frequency" of brain rhythms, defined as the frequency of spectral peaks. Oscillatory models of speech perception assume the engagement of endogenous brain rhythms in speech segmentation (Giraud and Poeppel, 2012). The preferred frequency is particularly relevant in this context because the success of synchronization between two oscillators or an oscillator and an external signal depends on the similarity of their eigenfrequencies (Fröhlich, 2015). Spectral analyses typically assume that oscillatory brain activity can be approximated by narrowband, sinusoidal fluctuations. However, recent research underscores the importance of also considering aperiodic brain activity (Donoghue et al., 2020) and the shape of the waveforms of ongoing activity (Cole and Voytek, 2017). Both aperiodic brain activity and non-sinusoidal waveforms have been proposed to reflect physiological properties of the underlying generators (Manning et al., 2009; Buzsáki et al., 2012). Furthermore, evidence suggests they vary as a function of task demand, perceptual and cognitive states (Buzsáki et al., 1985; He et al., 2010; Trimper et al., 2014), as well as in disease (Ouedraogo et al., 2016; Cole et al., 2017; Robertson et al., 2019; Molina et al., 2020). While this thesis linked the peak frequency of brain rhythms to speech processing, a comprehensive understanding of the neurophysiology of speech processing likely requires additionally considering aperiodic brain activity and waveform shape.

In Study 3, we quantified speech tracking using gaussian-copula mutual information (GCMI), an information theoretic measure quantifying the statistical dependencies between timeseries

(Shannon, 1948; Ince et al., 2017). GCMI provides a computationally efficient multivariate statistical framework, free of distributional assumptions for the variables (Ince et al., 2017; Chalas et al., 2022). While GCMI offers several advantages, numerous other techniques exist for quantifying brain-to-stimulus synchronization, such as phase-locking value, coherence, or phase-lag-index. Particularly in the context of naturalistic stimuli, there has been recent interest in system identification or linear modelling approaches. These methods model the impulse function, or temporal response function (TRF), exhibited by the brain to selected stimulus features (Crosse et al., 2016, 2021). However, while TRFs can capture the temporal evolution in neural responses elicited by speech (and other) features, they may not necessarily account for the temporal properties of the speech signal itself (Zoefel and Kösem, 2024). Given the particular interest of this thesis in the temporal alignment of the auditory cortex to speech, we used GCMI.

MEG is a method valued for its excellent temporal resolution, combined with decent source reconstruction. Nevertheless, compared to neuroimaging techniques, the localization is limited, so that the anatomical regions under investigation in this work have to be taken with caution. Furthermore, for brain parcellation we used the Automated Anatomical Labelling (AAL) atlas (Tzourio-Mazoyer et al., 2002; Rolls et al., 2020), which divides the brain into 116 brain areas and is widely used in neuroimaging research. One limitation of this choice, is the inability of distinguishing SMA-proper from pre-SMA. To further specify the contribution of SMA to speech segmentation, follow-up analyses could use a different atlas for parcellation.

Finally, a central assumption in this thesis is that behavioral preferred rates are the behavioral readouts of neural processes. Study 3 combines behavioral and neural estimates of the preferred and endogenous frequencies of the auditory and motor systems and revealed parallel effects across both measurements. Particularly, individuals with stronger behavioral auditory-motor synchronization, demonstrated enhanced recruitment of motor parameters on speech comprehension (behavior) and speech segmentation (brain). In contrast, individuals with lower auditory-motor synchronization relied more strongly on their auditory systems. Despite such parallel effects and contrary to our hypotheses, we observed no correlation between behavioral and neural measures. One possibility is that, although endogenous brain rhythms and speech tracking contribute to speech comprehension, they only reflect some of the multiple processes influencing human behavior.

6.6 Impact

The current thesis emphasizes the presence of individual differences in the recruitment of the motor system and the role of brain rhythms in the interaction between the auditory and motor systems during speech comprehension. The motor system exhibits notable susceptibility to training, as exemplified by the heightened auditory-motor synchronization observed in musicians (Zatorre et al.,

2007; Chen et al., 2008b). The findings presented here offer valuable insights into the processing level and neural mechanism underlying the effects of auditory-motor interaction, thereby advancing existing models of speech perception.

In addition to its contribution to fundamental research, the current findings may hold implications for clinical research, particularly for conditions including temporal processing deficits. Such deficits are evident, for example, in individuals with hearing impairments using cochlear implants (CI) or other hearing aids, as well as in children with developmental speech disorders or dyslexia. These populations often exhibit diminished performance in speech processing tasks, particularly when temporal processing is challenging, such as at faster syllabic rates (Guiraud et al., 2018). Beyond deficits in rhythmic auditory processing, impairments also occur in rhythmic motor processing among both CI users and children with specific language impairment, as evidenced by reduced tap-to-beat synchronization (Corriveau and Goswami, 2009; Phillips-Silver et al., 2015; Hidalgo et al., 2021). Neural evidence suggests that such behavioral deficits in CI users may be related to slowed or inefficient neural processing. Specifically, while speech tracking, quantified using TRFs, appears similar or even stronger in both pediatric and adult CI users compared to normally hearing controls, TRF latencies are longer (Gillis et al., 2022; Federici et al., 2024). Furthermore, elderly CI users present with altered connectivity patterns, particularly in a fronto-temporal network during speech perception (Bidelman et al., 2019). To enhance speech processing, both individuals with hearing impairments and children with developmental speech disorders may benefit from rhythmic and music training, as also noted by others (Patel, 2011; Hidalgo et al., 2021). The current findings highlight potential strategies to identify at what level processing is impaired in these populations, e.g. presenting as slowed individual auditory and motor rhythms or reduced auditory-motor coupling, which may help tailor interventions.

Another applied field potentially benefitting from the finding that individual differences in speech rate preferences influence speech comprehension is digital learning. Digital learning, involving audio or audio-visual materials and gamification, allows for personalized teaching materials beyond the scope of traditional classroom settings (Kebritchi et al., 2017; Shaikh and Asif, 2022). The results of the current thesis suggest that individual speech rate preferences relate to speech comprehension, thereby likely influencing learning outcomes. Optimal speech processing may align with one's individual preferred rate (Fröhlich, 2015; Notbohm et al., 2016), highlighting the potential advantages of dynamically adjusting the speech rate of learning materials. Preliminary evidence supports this, showing that video instructors' speech rates influence video popularity and student engagement. Some studies noted benefits for faster average speech rates (Guo et al., 2014; Ten Hove and van der Meji, 2015), while others have found that students' attitude and attention were most positively affected by individually adjustable speech rates (Díaz and Recabarren, 2024).

6.7 Conclusions

In three studies, this thesis elaborated on the question of whether neural speech processing is supported by an oscillatory model, engaging endogenous brain rhythms of both auditory and motor cortex, and their coupling. The question was approached by characterizing variation (due to plasticity-related changes) in the temporal dynamics of the preferred auditory and motor brain rhythms, as well as their interaction, and quantifying whether such variation affected both the comprehension and segmentation of speech.

We demonstrated plasticity of the identified preferred rhythms using two different approaches. Firstly, in a whole-brain data-driven approach probing plasticity due to life-long reorganization, HG and multiple other brain regions across cortex displayed spectral alterations in congenitally blind individuals. These alterations were observed particularly in the theta-to-beta frequencies. Secondly, a more nuanced approach capitalizing on individual differences revealed variability in behavioral preferred auditory rates amongst individuals which, notably, contributed to their speech comprehension ability. Neural data supported these effects, as individual differences in the peak frequencies of endogenous theta brain rhythms in pSTG affected speech tracking strength in pSTG, likely reflecting effects on syllabic segmentation. These results suggest that (1) endogenous auditory rhythms plastically adapt under different circumstances and, importantly, (2) that differences between individuals elucidate variation in speech processing, both behaviorally in terms of comprehension and neurally concerning speech tracking.

Endogenous motor rhythms not only influenced speech comprehension but also speech segmentation, measured as tracking. Both comprehension and segmentation were enhanced in individuals with stronger behavioral auditory-motor synchronization. Interestingly, these individuals also displayed a stronger relation of the preferred endogenous rates of the motor system to speech processing. In addition to motor parameters, comprehension was also affected by linguistic predictions and working memory capacity. The current results suggest that sentence predictability interacts synergistically with auditory-motor synchronization strength, while working memory may compensate for the absence or reduction of motor-based predictions.

In conclusion, the findings presented in this thesis endorse an oscillatory model of speech segmentation that leverages endogenous brain rhythms in both auditory and speech-motor cortex. The extend of motor system dynamics' contribution to the segmentation process may be contingent on the coupling strength between auditory and motor cortex, with modest yet consistent behavioral and neural processing advantages for individuals with stronger coupling. The role of both working memory and linguistic predictability underscores the complexity of speech processing, highlighting interesting future research directions aimed at uncovering a comprehensive account of speech processing.

7 References

Abadi M et al. (2016) TensorFlow: Large-scale machine learning on heterogeneous distributed systems. arXiv:160304467 [cs].

Abbasi O, Gross J (2020) Beta-band oscillations play an essential role in motor–auditory interactions. Hum Brain Mapp 41:656–665.

Abbasi O, Steingräber N, Chalas N, Kluger DS, Gross J (2023) Spatiotemporal dynamics characterise spectral connectivity profiles of continuous speaking and listening. PLoS Biol 21:e3002178.

Adank P (2012) The neural bases of difficult speech comprehension and speech production: Two Activation Likelihood Estimation (ALE) meta-analyses. Brain and Language 122:42–54.

Adrian ED, Matthews BH (1934) The Berger rhythm: potential changes from the occipital lobes in man. Brain 57:355–385.

Aguirre GK, Datta R, Benson NC, Prasad S, Jacobson SG, Cideciyan AV, Bridge H, Watkins KE, Butt OH, Dain AS, Brandes L, Gennatas ED (2016) Patterns of individual variation in visual pathway structure and function in the sighted and blind. PLoS ONE 11:e0164677.

Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM (2001) Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proceedings of the National Academy of Sciences 98:6.

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

Amedi A, Raz N, Pianka P, Malach R, Zohary E (2003) Early 'visual' cortex activation correlates with superior verbal memory performance in the blind. Nat Neurosci 6:758–766.

Andersson JLR, Jenkinson M, Smith S (2007) Non-linear registration, aka spatial normalisation. FMRIB Technial Report TR07JA2. (Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, Department of Clinical Neurology, Oxford University). 62.

Anurova I, Carlson S, Rauschecker JP (2019) Overlapping Anatomical Networks Convey Cross-Modal Suppression in the Sighted and Coactivation of "Visual" and Auditory Cortex in the Blind. Cerebral Cortex 29:4863–4876.

Anvari SH, Trainor LJ, Woodside J, Levy BA (2002) Relations among musical skills, phonological processing, and early reading ability in preschool children. Journal of Experimental Child Psychology 83:111–130.

Arai T, Pavel M, Hermansky H, Avendano C (1996) Intelligibility of speech with filtered time trajectories of spectral envelopes. Proceeding of Fourth International Conference on Spoken Language Processing 4:2490–2493.

Arieli A, Sterkin A, Grinvald A, Aertsen A (1996) Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. Science 273:1868–1871.

Arnal LH (2012) Predicting "when" using the motor system's beta-band oscillations. Front Hum Neurosci 6.

Arnal LH, Doelling KB, Poeppel D (2015) Delta–beta coupled oscillations underlie temporal prediction accuracy. Cerebral Cortex 25:3077–3085.

Arnal LH, Giraud A-L (2012) Cortical oscillations and sensory predictions. Trends in Cognitive Sciences 16:390–398.

Arnal LH, Wyart V, Giraud A-L (2011) Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. Nat Neurosci 14:797–801.

Aron AR, Robbins TW, Poldrack RA (2014) Inhibition and the right inferior frontal cortex: one decade on. Trends in Cognitive Sciences 18:177–185.

Assaneo MF, Poeppel D (2018) The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. Sci Adv 4:eaao3842.

Assaneo MF, Rimmele JM, Sanz Perl Y, Poeppel D (2021) Speaking rhythmically can shape hearing. Nat Hum Behav 5:71–82.

Assaneo MF, Ripollés P, Orpella J, Lin WM, de Diego-Balaguer R, Poeppel D (2019) Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. Nat Neurosci 22:627–632.

Aung WY, Mar S, Benzinger TL (2013) Diffusion tensor MRI as a biomarker in axonal and myelin damage. Imaging in Medicine 5:427–440.

Baayen RH, Davidson DJ, Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language 59:390–412.

Baillet S (2017) Magnetoencephalography for brain electrophysiology and imaging. Nat Neurosci 20:327–339.

Bangert M, Peschel T, Schlaug G, Rotte M, Drescher D, Hinrichs H, Heinze H-J, Altenmüller E (2006) Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. NeuroImage 30:917–926.

Barchet AV, Henry MJ, Pelofi C, Rimmele JM (2024) Auditory-motor synchronization and perception suggest partially distinct time scales in speech and music. Commun Psychol 2:2.

Baron SG, Osherson D (2011) Evidence for conceptual combination in the left anterior temporal lobe. NeuroImage 55:1847–1852.

Barr DJ, Levy R, Scheepers C, Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language 68:255–278.

Barzegaran E, Vildavski VY, Knyazeva MG (2017) Fine structure of posterior alpha rhythm in human EEG: Frequency components, their cortical sources, and temporal behavior. Sci Rep 7:8249.

Başar-Eroglu C, Başar E, Demiralp T, Schürmann M (1992) P300-response: possible psychophysiological correlates in delta and theta frequency channels. A review. International journal of psychophysiology 13:161–179.

Bastos AM, Briggs F, Alitto HJ, Mangun GR, Usrey WM (2014) Simultaneous recordings from the primary visual cortex and lateral geniculate nucleus reveal rhythmic interactions and a cortical source for gamma-band oscillations. Journal of Neuroscience 34:7639–7644.

Becker R, Hervais-Adelman A (2023) Individual theta-band cortical entrainment to speech in quiet predicts word-in-noise comprehension. Cerebral Cortex Communications 4:tgad001.

Bedny M (2017) Evidence from blindness for a cognitively pluripotent cortex. Trends in Cognitive Sciences 21:637–648.

Bedny M, Pascual-Leone A, Dodell-Feder D, Fedorenko E, Saxe R (2011) Language processing in the occipital cortex of congenitally blind adults. Proceedings of the National Academy of Sciences 108:4429–4434.

Berger H (1929) Ueber das Elektrenzephalogram des Menschen. Arch Psychit Nervenk 87:527–570.

Beukelman DR, Yorkston KM (1979) The relationship between information transfer and speech intelligibility of dysarthric speakers. Journal of Communication Disorders 12:189–196.

Bidelman GM, Mahmud MS, Yeasin M, Shen D, Arnott SR, Alain C (2019) Age-related hearing loss increases full-brain connectivity while reversing directed signaling within the dorsal–ventral pathway for speech. Brain Struct Funct 224:2661–2676.

Binder JR (2000) Human temporal lobe activation by speech and nonspeech sounds. Cerebral Cortex 10:512–528.

Binder JR, Desai RH, Graves WW, Conant LL (2009) Where Is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cerebral Cortex 19:2767–2796.

Binder JR, Frost JA, Hammeke TA, Bellgowan PSF, Rao SM, Cox RW (1999) Conceptual processing during the conscious resting state: A functional MRI study. Journal of Cognitive Neuroscience 11:80–93.

Birn RM, Diamond JB, Smith MA, Bandettini PA (2006) Separating respiratory-variation-related fluctuations from neuronal-activity-related fluctuations in fMRI. NeuroImage 31:1536–1548.

Bishop GeoH (1932) Cyclic changes in excitability of the optic pathway of the rabbit. American Journal of Physiology-Legacy Content 103:213–224.

Black JW (1951) The effect of delayed side-tone upon vocal rate and intensity. J Speech Hear Disord 16:56–60.

Blakemore S-J, Wolpert D, Frith C (2000) Why can't you tickle yourself? Neuroreport 11:R11–R16.

Blakemore S-J, Wolpert DM, Frith CD (1998) Central cancellation of self-produced tickle sensation. Nat Neurosci 1:635–640.

Blumstein SE, Baum SR (2016) Neurobiology of speech production. In: Neurobiology of Language, pp 689–699. Elsevier.

Boemio A, Fromm S, Braun A, Poeppel D (2005) Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat Neurosci 8:389–395.

Boersma P, Weenik D (2020) Praat: doing phonetics by computer [Computer program]. Version 6.0.40.

Boltz MG (1993) The generation of temporal and melodic expectancies during musical listening. Perception & Psychophysics 53:585–600.

Bradshaw JL, Nettleton NC, Geffen G (1971) Ear differences and delayed auditory feedback: Effects on a speech and a music task. Journal of Experimental Psychology 91:85–92.

Brennan JR (2022) Language and the brain: a slim guide to neurolinguistics. Oxford: Oxford University Press.

Broderick MP, Anderson AJ, Di Liberto GM, Crosse MJ, Lalor EC (2018) Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. Current Biology 28:803-809.e3.

Broderick MP, Anderson AJ, Lalor EC (2019) Semantic context enhances the early auditory encoding of natural speech. J Neurosci 39:7564–7575.

Brookes MJ, Woolrich M, Luckhoo H, Price D, Hale JR, Stephenson MC, Barnes GR, Smith SM, Morris PG (2011) Investigating the electrophysiological basis of resting state networks using magnetoencephalography. Proceedings of the National Academy of Sciences 108:16783–16788.

Brungart DS, van Wassenhove V, Brandewie E, Romigh G (2007) The effects of temporal acceleration and deceleration on AV speech perception. AVSP:27–34.

Buffalo EA, Fries P, Landman R, Buschman TJ, Desimone R (2011) Laminar differences in gamma and alpha coherence in the ventral stream. Proceedings of the National Academy of Sciences 108:11262–11267.

Bull R, Rathborn H, Clifford BR (1983) The voice-recognition accuracy of blind listeners. Perception 12:223–226.

Burton H (2003) Visual cortex activity in early and late blind people. J Neurosci 23:4005–4011.

Burton H, Snyder AZ, Raichle ME (2014) Resting state functional connectivity in early blind humans. Frontiers in Systems Neuroscience 8.

Buzsáki G (2004) Neuronal oscillations in cortical networks. Science 304:1926–1929.

Buzsáki G, Anastassiou CA, Koch C (2012) The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. Nat Rev Neurosci 13:407–420.

Buzsáki G, Draguhn A (2004) Neuronal oscillations in cortical networks. Science 304:1926–1929.

Buzsáki G, Logothetis N, Singer W (2013) Scaling brain size, keeping timing: evolutionary preservation of brain rhythms. Neuron 80:751–764.

Buzsáki G, Rappelsberger P, Kellényi L (1985) Depth profiles of hippocampal rhythmic slow activity ('theta rhythm') depend on behaviour. Electroencephalography and Clinical Neurophysiology 61:77–88.

Cabral-Calderin Y, Henry MJ (2022) Reliability of neural entrainment in the human auditory system. J Neurosci 42:894–908.

Cabral-Calderin Y, Van Hinsberg D, Thielscher A, Henry MJ (2024) Behavioral entrainment to rhythmic auditory stimulation can be modulated by tACS depending on the electrical stimulation field properties. eLife 12:RP87820.

Carrara-Augustenborg C, Schultz BG (2019) The implicit learning of metrical and non-metrical rhythms in blind and sighted adults. Psychological Research 83:907–923.

Cason N, Astésano C, Schön D (2015) Bridging music and speech rhythm: Rhythmic priming and audio–motor training affect speech perception. Acta Psychologica 155:43–50.

Catani M, Jones DK, Ffytche DH (2005) Perisylvian language networks of the human brain. Annals of Neurology 57:8–16.

Catani M, Mesulam M (2008) The arcuate fasciculus and the disconnection theme in language and aphasia: History and current state. Cortex 44:953–961.

Cavanagh JF, Frank MJ (2014) Frontal theta as a mechanism for cognitive control. Trends in Cognitive Sciences 18:414–421.

Cecere R, Rees G, Romei V (2015) Individual differences in alpha frequency drive crossmodal illusory perception. Current Biology 25:231–235.

Chalas N, Daube C, Kluger DS, Abbasi O, Nitsch R, Gross J (2022) Multivariate analysis of speech envelope tracking reveals coupling beyond auditory cortex. NeuroImage 258:119395.

Chalas N, Daube C, Kluger DS, Abbasi O, Nitsch R, Gross J (2023) Speech onsets and sustained speech contribute differentially to delta and theta speech tracking in auditory cortex. Cerebral Cortex:bhac502.

Chander BS, Witkowski M, Braun C, Robinson SE, Born J, Cohen LG, Birbaumer N, Soekadar SR (2016) tACS phase locking of frontal midline theta oscillations disrupts working memory performance. Front Cell Neurosci 10:1:10.

Chang C, Cunningham JP, Glover GH (2009) Influence of heart rate on the BOLD signal: The cardiac response function. NeuroImage 44:857–869.

Chao ZC, Takaura K, Wang L, Fujii N, Dehaene S (2018) Large-scale cortical networks for hierarchical prediction and prediction error in the primate brain. Neuron 100:1252-1266.e3.

Chen JL, Penhune VB, Zatorre RJ (2008a) Listening to musical rhythms recruits motor regions of the brain. Cerebral Cortex 18:2844–2854.

Chen JL, Penhune VB, Zatorre RJ (2008b) Moving on Time: Brain Network for Auditory–Motor Synchronization is Modulated by Rhythm Complexity and Musical Training. Journal of Cognitive Neuroscience 20:226–239.

Chen Y-P, Schmidt F, Keitel A, Rösch S, Hauswald A, Weisz N (2023) Speech intelligibility changes the temporal evolution of neural speech tracking. NeuroImage 268:119894.

Cheung C, Hamilton LS, Johnson K, Chang EF (2016) The auditory representation of speech sounds in human motor cortex. eLife 5:e12577.

Chi T, Gao Y, Guyton MC, Ru P, Shamma S (1999) Spectro-temporal modulation transfer functions and speech intelligibility. The Journal of the Acoustical Society of America 106:2719–2732.

Chien J-T, Ku Y-C (2016) Bayesian Recurrent Neural Network for Language Modeling. IEEE Trans Neural Netw Learning Syst 27:361–374.

Chollet F, others (2015) Keras. Available at: https://keras.io.

Cogan GB, Thesen T, Carlson C, Doyle W, Devinsky O, Pesaran B (2014) Sensory–motor transformations for speech occur bilaterally. Nature 507:94–98.

Cohen MX (2011) Error-related medial frontal theta activity predicts cingulate-related structural connectivity. NeuroImage 55:1373–1383.

Cohen MX (2014) Analyzing neural time series data: theory and practice. Cambridge, Massachusetts: The MIT Press.

Cohen MX, Ridderinkhof KR, Haupt S, Elger CE, Fell J (2008) Medial frontal cortex and response conflict: Evidence from human intracranial EEG and medial frontal cortex lesion. Brain Research 1238:127–142.

Cole SR, Van Der Meij R, Peterson EJ, De Hemptinne C, Starr PA, Voytek B (2017) Nonsinusoidal beta oscillations reflect cortical pathophysiology in Parkinson's disease. J Neurosci 37:4830–4840.

Cole SR, Voytek B (2017) Brain oscillations and the importance of waveform shape. Trends in Cognitive Sciences 21:137–149.

Cope TE, Sohoglu E, Peterson KA, Jones PS, Rua C, Passamonti L, Sedley W, Post B, Coebergh J, Butler CR, Garrard P, Abdel-Aziz K, Husain M, Griffiths TD, Patterson K, Davis MH, Rowe JB (2023) Temporal lobe perceptual predictions for speech are instantiated in motor cortex and reconciled by inferior frontal cortex. Cell Reports 42:112422.

Corriveau KH, Goswami U (2009) Rhythmic motor entrainment in children with speech and language impairments: Tapping to the beat. Cortex 45:119–130.

Coull JT, Cheng R-K, Meck WH (2011) Neuroanatomical and neurochemical substrates of timing. Neuropsychopharmacol 36:3–25.

Crosse MJ, Di Liberto GM, Bednar A, Lalor EC (2016) The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. Front Hum Neurosci 10.

Crosse MJ, Zuk NJ, Di Liberto GM, Nidiffer A, Molholm S, Lalor E (2021) Linear modeling of neurophysiological responses to naturalistic stimuli: Methodological considerations for applied research. PsyArXiv.

Cummins F, Port R (1998) Rhythmic constraints on stress timing in English. Journal of Phonetics 26:145–171.

Curio G, Neuloh G, Numminen J, Jousm□ki V, Hari R (2000) Speaking modifies voice-evoked activity in the human auditory cortex. Hum Brain Mapp 9:183–191.

D'Ausilio A, Pulvermüller F, Salmas P, Bufalari I, Begliomini C, Fadiga L (2009) The motor somatotopy of speech perception. Current Biology 19:381–385.

Davidesco I, Thesen T, Honey CJ, Melloni L, Doyle W, Devinsky O, Ghitza O, Schroeder C, Poeppel D, Hasson U (2018) Electrocorticographic responses to time-compressed speech vary across the cortical auditory hierarchy. Neuroscience.

de Jong NH, Wempe T (2009) Praat script to detect syllable nuclei and measure speech rate automatically. Behavior Research Methods 41:385–390.

de Pasquale F, Della Penna S, Snyder AZ, Lewis C, Mantini D, Marzetti L, Belardinelli P, Ciancetta L, Pizzella V, Romani GL, Corbetta M (2010) Temporal dynamics of spontaneous MEG activity in brain networks. Proceedings of the National Academy of Sciences 107:6040–6045.

Deco G, Jirsa VK, McIntosh AR (2011) Emerging concepts for the dynamical organization of resting-state activity in the brain. Nat Rev Neurosci 12:43–56.

DeLong KA, Urbach TP, Kutas M (2005) Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. Nat Neurosci 8:1117–1121.

Demerens C, Stankoff B, Logak M, Anglade P, Allinquant B, Couraud F, Zalc B, Lubetzki C (1996) Induction of myelination in the central nervous system by electrical activity. Proceedings of the National Academy of Sciences 93:9887–9892.

DeWitt I, Rauschecker JP (2012) Phoneme and word recognition in the auditory ventral stream. Proc Natl Acad Sci USA 109:E505–E514.

Díaz M, Recabarren M (2024) Beyond one-size-fits-all: speech rate personalization as a key to inclusive video lecture design. Univ Access Inf Soc:1–12.

Dick AS, Bernal B, Tremblay P (2014) The Language Connectome: New Pathways, New Concepts. Neuroscientist 20:453–467.

Dietrich S, Hertrich I, Ackermann H (2013) Ultra-fast speech comprehension in blind subjects engages primary visual cortex, fusiform gyrus, and pulvinar – a functional magnetic resonance imaging (fMRI) study. BMC Neuroscience 14:1–15.

Dietz KC, Polanco JJ, Pol SU, Sim FJ (2016) Targeting human oligodendrocyte progenitors for myelin repair. Experimental Neurology 283:489–500.

Ding N, Patel AD, Chen L, Butler H, Luo C, Poeppel D (2017) Temporal modulations in speech and music. Neuroscience & Biobehavioral Reviews 81:181–187.

Ding N, Simon JZ (2014) Cortical entrainment to continuous speech: functional roles and interpretations. Frontiers in Human Neuroscience 8:311.

Doelling KB, Arnal LH, Assaneo MF (2023) Adaptive oscillators support Bayesian prediction in temporal processing Graham LJ, ed. PLoS Comput Biol 19:e1011669.

Doelling KB, Arnal LH, Ghitza O, Poeppel D (2014) Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. NeuroImage 85:761–768.

Doelling KB, Herbst SK, Arnal LH, van Wassenhove V (2022) Psychological and neuroscientific foundations of rhythms and timing. PsyArXiv.

Doelling KB, Poeppel D (2015) Cortical entrainment to music and its modulation by expertise. Proc Natl Acad Sci USA 112:E6233–E6242.

Donhauser PW, Baillet S (2020) Two distinct neural timescales for predictive speech processing. Neuron 105:385-393.e9.

Donoghue T, Haller M, Peterson EJ, Varma P, Sebastian P, Gao R, Noto T, Lara AH, Wallis JD, Knight RT, Shestyuk A, Voytek B (2020) Parameterizing neural power spectra into periodic and aperiodic components. Nat Neurosci 23:1655–1665.

Dragoy O, Zyryanov A, Bronov O, Gordeyeva E, Gronskaya N, Kryuchkova O, Klyuev E, Kopachev D, Medyanik I, Mishnyakova L, Pedyash N, Pronin I, Reutov A, Sitnikov A, Stupina E, Yashin K, Zhirnova V, Zuev A (2020) Functional linguistic specificity of the left frontal aslant tract for spontaneous speech fluency: Evidence from intraoperative language mapping. Brain and Language 208:104836.

Drake C, Botte M-C (1993) Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. Perception & Psychophysics 54:277–286.

Drake C, Jones MR, Baruch C (2000) The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. Cognition 77:251–288.

Drullman R, Festen JM, Plomp R (1994a) Effect of temporal envelope smearing on speech reception. The Journal of the Acoustical Society of America 95:1053–1064.

Drullman R, Festen JM, Plomp R (1994b) Effect of reducing slow temporal modulations on speech reception. The Journal of the Acoustical Society of America 95:2670–2680.

Du Y, Buchsbaum BR, Grady CL, Alain C (2014) Noise differentially impacts phoneme representations in the auditory and speech motor systems. Proceedings of the National Academy of Sciences 111:7126–7131.

Du Y, Zatorre RJ (2017) Musical training sharpens and bonds ears and tongue to hear speech better. Proc Natl Acad Sci USA 114:13579–13584.

Dupoux E, Green K (1997) Perceptual adjustment to highly compressed speech. Journal of Experimental Psychology: Human Perception and Performance 23:914–927.

Edwards E, Chang EF (2013) Syllabic (~2–5 Hz) and fluctuation (~1–10 Hz) ranges in speech and auditory processing. Hearing Research 305:113–134.

Edwards E, Nagarajan SS, Dalal SS, Canolty RT, Kirsch HE, Barbaro NM, Knight RT (2010) Spatiotemporal imaging of cortical activation during verb generation and picture naming. Neuroimage 50:291–301.

Elbert T, Sterr A, Rockstroh B, Pantev C, Müller MM, Taub E (2002) Expansion of the tonotopic area in the auditory cortex of the blind. J Neurosci 22:9941–9944.

Elliott TM, Theunissen FE (2009) The modulation transfer function for speech intelligibility Friston KJ, ed. PLoS Comput Biol 5:e1000302.

Emmorey K, Giezen MR, Petrich JAF, Spurgeon E, O'Grady Farnady L (2017) The relation between working memory and language comprehension in signers and speakers. Acta Psychologica 177:69–77.

Engel AK, Fries P, Singer W (2001) Dynamic predictions: Oscillations and synchrony in top–down processing. Nat Rev Neurosci 2:704–716.

Engel AK, Gerloff C, Hilgetag CC, Nolte G (2013) Intrinsic coupling modes: Multiscale interactions in ongoing brain activity. Neuron 80:867–886.

Engel AK, König P, Kreiter AK, Schillen TB, Singer W (1992) Temporal coding in the visual cortex: new vistas on integration in the nervous system. Trends in Neurosciences 15:218–226.

Eschmann KCJ, Bader R, Mecklinger A (2018) Topographical differences of frontal-midline theta activity reflect functional differences in cognitive control abilities. Brain and Cognition 123:57–64.

Evans S, Davis MH (2015) Hierarchical organization of auditory and motor representations in speech perception: Evidence from searchlight similarity analysis. Cereb Cortex 25:4772–4788.

Fabbro F, Daro V (1995) Delayed auditory feedback in polyglot simultaneous interpreters. Brain and Language 48:309–319.

Federici A, Fantoni M, Pavani F, Handjaras G, Bednaya E, Martinelli A, Berto M, Ricciardi E, Nava E, Orzan E, Bianchi B, Bottari D (2024) Resilience and vulnerability of speech neural tracking to early auditory deprivation. bioRxiv.

Federmeier KD (2007) Thinking ahead: The role and roots of prediction in language comprehension. Psychophysiology 44:491–505.

Federmeier KD, Kutas M (1999) A rose by any other name: Long-term memory structure and sentence processing. Journal of Memory and Language 41:469–495.

Fernandez J, Downey D (2018) Sampling Informative Training Data for RNN Language Models. In: Proceedings of ACL 2018, Student Research Workshop, pp 9–13. Melbourne, Australia: Association for Computational Linguistics.

Ferrarelli F, Sarasso S, Guller Y, Riedner BA, Peterson MJ, Bellesi M, Massimini M, Postle BR, Tononi G (2012) Reduced natural oscillatory frequency of frontal thalamocortical circuits in schizophrenia. Arch Gen Psychiatry 69:766–774.

Fields RD (2004) Volume transmission in activity-dependent regulation of myelinating glia. Neurochemistry International 45:503–509.

Fischler I, Bloom PA (1979) Automatic and attentional processes in the effects of sentence contexts on word recognition. Journal of Verbal Learning and Verbal Behavior 18:1–20.

Flinker A, Chang EF, Kirsch HE, Barbaro NM, Crone NE, Knight RT (2010) Single-trial speech suppression of auditory cortex activity in humans. J Neurosci 30:16643–16650.

Foecker J, Best A, Hoelig C, Roeder B (2012) The superiority in voice processing of the blind arises from neural plasticity at sensory processing stages. Neuropsychologia 50:2056–2067.

Fontolan L, Morillon B, Liegeois-Chauvel C, Giraud A-L (2014) The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. Nat Commun 5:4694.

Forster KI (1981) Priming and the Effects of Sentence and Lexical Contexts on Naming Time: Evidence for Autonomous Lexical Processing. The Quarterly Journal of Experimental Psychology Section A 33:465–495.

Fox J, Weisberg S (2019) An R companion to applied regression, 3rd ed. Thousand Oaks CA: Sage Publications.

Fox MD, Raichle ME (2007) Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat Rev Neurosci 8:700–711.

Fox MD, Snyder AZ, Zacks JM, Raichle ME (2006) Coherent spontaneous activity accounts for trial-to-trial variability in human evoked brain responses. Nat Neurosci 9:23–25.

Frank SL, Otten LJ, Galli G, Vigliocco G (2015) The ERP response to the amount of information conveyed by words in sentences. Brain and Language 140:1–11.

Friederici AD (2015) White-matter pathways for speech and language processing. In: Handbook of Clinical Neurology, pp 177–186. Elsevier.

Friederici AD, Gierhan SM (2013) The language network. Current Opinion in Neurobiology 23:250–254.

Friederici AD, Pfeifer E, Hahne A (1993) Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. Cognitive Brain Research 1:183–192.

Fries P (2005) A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends in Cognitive Sciences 9:474–480.

Friston K (2005) A theory of cortical responses. Phil Trans R Soc B 360:815–836.

Fröhlich F (2015) Experiments and models of cortical oscillations as a target for noninvasive brain stimulation. In: Progress in Brain Research, pp 41–73. Elsevier.

Fujioka T, Trainor LJ, Large EW, Ross B (2012) Internalized Timing of Isochronous Sounds Is Represented in Neuromagnetic Beta Oscillations. Journal of Neuroscience 32:1791–1802.

Galambos R, Makeig S, Talmachoff PJ (1981) A 40-Hz auditory potential recorded from the human scalp. Proc Natl Acad Sci USA 78:2643–2647.

Garvey WD (1953) The intelligibility of speeded speech. Journal of Experimental Psychology 45:102–108.

Gasser T, Bächer P, Steinberg H (1985) Test-retest reliability of spectral parameters of the EEG. Electroencephalography and Clinical Neurophysiology 60:312–319.

Gautier HOB, Evans KA, Volbracht K, James R, Sitnikov S, Lundgaard I, James F, Lao-Peregrin C, Reynolds R, Franklin RJM, Káradóttir RT (2015) Neuronal activity regulates remyelination via glutamate signalling to oligodendrocyte progenitors. Nat Commun 6:8518.

Ghitza O (2011) Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. Front Psychology 2:130.

Ghitza O (2012) On the role of theta-driven syllabic parsing in decoding speech: Intelligibility of speech with a manipulated modulation spectrum. Front Psychology 3:238.

Ghitza O (2013) The theta-syllable: a unit of speech information defined by cortical function. Front Psychol 4:138.

Ghitza O (2014) Behavioral evidence for the role of cortical theta oscillations in determining auditory channel capacity for speech. Front Psychol 5:652.

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.

Gillis M, Decruy L, Vanthornhout J, Francart T (2022) Hearing loss is associated with delayed neural responses to continuous speech. Eur J of Neuroscience 55:1671–1690.

Giraud A-L, Kleinschmidt A, Poeppel D, Lund TE, Frackowiak RSJ, Laufs H (2007) Endogenous cortical rhythms determine cerebral specialization for speech perception and production. Neuron 56:1127–1134.

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.

Giraud A-L, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci 15:511–517.

Giroud J, Lerousseau JP, Pellegrino F, Morillon B (2023) The channel capacity of multilevel linguistic features constrains speech comprehension. Cognition 232:105345.

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 Griffiths TD, ed. PLoS Biol 18:e3000207.

Goldman-Eisler F (1961) The significance of changes in the rate of articulation. Language and Speech 4:171–174.

Goswami U, Leong V (2013) Speech rhythm and temporal structure: Converging perspectives? Laboratory Phonology 4:67–92.

Gougoux F, Belin P, Voss P, Lepore F, Lassonde M, Zatorre RJ (2009) Voice perception in blind persons: A functional magnetic resonance imaging study. Neuropsychologia 47:2967–2974.

Gougoux F, Lepore F, Lassonde M, Voss P, Zatorre RJ, Belin P (2004) Pitch discrimination in the early blind. Nature 430:309–309.

Gougoux F, Zatorre RJ, Lassonde M, Voss P, Lepore F (2005) A functional neuroimaging study of sound localization: visual cortex activity predicts performance in early-blind individuals Raichle M, ed. PLoS Biol 3:e27.

Grahn JA, Brett M (2007) Rhythm and beat perception in motor areas of the brain. Journal of Cognitive Neuroscience 19:893–906.

Grahn JA, Rowe JB (2009) Feeling the beat: Premotor and striatal interactions in musicians and nonmusicians during beat perception. Journal of Neuroscience 29:7540–7548.

Grant KW, Seitz PF (2000) The recognition of isolated words and words in sentences: Individual variability in the use of sentence context. The Journal of the Acoustical Society of America 107:1000–1011.

Greenberg S (1999) Speaking in shorthand – A syllable-centric perspective for understanding pronunciation variation. Speech Communication 29:159–176.

Greenberg S, Arai T (2001) What are the essential cues for understanding spoken language? The Journal of the Acoustical Society of America 109:2382–2382.

Greenberg S, Arai T, Silipo R (1998) Speech intelligibility derived from exceedingly sparse spectral information. In: 5th International Conference on Spoken Language Processing (ICSLP 1998), pp paper 0074-0. ISCA.

Greenberg S, Carvey H, Hitchcock L, Chang S (2003) Temporal properties of spontaneous speech—a syllable-centric perspective. Journal of Phonetics 31:465–485.

Greicius MD, Krasnow B, Reiss AL, Menon V (2003) Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. Proc Natl Acad Sci USA 100:253–258.

Groppe DM, Bickel S, Keller CJ, Jain SK, Hwang ST, Harden C, Mehta AD (2013) Dominant frequencies of resting human brain activity as measured by the electrocorticogram. NeuroImage 79:223–233.

Grosjean F (1980) Spoken word recognition processes and the gating paradigm. Perception & Psychophysics 28:267–283.

Gross J (2014) Analytical methods and experimental approaches for electrophysiological studies of brain oscillations. Journal of Neuroscience Methods 228:57–66.

Gross J (2019) Magnetoencephalography in cognitive neuroscience: A primer. Neuron 104:189–204.

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 Biol 11:e1001752.

Gudi-Mindermann H, Rimmele JM, Nolte G, Bruns P, Engel AK, Roeder B (2018) Working memory training in congenitally blind individuals results in an integration of occipital cortex in functional networks. Behavioural Brain Research 348:31–41.

Guiraud H, Bedoin N, Krifi-Papoz S, Herbillon V, Caillot-Bascoul A, Gonzalez-Monge S, Boulenger V (2018) Don't speak too fast! Processing of fast rate speech in children with specific language impairment Kotz S, ed. PLoS ONE 13:e0191808.

Gunasekaran H, Azizi L, Van Wassenhove V, Herbst SK (2023) Characterizing endogenous delta oscillations in human MEG. Sci Rep 13:11031.

Guo PJ, Kim J, Rubin R (2014) How video production affects student engagement: an empirical study of MOOC videos. In: Proceedings of the first ACM conference on Learning @ scale conference, pp 41–50. Atlanta Georgia USA: ACM.

Gwilliams L (2020) How the brain composes morphemes into meaning. Phil Trans R Soc B 375:20190311.

Gwilliams L, Davis MH (2022) Extracting language content from speech sounds: The information theoretic approach. In: Speech Perception (Holt LL, Peelle JE, Coffin AB, Popper AN, Fay RR, eds), pp 113–139 Springer Handbook of Auditory Research. Cham: Springer International Publishing.

Gwilliams L, King J-R, Marantz A, Poeppel D (2022) Neural dynamics of phoneme sequences reveal position-invariant code for content and order. Nat Commun 13:6606.

Gwilliams L, Marantz A, Poeppel D, King J-R (2023) Top-down information shapes lexical processing when listening to continuous speech. Language, Cognition and Neuroscience:1–14.

Haak KV, Beckmann CF (2019) Plasticity versus stability across the human cortical visual connectome. Nat Commun 10:3174.

Haegens S, Nacher V, Luna R, Romo R, Jensen O (2011) Alpha-oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. Proceedings of the National Academy of Sciences 108:19377–19382.

Haegens S, Zion Golumbic E (2018) Rhythmic facilitation of sensory processing: A critical review. Neuroscience & Biobehavioral Reviews 86:150–165.

Hagoort P (2014) Nodes and networks in the neural architecture for language: Broca's region and beyond. Current Opinion in Neurobiology 28:136–141.

Hagoort P, Hald L, Bastiaansen M, Petersson KM (2004) Integration of word meaning and world knowledge in language comprehension. Science 304:438–441.

Halgren M, Fabó D, Ulbert I, Madsen JR, Erőss L, Doyle WK, Devinsky O, Schomer D, Cash SS, Halgren E (2018) Superficial slow rhythms integrate cortical processing in humans. Sci Rep 8:2055.

Hämäläinen M, Hari R, Ilmoniemi RJ, Knuutila J, Lounasmaa OV (1993) Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev Mod Phys 65:413–497.

Hamilton LS, Huth AG (2020) The revolution will not be controlled: natural stimuli in speech neuroscience. Language, Cognition and Neuroscience 35:573–582.

Hamilton LS, Oganian Y, Hall J, Chang EF (2021) Parallel and distributed encoding of speech across human auditory cortex. Cell 184:4626-4639.e13.

Hari R, Puce A (2023) MEG - EEG Primer, 2nd ed. Oxford University Press.

Harms MP, Melcher JR (2002) Sound repetition rate in the human auditory pathway: Representations in the waveshape and amplitude of fMRI activation. Journal of Neurophysiology 88:1433–1450.

Hauswald A, Keitel A, Chen Y, Rösch S, Weisz N (2022) Degradation levels of continuous speech affect neural speech tracking and alpha power differently. Eur J of Neuroscience 55:3288–3302.

Hawellek DJ, Schepers IM, Roeder B, Engel AK, Siegel M, Hipp JF (2013) Altered intrinsic neuronal interactions in the visual cortex of the blind. The Journal of Neuroscience 33:17072–17080.

He BJ, Zempel JM, Snyder AZ, Raichle ME (2010) The Temporal Structures and Functional Significance of Scale-free Brain Activity. Neuron 66:353–369.

He D, Buder EH, Bidelman GM (2023) Effects of syllable rate on neuro-behavioral synchronization across modalities: Brain oscillations and speech productions. Neurobiology of Language 4:344–360.

Heilbron M, Armeni K, Schoffelen J-M, Hagoort P, De Lange FP (2022) A hierarchy of linguistic predictions during natural language comprehension. Proc Natl Acad Sci USA 119:e2201968119.

Henry MJ, Herrmann B, Obleser J (2014) Entrained neural oscillations in multiple frequency bands comodulate behavior. Proceedings of the National Academy of Sciences 111:14935–14940.

Henry MJ, Herrmann B, Obleser J (2016) Neural Microstates Govern Perception of Auditory Input without Rhythmic Structure. J Neurosci 36:860–871.

Henry MJ, Obleser J (2012) Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. Proceedings of the National Academy of Sciences 109:20095–20100.

Herbst SK, Stefanics G, Obleser J (2022) Endogenous modulation of delta phase by expectation–A replication of Stefanics et al., 2010. Cortex 149:226–245.

Herholz SC, Zatorre RJ (2012) Musical Training as a Framework for Brain Plasticity: Behavior, Function, and Structure. Neuron 76:486–502.

Herrmann CS (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.

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.

Hertrich I, Dietrich S, Ackermann H (2016) The role of the supplementary motor area for speech and language processing. Neuroscience & Biobehavioral Reviews 68:602–610.

Hertrich I, Dietrich S, Moos A, Trouvain J, Ackermann H (2009) Enhanced speech perception capabilities in a blind listener are associated with activation of fusiform gyrus and primary visual cortex. Neurocase 15:163–170.

Hervais-Adelman A, Moser-Mercer B, Michel CM, Golestani N (2015) fMRI of Simultaneous Interpretation Reveals the Neural Basis of Extreme Language Control. Cereb Cortex 25:4727–4739.

Heusser AC, Poeppel D, Ezzyat Y, Davachi L (2016) Episodic sequence memory is supported by a theta–gamma phase code. Nat Neurosci 19:1374–1380.

Hickok G, Houde J, Rong F (2011) Sensorimotor integration in speech processing: Computational basis and neural organization. Neuron 69:407–422.

Hickok G, Poeppel D (2004) Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 92:67–99.

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402.

Hidalgo C, Zécri A, Pesnot-Lerousseau J, Truy E, Roman S, Falk S, Dalla Bella S, Schön D (2021) Rhythmic abilities of children with hearing loss. Ear & Hearing 42:364–372.

Hill RA, Patel KD, Goncalves CM, Grutzendler J, Nishiyama A (2014) Modulation of oligodendrocyte generation during a critical temporal window after NG2 cell division. Nat Neurosci 17:1518–1527.

Hillebrand A, Barnes GR, Bosboom JL, Berendse HW, Stam CJ (2012) Frequency-dependent functional connectivity within resting-state networks: An atlas-based MEG beamformer solution. NeuroImage 59:3909–3921.

Hincapié Casas AS, 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.

Hipp JF, Engel AK, Siegel M (2011) Oscillatory synchronization in large-scale cortical networks predicts perception. Neuron 69:387–396.

Hipp JF, Hawellek DJ, Corbetta M, Siegel M, Engel AK (2012) Large-scale cortical correlation structure of spontaneous oscillatory activity. Nat Neurosci 15:884–890.

Ho HT, Leung J, Burr DC, Alais D, Morrone MC (2017) Auditory sensitivity and decision criteria oscillate at different frequencies separately for the two ears. Current Biology 27:3643-3649.e3.

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Computation 9:1735–1780.

Hoetting K, Roesler F, Roeder B (2004) Altered auditory-tactile interactions in congenitally blind humans: an event-related potential study. Experimental Brain Research 159:370–381.

Hofstetter S, Sabbah N, Mohand-Saïd S, Sahel J-A, Habas C, Safran AB, Amedi A (2019) The development of white matter structural changes during the process of deterioration of the visual field. Sci Rep 9:2085.

Holcombe AO (2009) Seeing slow and seeing fast: two limits on perception. Trends in Cognitive Sciences 13:216–221.

Houde JF, Jordan MI (1998) Sensorimotor adaptation in speech production. Science 279:1213–1216.

Houde JF, Nagarajan SS, Sekihara K, Merzenich MM (2002) Modulation of the auditory cortex during speech: An MEG study. Journal of Cognitive Neuroscience 14:1125–1138.

Houtgast T, Steeneken HJM (1973) The modulation transfer function in room acoustics as a predictor of speech intelligibility. The Journal of the Acoustical Society of America 54:557–557.

Houtgast T, Steeneken HJM, Plomp R (1980) Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function. I. General Room Acoustics. 46.

Howard MF, 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.

Hsieh L-T, Ranganath C (2014) Frontal midline theta oscillations during working memory maintenance and episodic encoding and retrieval. NeuroImage 85:721–729.

Hullett PW, Hamilton LS, Mesgarani N, Schreiner CE, Chang EF (2016) Human superior temporal gyrus organization of spectrotemporal modulation tuning derived from speech stimuli. J Neurosci 36:2014–2026.

Hutcheon B, Yarom Y (2000) Resonance, oscillation and the intrinsic frequency preferences of neurons. Trends in Neurosciences 23:216–222.

Ince RAA, Giordano BL, Kayser C, Rousselet GA, Gross J, Schyns PG (2017) A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula: Gaussian Copula Mutual Information. Hum Brain Mapp 38:1541–1573.

Ishibashi T, Dakin KA, Stevens B, Lee PR, Kozlov SV, Stewart CL, Fields RD (2006) Astrocytes promote myelination in response to electrical impulses. Neuron 49:823–832.

Itthipuripat S, Wessel JR, Aron AR (2013) Frontal theta is a signature of successful working memory manipulation. Exp Brain Res 224:255–262.

Iversen JR, Repp BH, Patel AD (2009) Top-down control of rhythm perception modulates early auditory responses. Annals of the New York Academy of Sciences 1169:58–73.

Izhikevich EM, Desai NS, Walcott EC, Hoppensteadt FC (2003) Bursts as a unit of neural information: selective communication via resonance. Trends in Neurosciences 26:161–167.

Janse E (2004) Word perception in fast speech: artificially time-compressed vs. naturally produced fast speech. Speech Communication 42:155–173.

Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM (2012) FSL. NeuroImage 62:782–790.

Jensen O, Bonnefond M, VanRullen R (2012) An oscillatory mechanism for prioritizing salient unattended stimuli. Trends in Cognitive Sciences 16:200–206.

Jensen O, Mazaheri A (2010) Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. Front Hum Neurosci 4.

Jiang J, Zhu W, Shi F, Liu Y, Li J, Qin W, Li K, Yu C, Jiang T (2009) Thick visual cortex in the early blind. Journal of Neuroscience 29:2205–2211.

Jones JA, Munhall KG (2000) Perceptual calibration of *F* 0 production: Evidence from feedback perturbation. The Journal of the Acoustical Society of America 108:1246–1251.

Jones MR (1976) Time, our lost dimension: Toward a new theory of perception, attention, and memory. Psychological Review 83:323–355.

Jones MR (2019) Time will tell: a theory of dynamic attending. New York, NY: Oxford University Press.

Jones MR, Boltz M, Kidd G (1982) Controlled attending as a function of melodic and temporal context. Perception & Psychophysics 32:211–218.

Jones MR, Kidd G, Wetzel R (1981) Evidence for rhythmic attention. Journal of Experimental Psychology: Human Perception and Performance 7:1059–1073.

Jones MR, Moynihan H, MacKenzie N, Puente J (2002) Temporal aspects of stimulus-driven attending in dynamic arrays. Psychological Science 13:7.

Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. Physiological Reviews 84:541–577.

Jurafsky D, Martin JH (2009) Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition, 2. ed. [Nachdr.]. Upper Saddle River, NJ: Prentice Hall.

Kaya E, Henry MJ (2022) Reliable estimation of internal oscillator properties from a novel, fast-paced tapping paradigm. Sci Rep 12:20466.

Kaya E, Kotz SA, Henry MJ (2023) Individual differences in internal oscillator properties that impact perception and production of rhythms. elife.

Kayser C (2019) Evidence for the rhythmic perceptual sampling of auditory scenes. Neuroscience.

Kebritchi M, Lipschuetz A, Santiague L (2017) Issues and Challenges for Teaching Successful Online Courses in Higher Education: A Literature Review. Journal of Educational Technology Systems 46:4–29.

Keitel A, Gross J (2016) Individual human brain areas can be identified from their characteristic spectral activation fingerprints Engel AK, ed. PLOS Biology 14:e1002498.

Keitel A, Gross J, Kayser C (2018) Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features Bizley J, ed. PLOS Biology 16:e2004473.

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

Kern P, Assaneo MF, Endres D, Poeppel D, Rimmele JM (2021) Preferred auditory temporal processing regimes and auditory-motor synchronization. Psychon Bull Rev 28:1860–1873.

Kidd G, Boltz M, Jones MR (1984) Some effects of rhythmic context on melody recognition. The American Journal of Psychology 97:153–173.

Kim H, Lee KM, Lee YS (2023) Sensorimotor and working memory systems jointly support development of perceptual rhythm processing. Developmental Science 26:e13261.

Kim JA, Davis KD (2021) Magnetoencephalography: physics, techniques, and applications in the basic and clinical neurosciences. Journal of Neurophysiology 125:938–956.

Klawiter EC, Schmidt RE, Trinkaus K, Liang H-F, Budde MD, Naismith RT, Song S-K, Cross AH, Benzinger TL (2011) Radial diffusivity predicts demyelination in ex vivo multiple sclerosis spinal cords. NeuroImage 55:1454–1460.

Klimesch W (1999) EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Research Reviews 29:169–195.

Klimesch W, Sauseng P, Hanslmayr S (2007) EEG alpha oscillations: The inhibition–timing hypothesis. Brain Research Reviews 53:63–88.

Klinge C, Eippert F, Roder B, Buchel C (2010) Corticocortical Connections Mediate Primary Visual Cortex Responses to Auditory Stimulation in the Blind. Journal of Neuroscience 30:12798–12805.

Komorowski MK, Aghabeig M, Nikadon J, Piotrowski T, Dreszer J, Bałaj B, Lewandowska M, Wojciechowski J, Pawlaczyk N, Szmytke M, Cichocki A, Duch W (2018) Multi-level Explanations in Neuroscience II: EEG Spectral Fingerprints and Tensor Decompositions for Understanding Brain Activity --- Initial Results. Acta Phys Pol B 49:2011.

Kondacs A, Szabó M (1999) Long-term intra-individual variability of the background EEG in normals. Clinical Neurophysiology 110:1708–1716.

Korte M, Rauschecker JP (1993) Auditory spatial tuning of cortical neurons is sharpened in cats with early blindness. Journal of Neurophysiology 70:1717–1721.

Kösem A, Bosker HR, Takashima A, Meyer A, Jensen O, Hagoort P (2018) Neural entrainment determines the words we hear. Current Biology 28:2867-2875.e3.

Kösem A, Dai B, McQueen JM, Hagoort P (2023) Neural tracking of speech envelope does not unequivocally reflect intelligibility. NeuroImage 272:120040.

Kösem A, Gramfort A, van Wassenhove V (2014) Encoding of event timing in the phase of neural oscillations. NeuroImage 92:274–284.

Kotz SA, Ravignani A, Fitch WT (2018) The evolution of rhythm processing. Trends in Cognitive Sciences 22:896–910.

Kotz SA, Schwartze M (2010) Cortical speech processing unplugged: a timely subcortico-cortical framework. Trends in Cognitive Sciences 14:392–399.

Kotz SA, Schwartze M, Schmidt-Kassow M (2009) Non-motor basal ganglia functions: A review and proposal for a model of sensory predictability in auditory language perception. Cortex 45:982–990.

Kraus N, Strait DL, Parbery-Clark A (2012) Cognitive factors shape brain networks for auditory skills: spotlight on auditory working memory. Annals of the New York Academy of Sciences 1252:100–107.

Kriegseis A, Hennighausen E, Roesler F, Roeder B (2006) Reduced EEG alpha activity over parieto-occipital brain areas in congenitally blind adults. Clinical Neurophysiology 117:1560–1573.

Kubota Y, Sato W, Toichi M, Murai T, Okada T, Hayashi A, Sengoku A (2001) Frontal midline theta rhythm is correlated with cardiac autonomic activities during the performance of an attention demanding meditation procedure. Cognitive Brain Research 11:281–287.

Kulashekhar S, Pekkola J, Palva JM, Palva S (2016) The role of cortical beta oscillations in time estimation: Oscillations in Time Estimation. Hum Brain Mapp 37:3262–3281.

Kuperberg GR, Jaeger TF (2016) What do we mean by prediction in language comprehension? Language, Cognition and Neuroscience 31:32–59.

Kutas M, Federmeier KD (2011) Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). Annu Rev Psychol 62:621–647.

Kutas M, Hillyard SA (1980) Reading senseless sentences: Brain potentials reflect semantic incongruity. Science 207:203–205.

Kutas M, Hillyard SA (1984) Brain potentials during reading reflect word expectancy and semantic association. Nature 307:161–163.

Lachaux J-P, Rodriguez E, Martinerie J, Varela FJ (1999) Measuring phase synchrony in brain signals. Human brain mapping 8:194–208.

Lakatos P, Barczak A, Neymotin SA, McGinnis T, Ross D, Javitt DC, O'Connell MN (2016) Global dynamics of selective attention and its lapses in primary auditory cortex. Nature Neuroscience 19:1707–1717.

Lakatos P, Gross J, Thut G (2019) A New Unifying Account of the Roles of Neuronal Entrainment. Current Biology 29:R890–R905.

Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. Science 320:110–113.

Lakatos P, Musacchia G, O'Connel MN, Falchier AY, Javitt DC, Schroeder CE (2013) The spectrotemporal flter mechanism of auditory selective attention. Neuron 77:750–761.

Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, Schroeder CE (2005) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. Journal of Neurophysiology 94:1904–1911.

Lane C, Kanjlia S, Richardson H, Fulton A, Omaki A, Bedny M (2017) Reduced left lateralization of language in congenitally blind individuals. Journal of Cognitive Neuroscience 29:65–78.

Lee BS (1950) Effects of delayed speech feedback. The Journal of the Acoustical Society of America 22:824–826.

Lee C, Federmeier KD (2009) Wave-ering: An ERP study of syntactic and semantic context effects on ambiguity resolution for noun/verb homographs. Journal of Memory and Language 61:538–555.

Lerens E, Araneda R, Renier L, De Volder AG (2014) Improved beat asynchrony detection in early blind individuals. Perception 43:1083–1096.

Lessard N, Paré M, Lepore F, Lassonde M (1998) Early-blind human subjects localize sound sources better than sighted subjects. Nature 395:278–280.

Lewis AG, Bastiaansen M (2015) A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. Cortex 68:155–168.

Lewis PA, Miall RC (2003a) Distinct systems for automatic and cognitively controlled time measurement: evidence from neuroimaging. Current Opinion in Neurobiology 13:250–255.

Lewis PA, Miall RC (2003b) Brain activation patterns during measurement of sub- and supra-second intervals.

Li X, Zatorre RJ, Du Y (2021) The microstructural plasticity of the arcuate fasciculus undergirds improved speech in noise perception in musicians. Cerebral Cortex 31:3975–3985.

Liegeois-Chauvel C, Lorenzi C, Trebuchon A, Regis J, Chauvel P (2004) Temporal envelope processing in the human left and right auditory cortices. Cerebral Cortex 14:731–740.

Liu Y, Yu C, Liang M, Li J, Tian L, Zhou Y, Qin W, Li K, Jiang T (2007) Whole brain functional connectivity in the early blind. Brain 130:2085–2096.

Lizcano-Cortés F, Gómez-Varela I, Mares C, Wallisch P, Orpella J, Poeppel D, Ripollés P, Assaneo MF (2022) Speech-to-Speech Synchronization protocol to classify human participants as high or low auditory-motor synchronizers. STAR Protocols 3:101248.

Loehr JD, Palmer C (2011) Temporal coordination between performing musicians. Quarterly Journal of Experimental Psychology 64:2153–2167.

Lopes da Silva F (1991) Neural mechanisms underlying brain waves: from neural membranes to networks. Electroencephalography and Clinical Neurophysiology 79:81–93.

Lopes da Silva F (2013) EEG and MEG: Relevance to neuroscience. Neuron 80:1112–1128.

Lőrincz ML, Kékesi KA, Juhász G, Crunelli V, Hughes SW (2009) Temporal framing of thalamic relay-mode firing by phasic inhibition during the alpha rhythm. Neuron 63:683–696.

Lubinus C, Keitel A, Obleser J, Poeppel D, Rimmele JM (2023) Explaining flexible continuous speech comprehension from individual motor rhythms. Proc R Soc B 290:20222410.

Lubinus C, Orpella J, Keitel A, Gudi-Mindermann H, Engel AK, Roeder B, Rimmele JM (2021) Data-driven classification of spectral profiles reveals brain region-specific plasticity in blindness. Cerebral Cortex 31:2505–2522.

Luo C, Guo Z, Lai Y, Liao W, Liu Q, Kendrick KM, Yao D, Li H (2012) Musical training induces functional plasticity in perceptual and motor networks: Insights from resting-state fMRI He Y, ed. PLoS ONE 7:e36568.

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

Lykken DT, Tellegen A, Thorkelson K (1974) Genetic determination of EEG frequency spectra. Biological Psychology 1:245–259.

Macar F, Lejeune H, Bonnet M, Ferrara A, Pouthas V, Vidal F, Maquet P (2002) Activation of the supplementary motor area and of attentional networks during temporal processing. Experimental Brain Research 142:475–485.

MacQueen J (1967) Some methods for classification and analysis of multivariate observations. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability 1:281–297.

Mahjoory K, Schoffelen J-M, Keitel A, Gross J (2020) The frequency gradient of human resting-state brain oscillations follows cortical hierarchies. eLife 9:e53715.

Manning JR, Jacobs J, Fried I, Kahana MJ (2009) Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. J Neurosci 29:13613–13620.

Mantini D, Perrucci MG, Del Gratta C, Romani GL, Corbetta M (2007) Electrophysiological signatures of resting state networks in the human brain. Proceedings of the National Academy of Sciences 104:13170–13175.

Marshall TR, den Boer S, Cools R, Jensen O, Fallon SJ, Zumer JM (2018) Occipital alpha and gamma oscillations support complementary mechanisms for processing stimulus value associations. Journal of Cognitive Neuroscience 30:119–129.

Marslen-Wilson W, Brown CM, Tyler LK (1988) Lexical representations in spoken language comprehension. Language and Cognitive Processes 3:1–16.

Mayhew JEW, Askew S, Zheng Y, Porrill J, Westby GWM, Redgrave P, Rector DM, Harper RM (1996) Cerebral vasomotion: A 0.1-Hz oscillation in reflected light imaging of neural activity. NeuroImage 4:183–193.

McAuley JD, Jones MR, Holub S, Johnston HM, Miller NS (2006) The time of our lives: Life span development of timing and event tracking. Journal of Experimental Psychology: General 135:348–367.

McPherson T, Berger D, Alagapan S, Fröhlich F (2018) Intrinsic Rhythmicity Predicts Synchronization-Continuation Entrainment Performance. Scientific Reports 8.

Mehler J, Sebastian N, Altmann G, Dupoux E, Christophe A, Pallier C (1993) Understanding compressed sentences: The role of rhythm and meaning. Annals of the New York Academy of Sciences 682:272–282.

Merchant H, Yarrow K (2016) How the motor system both encodes and influences our sense of time. Current Opinion in Behavioral Sciences 8:22–27.

Merity S, Keskar NS, Socher R (2017) Regularizing and Optimizing LSTM Language Models.

Meyer L (2018) The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms. Eur J Neurosci 48:2609–2621.

Meyer L, Sun Y, Martin AE (2020) Synchronous, but not entrained: exogenous and endogenous cortical rhythms of speech and language processing. Language, Cognition and Neuroscience 35:1089–1099.

Michaelis K, Wiener M, Thompson JC (2014) Passive listening to preferred motor tempo modulates corticospinal excitability. Front Hum Neurosci 8.

Michalareas G, Vezoli J, van Pelt S, Schoffelen J-M, Kennedy H, Fries P (2016) Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. Neuron 89:384–397.

Mikolov T (2010) Recurrent neural network based language model. Eleventh annual conference of the international speech communication association:24.

Miller GA, Isard S (1963) Some perceptual consequences of linguistic rules. Journal of Verbal Learning and Verbal Behavior 2:217–228.

Miller JL, Grosjean F, Lomanto C (1984) Articulation rates and its variability in spontaneous speech: A reanalysis and some implications. Phonetica 41:215–3225.

Mitsuya T, MacDonald EN, Munhall KG (2014) Temporal control and compensation for perturbed voicing feedback. The Journal of the Acoustical Society of America 135:2986–2994.

Molina JL, Voytek B, Thomas ML, Joshi YB, Bhakta SG, Talledo JA, Swerdlow NR, Light GA (2020) Memantine effects on electroencephalographic measures of putative excitatory/inhibitory balance in schizophrenia. Biol Psychiatry Cogn Neurosci Neuroimaging 5:562–568.

Molinaro N, Monsalve IF, Lizarazu M (2016) Is there a common oscillatory brain mechanism for producing and predicting language? Language, Cognition and Neuroscience 31:145–158.

Moos A, Trouvain J (2007) Comprehension of ultra-fast speech - blind vs. “normally-hearing” persons. Proceedings of the 16th International Congress of Phonetic Sciences 1:677–680.

Morillon B, Arnal LH, Schroeder CE, Keitel A (2019) Prominence of delta oscillatory rhythms in the motor cortex and their relevance for auditory and speech perception. Neuroscience & Biobehavioral Reviews 107:136–142.

Morillon B, Baillet S (2017) Motor origin of temporal predictions in auditory attention. Proceedings of the National Academy of Sciences 114:E8913–E8921.

Morillon B, Hackett TA, Kajikawa Y, Schroeder CE (2015) Predictive motor control of sensory dynamics in auditory active sensing. Current Opinion in Neurobiology 31:230–238.

Morillon B, Liégeois-Chauvel C, Arnal LH, Bénar C-G, Giraud A-L (2012) Asymmetric function of theta and gamma activity in syllable processing: An intra-cortical study. Front Psychology 3.

Morillon B, Schroeder CE, Wyart V (2014) Motor contributions to the temporal precision of auditory attention. Nature Communications 5:5255.

Morillon B, Schroeder CE, Wyart V, Arnal LH (2016) Temporal prediction in lieu of periodic stimulation. The Journal of Neuroscience 36:2342–2347.

Moulines E, Charpentier F (1990) Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. Speech Communication 9:453–467.

Murakami S, Okada Y (2006) Contributions of principal neocortical neurons to magnetoencephalography and electroencephalography signals. The Journal of Physiology 575:925–936.

Narici L, Romani GL (1989) Neuromagnetic investigation of synchronized spontaneous activity. Brain Topogr 2:19–30.

Nastase SA, Goldstein A, Hasson U (2020) Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. NeuroImage 222:117254.

Ng BSW, Schroeder T, Kayser C (2012) A precluding but not ensuring role of entrained low-frequency oscillations for auditory perception. Journal of Neuroscience 32:12268–12276.

Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: A primer with examples. Hum Brain Mapp 15:1–25.

Nieuwland MS et al. (2020) Dissociable effects of prediction and integration during language comprehension: evidence from a large-scale study using brain potentials. Phil Trans R Soc B 375:20180522.

Noebels JL, Roth WT, Kopell BS (1978) Cortical slow potentials and the occipital EEG in congenital blindness. Journal of the Neurological Sciences 37:51–58.

Nolte G, Ziehe A, Nikulin VV, Schlögl A, Krämer N, Brismar T, Müller K-R (2008) Robustly estimating the flow direction of information in complex physical systems. Phys Rev Lett 100:234101.

Noppeney U (2007) The effects of visual deprivation on functional and structural organization of the human brain. Neuroscience & Biobehavioral Reviews 31:1169–1180.

Noppeney U, Friston KJ, Ashburner J, Frackowiak R, Price CJ (2005) Early visual deprivation induces structural plasticity in gray and white matter. Current Biology 15:R488–R490.

Norcia AM, Appelbaum LG, Ales JM, Cottereau BR, Rossion B (2015) The steady-state visual evoked potential in vision research: A review. Journal of Vision 15:4.

Notbohm A, Kurths J, Herrmann CS (2016) Modification of brain oscillations via rhythmic light stimulation provides evidence for entrainment but not for superposition of event-related responses. Front Hum Neurosci 10:10.

Nourski KV, Reale RA, Oya H, Kawasaki H, Kovach CK, Chen H, Howard MA, Brugge JF (2009) Temporal Envelope of Time-Compressed Speech Represented in the Human Auditory Cortex. Journal of Neuroscience 29:15564–15574.

Obleser J, Henry MJ, Lakatos P (2017) What do we talk about when we talk about rhythm? PLoS Biol 15:e2002794.

Obleser J, Kayser C (2019) Neural entrainment and attentional selection in the listening brain. Trends in Cognitive Sciences 23:913–926.

Obleser J, Kotz SA (2010) Expectancy constraints in degraded speech modulate the language comprehension network. Cerebral Cortex 20:633–640.

Obleser J, Wise RJS, Alex Dresner M, Scott SK (2007) Functional integration across brain regions improves speech perception under adverse listening conditions. Journal of Neuroscience 27:2283–2289.

Oechslin MS (2010) The plasticity of the superior longitudinal fasciculus as a function of musical expertise: a diffusion tensor imaging study. Front Hum Neurosci 3:1076.

Oganian Y, Chang EF (2019) A speech envelope landmark for syllable encoding in human superior temporal gyrus. Sci Adv 5:eaay6279.

Olsthoorn NM, Andringa S, Hulstijn JH (2014) Visual and auditory digit-span performance in native and non-native speakers. International Journal of Bilingualism 18:663–673.

Onton J, Delorme A, Makeig S (2005) Frontal midline EEG dynamics during working memory. NeuroImage 27:341–356.

Oostenveld R, Fries P, Maris E, Schoffelen J-M (2011) FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational Intelligence and Neuroscience 2011:1–9.

Orpella J, Assaneo MF, Ripollés P, Noejovich L, López-Barroso D, Diego-Balaguer RD, Poeppel D (2022) Differential activation of a frontoparietal network explains population-level differences in statistical learning from speech Rushworth MFS, ed. PLoS Biol 20:e3001712.

Ortiz-Terán L, Ortiz T, Perez DL, Aragón JI, Diez I, Pascual-Leone A, Sepulcre J (2016) Brain plasticity in blind subjects centralizes beyond the modal cortices. Front Syst Neurosci 10:61.

Ossandón JP, Stange L, Gudi-Mindermann H, Rimmele JM, Sourav S, Bottari D, Kekunnaya R, Roeder B (2023) The development of oscillatory and aperiodic resting state activity is linked to a sensitive period in humans. NeuroImage 275:120171.

Ottosson J, Grahn P (2005) A comparison of leisure time spent in a garden with leisure time spent indoors: On measures of restoration in residents in geriatric care. Landscape Research 30:23–55.

Ouedraogo DW, Lenck-Santini P-P, Marti G, Robbe D, Crépel V, Epsztein J (2016) Abnormal UP/DOWN membrane potential dynamics coupled with the neocortical slow oscillation in dentate granule cells during the latent phase of temporal lobe epilepsy. eneuro 3:ENEURO.0017-16.2016.

Park H, Ince RAA, Schyns PG, 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.

Park H, Thut G, Gross J (2018) Predictive entrainment of natural speech through two fronto-motor top-down channels. Language, Cognition and Neuroscience 35:739–751.

Park H-J, Friston K (2013) Structural and functional brain networks: From connections to cognition. Science 342:1238411–1238411.

Park H-J, Jeong S-O, Kim EY, Kim D-J, Kim SY, Lee SC, Lee JD (2007) Reorganization of neural circuits in the blind on diffusion direction analysis. Neuroreport 18:1757–1760.

Park H-J, Lee JD, Kim EY, Park B, Oh M-K, Lee S, Kim J-J (2009) Morphological alterations in the congenital blind based on the analysis of cortical thickness and surface area. NeuroImage 47:98–106.

Pascual-Leone A, Amedi A, Fregni F, Merabet LB (2005) The plastic human brain cortex. Annu Rev Neurosci 28:377–401.

Pascual-Leone A, Torres F (1993) Plasticity of the sensorimotor cortex representation of the reading finger in Braille readers. Brain 116:39–52.

Pastor MA, Artieda J, Arbizu J, Marti-Climent JM, Peñuelas I, Masdeu JC (2002) Activation of human cerebral and cerebellar cortex by auditory stimulation at 40 Hz. J Neurosci 22:10501–10506.

Patel AD (2011) Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. Front Psychology 2.

Peelle JE, Davis MH (2012) Neural oscillations carry speech rhythm through to comprehension. Frontiers in Psychology 3:320.

Peelle JE, Gross J, Davis MH (2013) Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cerebral Cortex 23:1378–1387.

Pefkou M, Arnal LH, Fontolan L, Giraud A-L (2017) θ-band and β-band neural activity reflects independent syllable tracking and comprehension of time-compressed speech. The Journal of Neuroscience 37:7930–7938.

Pelland M, Orban P, Dansereau C, Lepore F, Bellec P, Collignon O (2017) State-dependent modulation of functional connectivity in early blind individuals. NeuroImage 147:532–541.

Pellegrino F, Coupé C, Marsico E (2011) Across-language perspective on speech information rate. Language 87:539–558.

Peña M, Melloni L (2012) Brain Oscillations during Spoken Sentence Processing. Journal of Cognitive Neuroscience 24:1149–1164.

Penn LR, Ayasse ND, Wingfield A, 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 144:2088–2094.

Phillips-Silver J, Toiviainen P, Gosselin N, Turgeon C, Lepore F, Peretz I (2015) Cochlear implant users move in time to the beat of drum music. Hearing Research 321:25–34.

Pickering MJ, Gambi C (2018) Predicting while comprehending language: A theory and review. Psychological Bulletin 144:1002–1044.

Picton TW, John MS, Dimitrijevic A, Purcell D (2003) Human auditory steady-state responses: Respuestas auditivas de estado estable en humanos. International Journal of Audiology 42:177–219.

Picton TW, Skinner CR, Champagne SC, Kellett AJC, Maiste AC (1987) Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone. The Journal of the Acoustical Society of America 82:165–178.

Pinneo LR (1966) On noise in the nervous system. Psychological Review 73:242–247.

Pittman-Polletta BR, Wang Y, Stanley DA, Schroeder CE, Whittington MA, Kopell NJ (2021) Differential contributions of synaptic and intrinsic inhibitory currents to speech segmentation via flexible phase-locking in neural oscillators Gutkin BS, ed. PLoS Comput Biol 17:e1008783.

Plöchl M, Fiebelkorn I, Kastner S, Obleser J (2022) Attentional sampling of visual and auditory objects is captured by theta-modulated neural activity. Eur J of Neuroscience 55:3067–3082.

Poeppel D (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time.' Speech Communication 41:245–255.

Poeppel D, Assaneo MF (2020) Speech rhythms and their neural foundations. Nat Rev Neurosci 21:322–334.

Poeppel D, Idsardi WJ, van Wassenhove V (2008) Speech perception at the interface of neurobiology and linguistics. Philosophical Transactions of the Royal Society B: Biological Sciences 363:1071–1086.

Politimou N, Dalla Bella S, Farrugia N, Franco F (2019) Born to speak and sing: Musical predictors of language development in pre-schoolers. Front Psychol 10:948.

Portoles O, Borst JP, van Vugt MK (2018) Characterizing synchrony patterns across cognitive task stages of associative recognition memory. Eur J Neurosci 48:2759–2769.

Provasi J, Anderson DI, Barbu-Roth M (2014) Rhythm perception, production, and synchronization during the perinatal period. Front Psychol 5:105881.

Ptito M, Schneider FCG, Paulson OB, Kupers R (2008) Alterations of the visual pathways in congenital blindness. Exp Brain Res 187:41–49.

Pulvermüller F, Fadiga L (2010) Active perception: sensorimotor circuits as a cortical basis for language. Nat Rev Neurosci 11:351–360.

Pulvermüller F, Huss M, Kherif F, Moscoso Del Prado Martin F, Hauk O, Shtyrov Y (2006) Motor cortex maps articulatory features of speech sounds. Proc Natl Acad Sci USA 103:7865–7870.

Raichle ME (2011) The restless brain. Brain Connectivity 1:3–12.

Raichle ME (2015) The restless brain: how intrinsic activity organizes brain function. Phil Trans R Soc B 370:20140172.

Rao RPN, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87.

Rauschecker JP (2008) Plasticity of cortical maps in visual deprivation. In: Blindness and brain plasticity in navigation and object perception, pp 43–66. New York: Taylor & Francis.

Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12:718–724.

Rauschecker JP, Tian B, Korte M, Egert U (1992) Crossmodal changes in the somatosensory vibrissa/barrel system of visually deprived animals. Proceedings of the National Academy of Sciences 89:5063–5067.

Reislev NL, Dyrby TB, Siebner HR, Kupers R, Ptito M (2016) Simultaneous assessment of white matter changes in microstructure and connectedness in the blind brain. Neural Plasticity 2016:1–12.

Repp BH (2005) Sensorimotor synchronization: A review of the tapping literature. Psychonomic Bulletin & Review 12:969–992.

Repp BH, Su Y-H (2013) Sensorimotor synchronization: A review of recent research (2006–2012). Psychon Bull Rev 20:403–452.

Restani L, Caleo M (2016) Reorganization of visual callosal connections following alterations of retinal input and brain damage. Front Syst Neurosci 10:86.

Reynolds DA, Rose RC (1995) Robust text-independent speaker identification using Gaussian mixture speaker models. IEEE Trans Speech Audio Process 3:72–83.

Richardson JTE (2007) Measures of short-term memory: A historical review. Cortex 43:635–650.

Riecke L, Formisano E, Sorger B, Başkent D, Gaudrain E (2018) Neural entrainment to speech modulates speech intelligibility. Current Biology 28:161-169.e5.

Rimmele J, Keitel A (2023) Region-specific endogenous brain rhythms and their role for speech and language. PsyArXiv.

Rimmele JM, Gudi-Mindermann H, Nolte G, Roeder B, Engel AK (2019) Working memory training integrates visual cortex into beta-band networks in congenitally blind individuals. NeuroImage 194:259–271.

Rimmele JM, Kern P, Lubinus C, Frieler K, Poeppel D, Assaneo MF (2022) Musical sophistication and speech auditory-motor coupling: Easy tests for quick answers. Front Neurosci 15:764342.

Rimmele JM, Morillon B, Poeppel D, Arnal LH (2018) Proactive sensing of periodic and aperiodic auditory patterns. Trends in Cognitive Sciences 22:870–882.

Rimmele JM, Poeppel D, Ghitza O (2021) Acoustically driven cortical δ oscillations underpin prosodic chunking. eNeuro 8:ENEURO.0562-20.2021.

Rimmele JM, 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.

Robertson MM, Furlong S, Voytek B, Donoghue T, Boettiger CA, Sheridan MA (2019) EEG power spectral slope differs by ADHD status and stimulant medication exposure in early childhood. Journal of Neurophysiology 122:2427–2437.

Rodenburg M (1977) Investigation of temporal effects with amplitude modulated signals. In: Psychophysics and physiology of hearing: an international symposium, pp 429–439. London, New York: Academic Press.

Rodenburg M, Verweij C, Van Der Brink G (1972) Analysis of evoked responses in man elicited by sinusoidally modulated noise. Int J Audiol 11:283–293.

Roeder B, Demuth L, Streb J, Roesler F (2003) Semantic and morpho-syntactic priming in auditory word recognition in congenitally blind adults. Language and Cognitive Processes 18:1–20.

Roeder B, Kraemer UM, Lange K (2007) Congenitally blind humans use different stimulus selection strategies in hearing: An ERP study of spatial and temporal attention. Restorative Neurology and Neuroscience 25:311–322.

Roeder B, Neville H (2003) Developmental functional plasticity Grafman J, Robertson I, eds. Handbook of neuropsychology:231–270.

Roeder B, Roesler F (2003) Memory for environmental sounds in sighted, congenitally blind and late blind adults: evidence for cross-modal compensation. International Journal of Psychophysiology 50:27–39.

Roeder B, Roesler F, Hennighausen E, Nicker F (1996) Event-related potentials during auditory and somatosensory discrimination in sighted and blind human subjects. Cognitive Brain Research 4:17.

Roeder B, Roesler F, Neville HJ (1999) Effects of interstimulus interval on auditory event-related potentials in congenitally blind and normally sighted humans. Neuroscience Letters 264:53–56.

Roeder B, Roesler F, Neville HJ (2000) Event-related potentials during auditory language processing in congenitally blind and sighted people. Neuropsychologia 38:1482–1502.

Roeder B, Roesler F, Neville HJ (2001) Auditory memory in congenitally blind adults: A behavioral-electrophysiological investigation. Cognitive Brain Research 11:289–303.

Roeder B, Roesler F, Spence C (2004) Early vision impairs tactile perception in the blind. Current Biology 14:121–124.

Roeder B, Stock O, Bien S, Neville H, Roesler F (2002) Speech processing activates visual cortex in congenitally blind humans: Plasticity of language functions in blind adults. European Journal of Neuroscience 16:930–936.

Rogalsky C, Basilakos A, Rorden C, Pillay S, LaCroix AN, Keator L, Mickelsen S, Anderson SW, Love T, Fridriksson J, Binder J, Hickok G (2022) The neuroanatomy of speech processing: A large-scale lesion study. Journal of Cognitive Neuroscience:1–21.

Rolls ET, Huang C-C, Lin C-P, Feng J, Joliot M (2020) Automated anatomical labelling atlas 3. NeuroImage 206:116189.

Roman IR, Roman AS, Kim JC, Large EW (2023) Hebbian learning with elasticity explains how the spontaneous motor tempo affects music performance synchronization. PLoS Comput Biol 19:e1011154.

Rosanova M, Casali A, Bellina V, Resta F, Mariotti M, Massimini M (2009) Natural frequencies of human corticothalamic circuits. Journal of Neuroscience 29:7679–7685.

Rosen S (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences 336:367–373.

Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20:53–65.

Roux F, Parish G, Chelvarajah R, Rollings DT, Sawlani V, Hamer H, Gollwitzer S, Kreiselmeyer G, Ter Wal MJ, Kolibius L, Staresina BP, Wimber M, Self MW, Hanslmayr S (2022) Oscillations support short latency co-firing of neurons during human episodic memory formation. eLife 11:e78109.

Samaha J, Postle BR (2015) The speed of alpha-band oscillations predicts the temporal resolution of visual perception. Current Biology 25:2985–2990.

Sassenhagen J, Alday PM (2016) A common misapplication of statistical inference: Nuisance control with null-hypothesis significance tests. Brain and Language 162:42–45.

Saur D, Kreher BW, 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. Proc Natl Acad Sci USA 105:18035–18040.

Schepers IM, Hipp JF, Schneider TR, Roeder B, Engel AK (2012) Functionally specific oscillatory activity correlates between visual and auditory cortex in the blind. Brain 135:922–934.

Scheurich R, Zamm A, Palmer C (2018) Tapping into rate flexibility: Musical training facilitates synchronization around spontaneous production rates. Front Psychol 9:458.

Schiavetti N (1992) Scaling procedures for the measurement of speech intelligibility. In: Studies in Speech Pathology and Clinical Linguistics, pp 11–34. Amsterdam: John Benjamins Publishing Company.

Schiavetti N, Sitler RW, Metz DE, Houde RA (1984) Prediction of contextual speech intelligibility from isolated word intelligibility measures. J Speech Lang Hear Res 27:623–626.

Schielzeth H (2010) Simple means to improve the interpretability of regression coefficients: *Interpretation of regression coefficients*. Methods in Ecology and Evolution 1:103–113.

Schmidt B, Kanis H, Holroyd CB, Miltner WHR, Hewig J (2018) Anxious gambling: Anxiety is associated with higher frontal midline theta predicting less risky decisions. Psychophysiology 55:e13210.

Schneider M, Broggini AC, Dann B, Tzanou A, Uran C, Sheshadri S, Scherberger H, Vinck M (2021) A mechanism for inter-areal coherence through communication based on connectivity and oscillatory power. Neuron 109:4050-4067.e12.

Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. Trends in Neurosciences 32:9–18.

Schroeder CE, Steinschneider M, Javitt DC, Tenke CE, Givre SJ, Mehta AD, Simpson GV, Arezzo JC, Vaughan HG (1995) Localization of ERP generators and identification of underlying neural processes. Electroencephalography and clinical neurophysiology Supplement 44:55–75.

Schroeder CE, Wilson DA, Radman T, Scharfman H, Lakatos P (2010) Dynamics of Active Sensing and perceptual selection. Current Opinion in Neurobiology 20:172–176.

Schroeder SCY, Ball F, Busch NA (2018) The role of alpha oscillations in distractor inhibition during memory retention. Eur J Neurosci 48:2516–2526.

Schubert J, Schmidt F, Gehmacher Q, Bresgen A, Weisz N (2023) Cortical speech tracking is related to individual prediction tendencies. Cerebral Cortex 33:6608–6619.

Schubert JTW, Buchholz VN, Foecker J, Engel AK, Roeder B, Heed T (2015) Oscillatory activity reflects differential use of spatial reference frames by sighted and blind individuals in tactile attention. NeuroImage 117:417–428.

Schubotz RI (2007) Prediction of external events with our motor system: towards a new framework. Trends in Cognitive Sciences 11:211–218.

Schwartze M, Rothermich K, Kotz SA (2012a) Functional dissociation of pre-SMA and SMA-proper in temporal processing. NeuroImage 60:290–298.

Schwartze M, Tavano A, Schröger E, Kotz SA (2012b) Temporal aspects of prediction in audition: Cortical and subcortical neural mechanisms. International Journal of Psychophysiology 83:200–207.

Scott SK, McGettigan C, Eisner F (2009) A little more conversation, a little less action — candidate roles for the motor cortex in speech perception. Nat Rev Neurosci 10:295–302.

Sedley W, Gander PE, Kumar S, Kovach CK, Oya H, Kawasaki H, Howard MA, Griffiths TD (2016) Neural signatures of perceptual inference. Elife 5:e11476.

Shackman AJ, Salomons TV, Slagter HA, Fox AS, Winter JJ, Davidson RJ (2011) The integration of negative affect, pain and cognitive control in the cingulate cortex. Nat Rev Neurosci 12:154–167.

Shaikh UU, Asif Z (2022) Persistence and dropout in higher online education: Review and categorization of factors. Front Psychol 13:902070.

Shamma S (2001) On the role of space and time in auditory processing. Trends in Cognitive Sciences 5:340–348.

Shannon CE (1948) A mathematical theory of communication. The Bell system technical journal 27:279–423.

Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M (1995) Speech Recognition with Primarily Temporal Cues. Science 270:303–304.

Sharp DJ, Bonnelle V, De Boissezon X, Beckmann CF, James SG, Patel MC, Mehta MA (2010) Distinct frontal systems for response inhibition, attentional capture, and error processing. Proc Natl Acad Sci USA 107:6106–6111.

Shimony JS, Burton H, Epstein AA, McLaren DG, Sun SW, Snyder AZ (2006) Diffusion Tensor Imaging Reveals White Matter Reorganization in Early Blind Humans. Cerebral Cortex 16:1653–1661.

Siegel M, Donner TH, Engel AK (2012) Spectral fingerprints of large-scale neuronal interactions. Nature Reviews Neuroscience 13:121–134.

Singer W (2013) Cortical dynamics revisited. Trends in Cognitive Sciences 17:616–626.

Singer W (2018) Neuronal oscillations: unavoidable and useful? Eur J Neurosci 48:2389–2398.

Singer W, Gray CM (1995) Visual feature integration and the temporal correlation hypothesis. Annual review of neuroscience 18:555–586.

Smit CM, Wright MJ, Hansell NK, Geffen GM, Martin NG (2006) Genetic variation of individual alpha frequency (IAF) and alpha power in a large adolescent twin sample. International Journal of Psychophysiology 61:235–243.

Smith NJ, Levy R (2013) The effect of word predictability on reading time is logarithmic. Cognition 128:302–319.

Smith S, Nichols T (2009) Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. NeuroImage 44:83–98.

Smith SM (2002) Fast robust automated brain extraction. Hum Brain Mapp 17:143–155.

Smith SM, Fox PT, Miller KL, Glahn DC, Fox PM, Mackay CE, Filippini N, Watkins KE, Toro R, Laird AR, Beckmann CF (2009) Correspondence of the brain's functional architecture during activation and rest. Proceedings of the National Academy of Sciences 106:13040–13045.

Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, Watkins KE, Ciccarelli O, Cader MZ, Matthews PM, Behrens TEJ (2006) Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. NeuroImage 31:1487–1505.

Smith ZM, Delgutte B, Oxenham AJ (2002) Chimaeric sounds reveal dichotomies in auditory perception. Nature 416:87–90.

Sobin C, Alpert M (1999) Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. Journal of psycholinguistic research 28:347–365.

Song J, Iverson P (2018) Listening effort during speech perception enhances auditory and lexical processing for non-native listeners and accents. Cognition 179:163–170.

Song S-K, Sun S-W, Ramsbottom MJ, Chang C, Russell J, Cross AH (2002) Dysmyelination revealed through MRI as increased radial (but unchanged axial) diffusion of water. NeuroImage 17:1429–1436.

Song S-K, Yoshino J, Le TQ, Lin S-J, Sun S-W, Cross AH, Armstrong RC (2005) Demyelination increases radial diffusivity in corpus callosum of mouse brain. NeuroImage 26:132–140.

Sormaz M, Murphy C, Wang H, Hymers M, Karapanagiotidis T, Poerio G, Margulies DS, Jefferies E, Smallwood J (2018) Default mode network can support the level of detail in experience during active task states. Proc Natl Acad Sci USA 115:9318–9323.

Spivey-Knowlton MJ, Trueswell JC, Tanenhaus MK (1993) Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale 47:276–309.

Stefanics G, Hangya B, Hernádi I, Winkler I, Lakatos P, Ulbert I (2010) Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. J Neurosci 30:13578–13585.

Steinmetzger K, Rosen S (2017) Effects of acoustic periodicity and intelligibility on the neural oscillations in response to speech. Neuropsychologia 95:173–181.

Sterr A, Müller MM, Elbert T, Rockstroh B, Pantev C, Taub E (1998a) Perceptual correlates of changes in cortical representation of fingers in blind multifinger Braille readers. J Neurosci 18:4417–4423.

Sterr A, Müller MM, Elbert T, Rockstroh B, Pantev C, Taub E (1998b) Changed perceptions in Braille readers. Nature 391:134–135.

Stevens AA, Weaver K (2005) Auditory perceptual consolidation in early-onset blindness. Neuropsychologia 43:1901–1910.

Stevens B, Porta S, Haak LL, Gallo V, Fields RD (2002) Adenosine: A neuron-glial transmitter promoting myelination in the CNS in response to action potentials. Neuron 36:855–868.

Stokes RC, Venezia JH, Hickok G (2019) The motor system's [modest] contribution to speech perception. Psychon Bull Rev 26:1354–1366.

Strait DL, Hornickel J, Kraus N (2011) Subcortical processing of speech regularities underlies reading and music aptitude in children. Behav Brain Funct 7:44.

Strauß A, Schwartz J-L (2017) The syllable in the light of motor skills and neural oscillations. Language, Cognition and Neuroscience 32:562–569.

Striem-Amit E, Ovadia-Caro S, Caramazza A, Margulies DS, Villringer A, Amedi A (2015) Functional connectivity of visual cortex in the blind follows retinotopic organization principles. Brain 138:1679–1695.

Striem-Amit E, Wang X, Bi Y, Caramazza A (2018) Neural representation of visual concepts in people born blind. Nature Communications 9:1–12.

Strogatz SH, Stewart I (1993) Coupled oscillators and biological synchronization. Scientific American 269:102–109.

Swann NC, Cai W, Conner CR, Pieters TA, Claffey MP, George JS, Aron AR, Tandon N (2012) Roles for the pre-supplementary motor area and the right inferior frontal gyrus in stopping action: Electrophysiological responses and functional and structural connectivity. NeuroImage 59:2860–2870.

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.

Taylor WL (1953) "Cloze Procedure": A new tool for measuring readability. Journalism Quarterly 30:415–433.

Ten Hove P, van der Meji H (2015) Like it or not. What characterizes YouTube's more popular instructional videos? Technical communication 62:48–62.

ten Oever S, Martin AE (2021) An oscillating computational model can track pseudo-rhythmic speech by using linguistic predictions. eLife 10:e68066.

ten Oever S, Sack AT (2015) Oscillatory phase shapes syllable perception. Proceedings of the National Academy of Sciences 112:15833–15837.

ten Oever S, Schroeder CE, Poeppel D, Van Atteveldt N, Mehta AD, Mégevand P, Groppe DM, Zion-Golumbic E (2017) Low-frequency cortical oscillations entrain to subthreshold rhythmic auditory stimuli. J Neurosci 37:4903–4912.

ten Oever S, Van Atteveldt N, Sack AT (2015) Increased stimulus expectancy triggers low-frequency phase reset during restricted vigilance. Journal of Cognitive Neuroscience 27:1811–1822.

Teng X, Cogan GB, Poeppel D (2019) Speech fine structure contains critical temporal cues to support speech segmentation. NeuroImage 202:116152.

Teng X, Poeppel D (2020) Theta and gamma bands encode acoustic dynamics over wide-ranging timescales. Cerebral Cortex 30:2600–2614.

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 Jensen O, ed. PLOS Biology 15:e2000812.

Thiebaut de Schotten M, ffytche DH, Bizzi A, Dell'Acqua F, Allin M, Walshe M, Murray R, Williams SC, Murphy DGM, Catani M (2011) Atlasing location, asymmetry and inter-subject variability of white matter tracts in the human brain with MR diffusion tractography. NeuroImage 54:49–59.

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.

Töllner T, Wang Y, Makeig S, Müller HJ, Jung T-P, Gramann K (2017) Two independent frontal midline theta oscillations during conflict detection and adaptation in a Simon-type manual reaching task. J Neurosci 37:2504–2515.

Tourville JA, Guenther FH (2011) The DIVA model: A neural theory of speech acquisition and production. Language and Cognitive Processes 26:952–981.

Trimper JB, Stefanescu RA, Manns JR (2014) Recognition memory and theta–gamma interactions in the hippocampus. Hippocampus 24:341–353.

Trouvain J (2007) On the comprehension of extremely fast synthetic speech. Saarland Working Papers in Linguistics 1:5–13.

Tsao Y-C, Weismer G (1997) Interspeaker variation in habitual speaking rate: Evidence for a neuromuscular component. 40:9.

Tulving E, Gold C (1963) Stimulus information and contextual information as determinants of tachistoscopic recognition of words. Journal of Experimental Psychology 66:319–327.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage 15:273–289.

Van Ackeren MJ, Barbero FM, 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.

Van Beijsterveldt CEM, Van Baal GCM (2002) Twin and family studies of the human electroencephalogram: a review and a meta-analysis. Biological Psychology 61:111–138.

Van Bree S, Alamia A, Zoefel B (2022) Oscillation or not—Why we can and need to know (commentary on Doelling and Assaneo, 2021). Eur J of Neuroscience 55:201–204.

Van Bree S, Levenstein D, Krause M, Voytek B, Gao R (2024) Decoupling Measurements and Processes: On the Epiphenomenon Debate Surrounding Brain Oscillations in Field Potentials.

Van Der Horst R, Leeuw AR, Dreschler WA (1999) Importance of temporal-envelope cues in consonant recognition. The Journal of the Acoustical Society of America 105:1801–1809.

van Kerkoerle T, Self MW, Dagnino B, Gariel-Mathis M-A, Poort J, van der Togt C, Roelfsema PR (2014) Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. Proc Natl Acad Sci USA 111:14332–14341.

Van Petten C, Luka BJ (2012) Prediction during language comprehension: Benefits, costs, and ERP components. International Journal of Psychophysiology 83:176–190.

Van Veen BD, Van Drongelen W, Yuchtman M, Suzuki A (1997) Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans Biomed Eng 44:867–880.

VanRullen R (2016) Perceptual Cycles. Trends in Cognitive Sciences 20:723–735.

VanRullen R (2018) Perceptual rhythms. In: Stevens' handbook of experimental psychology and cognitive neuroscience (Wixted JT, ed), pp 1–44. Hoboken, NJ, USA: John Wiley & Sons, Inc.

VanRullen R, Koch C (2003) Is perception discrete or continuous? Trends in Cognitive Sciences 7:207–213.

Varnet L, Ortiz-Barajas MC, Erra RG, Gervain J, Lorenzi C (2017) A cross-linguistic study of speech modulation spectra. The Journal of the Acoustical Society of America 142:1976–1989.

Verschueren E, Gillis M, Decruy L, Vanthornhout J, Francart T (2022) Speech understanding oppositely affects acoustic and linguistic neural tracking in a speech rate manipulation paradigm. J Neurosci 42:7442–7453.

Viemeister NF (1979) Temporal modulation transfer functions based upon modulation thresholds. The Journal of the Acoustical Society of America 66:1364–1380.

Vincent JL, Patel GH, Fox MD, Snyder AZ, Baker JT, Van Essen DC, Zempel JM, Snyder LH, Corbetta M, Raichle ME (2007) Intrinsic functional architecture in the anaesthetized monkey brain. Nature 447:83–86.

Visser M, Lambon Ralph MA (2011) Differential contributions of bilateral ventral anterior temporal lobe and left anterior superior temporal gyrus to semantic processes. Journal of Cognitive Neuroscience 23:3121–3131.

Voss P (2019) Brain (re)organization following visual loss. WIREs Cogn Sci 10:e1468.

Voss P, Zatorre RJ (2012) Organization and reorganization of sensory-deprived cortex. Current Biology 22:R168–R173.

Wake H, Ortiz FC, Woo DH, Lee PR, Angulo MC, Fields RD (2015) Nonsynaptic junctions on myelinating glia promote preferential myelination of electrically active axons. Nat Commun 6:7844.

Wang D, Qin W, Liu Y, Zhang Y, Jiang T, Yu C (2013) Altered white matter integrity in the congenital and late blind people. Neural Plasticity 2013:1–8.

Warren JE, Sauter DA, Eisner F, Wiland J, Dresner MA, Wise RJS, Rosen S, Scott SK (2006) Positive emotions preferentially engage an auditory-motor "mirror" system. Journal of Neuroscience 26:13067–13075.

Watkins KE, Shakespeare TJ, O'Donoghue MC, Alexander I, Ragge N, Cowey A, Bridge H (2013) Early auditory processing in area V5/MT+ of the congenitally blind brain. Journal of Neuroscience 33:18242–18246.

Westner BU, Dalal SS, Gramfort A, Litvak V, Mosher JC, Oostenveld R, Schoffelen J-M (2022) A unified view on beamformers for M/EEG source reconstruction. NeuroImage 246:118789.

Will U, Berg E (2007) Brain wave synchronization and entrainment to periodic acoustic stimuli. Neuroscience Letters 424:55–60.

Wilsch A, Neuling T, Obleser J, Herrmann CS (2018) Transcranial alternating current stimulation with speech envelopes modulates speech comprehension. NeuroImage 172:766–774.

Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004) Listening to speech activates motor areas involved in speech production. Nat Neurosci 7:701–702.

Wu Z-M, Chen M-L, Wu X-H, Li L (2014) Interaction between auditory and motor systems in speech perception. Neurosci Bull 30:490–496.

Yang X, Wang K, Shamma SA (1992) Auditory representations of acoustic signals. IEEE Trans Inform Theory 38:824–839.

Yu C, Liu Y, Li J, Zhou Y, Wang K, Tian L, Qin W, Jiang T, Li K (2008) Altered functional connectivity of primary visual cortex in early blindness. Hum Brain Mapp 29:533–543.

Zalta A, Petkoski S, Morillon B (2020) Natural rhythms of periodic temporal attention. Nat Commun 11:1051.

Zamm A, Wellman C, Palmer C (2016) Endogenous rhythms influence interpersonal synchrony. Journal of Experimental Psychology: Human Perception and Performance 42:611–616.

Zamorano AM, Cifre I, Montoya P, Riquelme I, Kleber B (2017) Insula-based networks in professional musicians: Evidence for increased functional connectivity during resting state fMRI. Human Brain Mapping 38:4834–4849.

Zatorre R, Evans A, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch discrimination in speech processing. Science 256:846–849.

Zatorre RJ, Chen JL, Penhune VB (2007) When the brain plays music: auditory–motor interactions in music perception and production. Nat Rev Neurosci 8:547–558.

Zatorre RJ, Fields RD, Johansen-Berg H (2012) Plasticity in gray and white: Neuroimaging changes in brain structure during learning. Nat Neurosci 15:528–536.

Zhang L, Jiang W, Shu H, Zhang Y (2019) Congenital blindness enhances perception of musical rhythm more than melody in Mandarin speakers. The Journal of the Acoustical Society of America 145:EL354–EL359.

Zhang LI, Bao S, Merzenich MM (2001) Persistent and specific influences of early acoustic environments on primary auditory cortex. Nat Neurosci 4:1123–1130.

Zhang Y, Frassinelli D, Tuomainen J, Skipper JI, Vigliocco G (2021) More than words: word predictability, prosody, gesture and mouth movements in natural language comprehension. Proc R Soc B 288:20210500.

Zhang Y, Zou J, 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.

Zhao L, Silva AB, Kurteff GL, Chang EF (2023) Inhibitory control of speech production in the human premotor frontal cortex.

Zion Golumbic E, Cogan GB, Schroeder CE, Poeppel D (2013) Visual input enhances selective speech envelope tracking in auditory cortex at a “Cocktail Party.” Journal of Neuroscience 33:1417–1426.

Zion Golumbic EM, Poeppel D, Schroeder CE (2012) Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective. Brain and Language 122:151–161.

Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ, Poeppel D, Schroeder CE (2013) Mechanisms underlying selective neuronal tracking of attended speech at a “Cocktail Party.” Neuron 77:980–991.

Zoefel B, Allard I, Anil M, Davis MH (2020) Perception of rhythmic speech is modulated by focal bilateral transcranial alternating current stimulation. Journal of Cognitive Neuroscience 32:226–240.

Zoefel B, Archer-Boyd A, Davis MH (2018a) Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. Current Biology 28:401-408.e5.

Zoefel B, Kösem A (2024) Neural tracking of continuous acoustics: properties, speech-specificity and open questions. Eur J of Neuroscience 59:394–414.

Zoefel B, Ten Oever S, Sack AT (2018b) The involvement of endogenous neural oscillations in the processing of rhythmic input: More than a regular repetition of evoked neural responses. Front Neurosci 12:95.

Zoefel B, VanRullen R (2015) The role of high-level processes for oscillatory phase entrainment to speech sound. Frontiers in Human Neuroscience 9:651.

Zuk J, Loui P, Guenther F (2022) Neural control of speaking and singing: The DIVA model for singing. PsyArXiv.

8 List of figures

9 List of tables

10 List of abbreviations

2IFC ..... *two-interval forced choice*
AAL ..... *automated anatomical labelling*
BA ..... *brodman area*
CB ..... *congenitally blind*
DAT ..... *dynamic attending theory*
DTI ..... *diffusion-tensor imaging*
ECoG ..... *electrocorticography*
EEG ..... *electroencephalography*
FA ..... *fractional anisotropy*
FDR ..... *false discovery rate*
FWE ..... *family-wise error*
GAMM ..... *generalized additive mixed-effects model*
GLMM ..... *generalized linear mixed-effects model*
GMM ..... *gaussian mixture model*
HG ..... *heschl's gyrus*
ICA ..... *independent component analysis*
IFG ..... *inferior frontal gyrus*
LCMV ..... *linearly-constrained minimum variance*
MEG ..... *magnetoencephalography*
MNI ..... *montreal neurological institute*
MRI ..... *magnetic resonance imaging*
PC ..... *principal components*
PLV ..... *phase-locking value*
PSOLA ..... *pitch synchronous overlap and add algorithm*
pSTG ..... *posterior superior temporal gyrus*
RD ..... *radial diffusivity*
RNN ..... *recurrent neural network*
ROI ..... *regions of interest*
SMA ..... *supplementary motor area*
SMC ..... *sensory motor cortex*
SSER ..... *steady-state evoked responses*
SSS-test ..... spontaneous speech synchronization test
TRF ..... *temporal response function*

11 Summary

Introduction

Brain rhythms, pervasive throughout cortex, are linked to cognitive functions (Buzsáki, 2004; Buzsáki et al., 2013; Singer, 2018). In auditory cortex, two distinct regimes of brain rhythms prevail: slower delta-theta (~2-8 Hz) and faster gamma (~30-40 Hz) rhythms (Boemio et al., 2005; Giraud et al., 2007; for review see: Edwards and Chang, 2013). Speech signals contain information at similar timescales, with syllables occurring at 4-5 Hz and phonemes in the gamma range. During speech perception, auditory theta rhythms are proposed to synchronize to temporal fluctuations in the speech amplitude envelope by anticipating the envelope's phase, presumably facilitating syllabic segmentation (Giraud and Poeppel, 2012; Gross et al., 2013). This "speech entrainment" may be optimal around the nominal speech rate (~5 Hz).

Despite the apparent (quasi-)rhythmicity in speech, there is considerable variability in its temporal regularity, challenging the assumption of oscillatory auditory units synchronizing to the amplitude envelope. Although the auditory system has a preference for certain frequencies, it demonstrates flexibility by tracking sentences exceeding the average speech modulation frequency. However, this flexibility appears to be constrained, as both speech tracking and comprehension decline beyond approximately ~8-9 Hz (Ahissar et al., 2001; Brungart et al., 2007; Ghitza and Greenberg, 2009; Hincapié Casas et al., 2021). Beyond internal auditory characteristics, we hypothesize that the flexibility of auditory cortex may be influenced by other modalities, e.g. the motor cortex.

The motor system's role in speech perception is subject to ongoing debates (Pulvermüller and Fadiga, 2010; Hickok et al., 2011; Pickering and Gambi, 2018). One hypothesis posits that motor areas contribute to perception by predictively modulating sensory excitability in auditory cortex through oscillatory coupling mechanisms, leveraging the motor system's temporal processing capabilities (Arnal and Giraud, 2012; Park et al., 2015; Morillon and Baillet, 2017; Haegens and Zion Golumbic, 2018; Rimmele et al., 2018). This auditory-motor interaction may be constrained in the frequency domain, with strongest coupling—and likely temporal predictions—occurring around 4.5 Hz (Assaneo and Poeppel, 2018).

A major challenge to current oscillatory speech theories consists in demonstrating actual speech entrainment, i.e. a process in which endogenous theta rhythms align to the amplitude envelope's phase, instead of passively tracking it (Zoefel et al., 2018b; Obleser and Kayser, 2019) This thesis aims to contribute to this ongoing debate by answering the following questions: (1) Do individual differences in endogenous theta brain rhythms account for behavioral and neural differences among individuals? More specifically, do spectral profiles change with brain reorganization, particularly in populations exhibiting adaptive behaviors such as congenitally blind individuals? (2) Do

we observe individual differences in the peak frequencies of preferred auditory and motor rhythms and in auditory-motor synchronization? Are speech comprehension abilities affected by these individual peak parameters of the auditory and motor systems? (3) Are the effects of endogenous brain rhythms in auditory and motor cortices on speech comprehension related to speech tracking?

Experiments and Results

The first study of this thesis quantified how brain reorganization in congenital blindness affects endogenous brain rhythms using resting-state MEG data. Employing spectral profile and classification analyses, altered spectral properties were observed in visual brain areas among congenitally blind individuals, indicating changed alpha-gamma circuits. Additionally, auditory and right frontal brain areas displayed increased power in the theta-to-beta frequencies in the congenitally blind compared to the sighted individuals. Notably, a second, faster theta cluster was identified in Heschl's gyrus (HG) in congenitally blind individuals. This finding together with the ability of congenitally blind individuals to comprehend ultra-fast speech provides some initial support for the hypothesis that individual theta peak frequencies in HG determine the temporal resolution of speech comprehension, with faster peak frequencies enabling comprehension of faster speech. Overall, this suggests that visual deprivation alters spectral profiles not only in visual but also other cortical regions, potentially reflecting adaptations in both structure and function.

Study 2 evaluated participants' ability to comprehend speech across various syllabic rates in two behavioral experiments. The primary goal was to investigate how speech comprehension related to individual differences in auditory-motor synchronization, preferred auditory and motor frequencies, and higher-level factors. Across experiments and participants, speech comprehension declined with increasing syllabic rate of speech. However, individuals with faster preferred motor rates, i.e. spontaneous motor production rate, and stronger auditory-motor synchronization demonstrated enhanced comprehension. Similarly, better working memory capacity (Digit span) and increased sentence predictability facilitated comprehension, especially at higher syllabic rates. These findings support a model wherein the motor system's temporal dynamics and auditory-motor synchronization affect speech processing. They also underscore the complexity of naturalistic speech comprehension, considering the impact of other cognitive domains (working memory) and linguistic predictability.

Having established the effects of temporal motor dynamics and auditory-motor synchronization, Study 3 sought to specify the computational level at which these parameters facilitate comprehension. Specifically, if endogenous brain rhythms and auditory-motor synchronization predict speech tracking, it is plausible that this tracking reflects speech entrainment in the narrow sense (Obleser and Kayser, 2019). In a combined behavioral and MEG experiment we revealed that speech tracking in posterior temporal gyrus (pSTG) was predicted by the endogenous theta peak frequency of

pSTG. Furthermore, peak frequencies of endogenous motor rhythms (IFG and SMA) and phase-phase auditory-motor coupling influenced speech tracking. These motor effects were evident only in individuals with strong behavioral auditory-motor synchronization. This link between the peak frequency and phase-phase coupling of endogenous auditory and motor brain rhythms and speech tracking supports the hypothesis of an oscillatory process underlying speech comprehension. Importantly, the neurophysiological results align with the behavioral findings observed in Studies 2 and 3, where the spontaneous motor production rate and auditory-motor synchronization predicted comprehension. In addition, Study 3 also revealed a main effect of the preferred auditory rate. Overall, our findings endorse an oscillatory model of speech perception and auditory-motor interactions, highlighting distinct processes for individuals with high versus low behavioral speech production-perception synchronization.

Discussion

The findings presented in this thesis demonstrate that (1) brain rhythms exhibit plasticity, reflected in changed spectral profiles alongside brain reorganization and in nuanced individual differences in spectral peak parameters, (2) preferred behavioral and endogenous neural auditory and motor brain rhythms predict speech comprehension and tracking, respectively, and (3) individual differences highlight the possibility of diverse processing strategies.

Brain rhythms at rest presumably reflect the functional neural architecture that subserves cognition (Singer, 2013), with spectral fingerprints reflecting distinct spectral components present in brain regions (Keitel and Gross, 2016). Here, auditory theta brain rhythms demonstrated adaptation in congenital blindness, differed in peak frequency between normally sighted individuals, and predicted speech segmentation. As such, our findings support the existence of endogenous theta brain rhythms in human auditory cortex and, importantly, endorse their engagement in speech segmentation, a cornerstone of oscillatory speech theory (Giraud and Poeppel, 2012).

Speech processing is shaped by different top-down predictions (phonetic, semantic, syntactic predictions, temporal predictions) at multiple computational levels along the speech processing hierarchy. Our results suggest that motor activity, possibly reflecting temporal predictions, influences speech segmentation in auditory cortex (pSTG) through an intricate auditory-motor network including, but likely not limited to, IFG and SMA. Supporting oscillatory models of speech segmentation, we demonstrate an interaction between oscillatory phase dynamics in auditory (pSTG, HG) and motor (IFG, SMA) cortex, particularly highlighting enhanced tracking modulation through the motor system when auditory-motor phase coupling is strong.

The differential motor recruitment, contingent on behavioral synchronization strength, suggests diverse processing strategies among individuals, underscoring the importance of considering individual differences. Lower behavioral auditory-motor synchronization led to stronger reliance on the

auditory system, possibly indicating closer monitoring of the auditory signal and perhaps a greater emphasis on bottom-up processing. Conversely, stronger synchronization was associated with additional recruitment of the motor system, likely reflecting temporal predictions. Apart from these auditory-motor dynamics, speech comprehension likely benefits from synergistic or compensatory mechanisms. While a synergy of motor-based temporal and linguistic predictions may optimize comprehension, working memory likely compensates in their absence. We speculate that different predictions may arise from distinct mechanisms, such as semantic predictions enhancing processing via gain modulation and temporal predictions via phase-resetting of slow rhythms in auditory cortex.

12 Zusammenfassung

Einleitung

Hirnrhythmen sind im gesamten Kortex verbreitet und mit kognitiven Funktionen verbunden (Buzsáki, 2004; Buzsáki et al., 2013; Singer, 2018). Im auditiven Kortex herrschen zwei verschiedene Regime von Hirnrhythmen vor: langsamere Delta-Theta- (~2-8 Hz) und schnelleren Gamma-Rhythmen (~30-40 Hz)(Boemio et al., 2005; Giraud et al., 2007; for review see: Edwards and Chang, 2013). Sprachsignale enthalten eine zeitliche Struktur auf ähnlichen Zeitskalen, wobei Silben mit Frequenzen von 4-5 Hz und Phoneme im Gamma-Bereich auftreten. Es wird angenommen, dass auditive Theta-Rhythmen während der Sprachwahrnehmung durch Phasen-Antizipation mit zeitlichen Fluktuationen in der Amplitudenhüllkurve des Sprachsignals synchronisieren. Dieses „Sprach-Entrainment“ erleichtert vermutlich die Segmentierung des Sprachsignals (Giraud and Poeppel, 2012; Gross et al., 2013) und ist optimal um die nominale Sprechrate (~5 Hz).

Trotz der oft angenommenen (quasi-)rhythmischen Struktur von Sprache, weist ihre zeitlichen Struktur erhebliche Variation auf. Dies stellt die Annahme, dass ein oszillierender auditiver Kortex mit der Amplitudenhüllkurve synchronisiert, in Frage. Obwohl das auditive System verstärkt auf gewisse Stimulationsfrequenzen reagiert, weist es ebenso Flexibilität auf, da eine Synchronisierung mit Sätzen oberhalb der durchschnittlichen Sprechmodulationsfrequenz möglich ist. Diese Flexibilität scheint jedoch limitiert, da sowohl das Synchronisieren zu als auch das Verstehen von Sprache jenseits von etwa ~8-9 Hz einbrechen (Ahissar et al., 2001; Brungart et al., 2007; Ghitza and Greenberg, 2009; Hincapié Casas et al., 2021). Neben den temporalen Eigenschaften des auditiven Systems stellen wir die Hypothese auf, dass der Grad seiner Flexibilität von anderen Modalitäten beeinflusst werden kann, zum Beispiel vom motorischen Kortex.

Die Rolle des motorischen Systems bei der Sprachwahrnehmung ist Gegenstand anhaltender Debatten (Pulvermüller and Fadiga, 2010; Hickok et al., 2011; Pickering and Gambi, 2018). Laut einer Hypothese tragen motorische Hirnareale zur Wahrnehmung bei, indem sie die sensorische Erregbarkeit des auditiven Systems durch prädiktive oszillatorische Kopplungsmechanismen zwischen den beiden Systemen beeinflussen (Arnal and Giraud, 2012; Park et al., 2015; Morillon and Baillet, 2017; Haegens and Zion Golumbic, 2018; Rimmele et al., 2018). Dabei wird vermutet, dass die Fähigkeit des motorischen Systems zeitliche Vorhersagen zu generieren eine zentrale Rolle für diesen prädiktiven Mechanismus spielen. Diese audio-motor Interaktion scheint jedoch nicht gleichermaßen über alle Stimulationsfrequenzen zu funktionieren. Untersuchungen haben stattdessen gezeigt, dass die Interaktion—und wahrscheinlich die Übermittlung zeitlicher Vorhersagen—bei einer Frequenz von 4.5 Hz optimal ist (Assaneo and Poeppel, 2018).

Eine zentrale Herausforderung für aktuelle oszillatorische Sprachtheorien besteht darin, wahrhaftiges Sprach-Entrainment nachzuweisen, d.h. einen Prozess, in dem endogene Theta-Hirnrhythmen die Phase der Amplitudenhüllkurve antizipieren, anstatt ihr passiv zu folgen (Zoefel et al., 2018b; Obleser and Kayser, 2019). Diese Dissertation soll einen Beitrag zu dieser anhaltenden Debatte leisten, indem sie folgende Fragen beantwortet: (1) Sind individuelle Unterschiede in den endogenen Theta-Gehirnrhythmen für Verhaltensunterschiede zwischen Individuen verantwortlich? Genauer gesagt, verändern sich die spektralen Profile mit der Reorganisation des Gehirns, insbesondere bei Populationen, die adaptive Verhaltensweisen zeigen, wie z.B. geburtsblinde Individuen? (2) Beobachten wir individuelle Unterschiede in den Peak-Frequenzen der bevorzugten auditiven und motorischen Rhythmen und in der audio-motor Synchronisierung? Beeinflussen diese individuellen Peakparameter der auditiven und motorischen Systeme das Sprachverstehens? (3) Stehen die Auswirkungen der endogenen Hirnrhythmen der auditiven und motorischen Kortizes auf das Sprachverständnis in Zusammenhang mit dem Sprachtracking?

Experimente und Ergebnisse

In der ersten Studie dieser Arbeit wurden MEG-Ruhemessungsdaten analysiert, um zu quantifizieren, wie sich die Reorganisation des Gehirns bei geburtsblinden Individuen auf endogene Hirnrhythmen auswirkt. Die Verwendung von Spektralprofil- und Klassifikationsanalysen zeigte veränderte spektrale Eigenschaften in visuellen Hirnarealen in der Gruppe Geburtsblinder, was auf veränderte Alpha-Gamma-Schaltkreise hinweist. Zusätzlich zeigten auditive und rechtshemisphärische frontale Hirnareale eine erhöhte Amplitude in den Theta-bis-Beta-Frequenzen in der blinden relativ zur sehenden Gruppe. Zudem wies die Gruppe geburtsblinder Studienteilnehmer ein zweites, schnelleres Theta-Cluster im Heschl's Gyrus (HG) auf. Dieser Befund unterstützt die Hypothese, dass individuelle Theta-Peak-Frequenzen im HG die zeitliche Auflösung der Sprachwahrnehmung bestimmen, wobei schnellere Peak-Frequenzen ein Verständnis von schnellerer Sprache ermöglichen. Insgesamt legen die Ergebnisse nahe, dass visuelle Deprivation nicht nur spektrale Profile in visuellen, sondern auch in anderen kortikalen Regionen verändert. Eine solche Veränderung könnte möglicherweise Anpassungen in Struktur und Funktion widerspiegeln.

Studie 2 bewertete in zwei Verhaltensexperimenten, wie Teilnehmer Sprache mit verschiedenen Silbenraten verstehen. Das Hauptziel bestand darin, zu untersuchen, wie das Sprachverständnis durch individuelle Unterschiede in der audio-motor Synchronisation, präferierten auditiven und motorischen Frequenzen sowie anderen kognitiven Faktoren beeinflusst wird. Über Experimente und Teilnehmer hinweg wurde beobachtet, dass das Sprachverständnis mit zunehmender Silbenrate abnahm. Personen mit schnelleren bevorzugten Motorraten, gemessen als spontane Sprechrate, und stärkerer audio-motor Synchronisation zeigten jedoch eine verbesserte Verständnisleistung. Ebenso erleichterten eine bessere Arbeitsgedächtniskapazität (Digit Span) und eine höhere

Satzvorhersagbarkeit das Verständnis, insbesondere bei höheren Silbenraten. Diese Ergebnisse unterstützen ein Modell, in dem die zeitlichen Dynamiken des motorischen Systems und die audio-motor Synchronisation die auditive Sprachverarbeitung beeinflussen. Sie betonen ebenso die Komplexität des natürlichen Sprachverstehens, welches mit anderen kognitiven Domänen (Arbeitsgedächtnis) und der Vorhersagbarkeit von Sprache zu interagieren scheint.

Nachdem die Auswirkungen der temporalen Dynamiken des Motorsystems und der audio-motor Synchronisation etabliert wurden, strebte Studie 3 an, die computationale Ebene zu spezifizieren, auf der diese Parameter das Sprachverstehen beeinflussen. Wenn endogene Hirnrhythmen und audio-motor Synchronisation das Sprachtracking vorhersagen, ist es plausibel, dass dieses Tracking wahrhaftiges Sprach-Entrainment widerspiegelt (Giraud and Poeppel, 2012; Obleser and Kayser, 2019). In einem kombinierten Verhaltens- und MEG-Experiment zeigten wir, dass das Sprachtracking im posterioren superioren temporalen Gyrus (pSTG) von der endogenen Theta-Peak-Frequenz des pSTG vorhergesagt wird. Darüber hinaus beeinflussten Peak-Frequenzen endogener Motor-Rhythmen (IFG und SMA) und die audio-motor Kopplung (Phase-zu-Phase) das Sprachtracking. Eine Beteiligung des Motorsystems wurde jedoch lediglich bei Personen festgestellt, die starke audio-motor Synchronisation im Verhalten zeigten. Der Zusammenhang zwischen der Peak-Frequenz von endogenen auditiven und motorischen Hirnrhythmen und Sprachtracking unterstützen die Hypothese, dass dem Sprachverstehen ein Oszillationsprozess unterliegt. Die neurophysiologischen Ergebnisse stimmen mit den Verhaltensergebnissen der zweiten und der aktuellen Studie überein, bei denen die spontane Sprechrate und die audio-motor Synchronisation das Sprachverstehen vorhersagten. Darüber hinaus wurde in Studie 3 einen Haupteffekt der präferierten auditiven Rate beobachtet. Insgesamt sind die Ergebnisse im Einklang mit Oszillationsmodellen von Sprachwahrnehmung und audio-motor Interaktionen. Zudem heben sie hervor, dass Sprachwahrnehmung in Individuen mit starker und schwacher audio-motor Synchronisierung durch unterschiedliche Prozesse dominiert sein könnte.

Diskussion

Die Ergebnisse dieser Arbeit zeigen, dass (1) Hirnrhythmen Plastizität aufweisen, die sich in veränderten spektralen Profilen widerspiegelt und mit der Reorganisation des Gehirns einhergeht, sowie in nuancierten individuellen Unterschieden in den spektralen Peak-Parametern, (2) präferierte behaviorale und endogene neurale auditive und motorische Rhythmen Rückschlüsse auf Sprachverstehen und -tracking erlauben und (3) individuelle Unterschiede die Möglichkeit verschiedener Verarbeitungsstrategien aufzeigen.

Hirnrhythmen im Ruhezustand spiegeln die funktionelle neurale Architektur wider, welche kognitiven Prozessen unterliegt (Singer, 2013), wobei spektrale Profile unterschiedliche spektrale Komponenten individueller Hirnregionen widerspiegeln (Keitel and Gross, 2016). Diese Arbeit zeigte,

dass auditive Theta-Hirnrhythmen in kongenitaler Blindheit eine Veränderung durchliefen, sich in der Peak-Frequenz zwischen normal sehenden Individuen unterschieden und mit dem Prozess der Sprachsegmentation assoziiert waren. Unsere Ergebnisse unterstützen die Existenz von endogenen Theta-Rhythmen im menschlichen auditiven Kortex, sowie ihre Beteiligung an der Sprachsegmentierung, einem Eckpfeiler oszillatorischer Sprachtheorien (Giraud and Poeppel, 2012).

Die Sprachverarbeitung wird durch verschiedene Top-down Vorhersagen (phonetische, semantische, syntaktische und zeitliche Vorhersagen) auf mehreren computationalen Ebenen entlang der Sprachverarbeitungshierarchie beeinflusst. Unsere Ergebnisse deuten darauf hin, dass Motoraktivität, welche vermutlich zeitliche Vorhersagen reflektiert, die Sprachsegmentierung im auditiven Kortex (pSTG) durch ein komplexes auditiv-motorisches Netzwerk beeinflusst, welches IFG und SMA–und vermutlich weitere Areale–umfasst. Im Einklang mit Oszillationsmodellen der Sprachsegmentierung zeigen wir hier eine Interaktion zwischen oszillatorischen Phasendynamiken im auditiven (pSTG, HG) und motorischen (IFG, SMA) Kortex. Dabei konnte verstärktes Tracking durch das motorische System insbesondere beobachtet werden, wenn die audio-motor Phasenkopplung stark ist.

Die unterschiedliche Rekrutierung des Motorsystems, die von der Stärke der audio-motor Synchronisation im Verhalten abhängt, deutet auf verschiedene Verarbeitungsstrategien zwischen Individuen hin und unterstreicht die Relevanz individuelle Unterschiede zu berücksichtigen. Eine geringere audio-motor Synchronisation führte zu einer stärkeren Gewichtung des auditiven Systems, was möglicherweise auf einen stärkeren Fokus auf das akustische Signal und Bottom-up-Verarbeitung hinweisen könnte. Hingegen war eine stärkere Synchronisation im Verhalten mit einer zusätzlichen Rekrutierung des Motorsystems verbunden, wodurch vermutlich vermehrt zeitliche Vorhersagen übermittelt werden. Abgesehen von dieser audio-motor Dynamik wird das Sprachverstehen potenziell durch synergetische oder kompensatorische Mechanismen begünstigt. Während eine Synergie von motor-basierten zeitlichen und linguistischen Vorhersagen das Verstehen optimieren kann, könnte das Arbeitsgedächtnis ein Fehlen dieser Vorhersagen kompensieren. Es ist denkbar, dass verschiedenen Vorhersagen diverse Mechanismen unterliegen, wobei semantische Vorhersagen die Verarbeitung durch „modulation gain“ und zeitliche Vorhersagen durch „phase reset“ langsamer Rhythmen im auditorischen Kortex verbessern könnten.

13 List of publications

Lubinus, C., Keitel, A., Obleser, J., Poeppel, D., & Rimmele, J. M. (2023). Explaining flexible continuous speech comprehension from individual motor rhythms. Proceedings of the Royal Society B, 290(1994), 20222410.

Lubinus, C., Einhaeuser, W., Schiller, F., Kircher, T., Straube, B., & van Kemenade, B. M. (2022). Action-based predictions affect visual perception, neural processing, and pupil size, regardless of temporal predictability. *NeuroImage, 263, 119601.*

Rimmele, J.M., Kern, P., **Lubinus, C.**, Frieler, K., Poeppel, D., & Assaneo, M. F. (2021). Musical sophistication and speech auditory-motor coupling: easy tests for quick answers. Frontiers in Neuroscience, *in press*.

Lubinus, C., Orpella, J., Keitel, A., Gudi-Mindermann, H., Engel, A. K., Roeder, B., & Rimmele, J. M. (2021). Data-driven classification of spectral profiles reveals brain region-specific plasticity in blindness. *Cerebral Cortex*, *31*(5), 2505-2522.

Talks

2023 *Explaining flexible continuous speech comprehension from individual motor rhythms*, European Society for Cognitive Psychology (ESCOP), Porto, Portugal

2022 *Individual rhythms in speech comprehension: contributions of the motor system*, Linguistic colloquium, University of Mainz, Germany

2022 *Explaining flexible continuous speech comprehension from individual motor rhythms,* BAND Lab (Sonia Kotz), University of Maastricht, Netherlands

2021 *Individual rhythms during the comprehension of continuous speech*, CLaMe lab meeting, NYU, USA

Conference poster presentations

2023 *Comprehension and neuronal tracking of speeded speech – contributions of preferred rates and auditory-motor synchronization.* Society for the neurobiology of language (SNL), Marseille, France

2023 *Influence of auditory-motor synchronization and rate preferences during the comprehension of speeded speech*. Symposium for audio-motor integration for cognition, Queretaro, Mexico

2022 *Explaining flexible continuous speech comprehension from individual motor rhythms*. International Conference of Cognitive Neuroscience (ICON), Helsinki, Finland

2020 *Data-driven classification of spectral profiles reveals brain region-specific plasticity*. Annual meeting of the Cognitive Neuroscience Society (CNS), Boston, USA (virtual)

2019 *Data-driven classification of oscillatory profiles reveals region-specific plasticity*. Brain rhythms and cortical communication (BryCoCo), Tel-Aviv, Israel

2019 *Data-driven classification of region-specific spectral profiles in congenitally blind and sighted individuals*. Salzburg Mind and brain annual meeting (SAMBA), Salzburg, Austria.

2018 *Action-based predictions affect visual perception, neural processing, and pupil size, regardless of temporal predictability*. Psychology and Brain conference, Gießen, Germany