A comparison and classification of oscillatory characteristics in speech perception and covert speech

A comparison and classification of oscillatory characteristics in speech perception and covert speech[†]

Abstract

Covert speech, the mental imagery of speaking, has been studied increasingly to understand and decode thoughts in the context of brain-computer interfaces. In studies of speech comprehension, neural oscillations are thought to play a key role in the temporal encoding of speech. However, little is known about the role of oscillations in covert speech. In this study, we investigated the oscillatory involvements in covert speech and speech perception. Data were collected from 10 participants with 64 channel EEG. Participants heard the words, 'blue' and 'orange', and subsequently mentally rehearsed them. First, continuous wavelet transform was performed on epoched signals and subsequently two-tailed t-tests between two classes were conducted to determine statistical differences in frequency and time (t-CWT). Features were also extracted using t-CWT and subsequently classified using a support vector machine. θ and γ phase amplitude coupling (PAC) was also assessed within and between tasks. All binary classifications produced accuracies significantly greater (80-90%) than chance level, supporting the use of t-CWT in determining relative oscillatory involvements. While the perception task dynamically invoked all frequencies with more prominent θ and α activity, the covert task favoured higher frequencies with significantly higher γ activity than perception. Moreover, the perception condition produced significant θ-γ PAC, corroborating a reported linkage between syllabic and phonemic sampling. Although this coupling was found to be suppressed in the covert condition, we found significant cross-task coupling between perception θ and covert speech γ. Covert speech processing appears to be largely associated with higher frequencies of EEG. Importantly, the significant cross-task coupling between speech perception and covert speech, in the absence of within-task covert speech PAC, supports the notion that the γ- and θ-bands subserve, respectively, shared and unique encoding processes across tasks.

***Keywords*—** EEG, Neural Oscillation, Phase-Amplitude Coupling, Speech Perception, Covert Speech, Brain-computer interface

[†]Draft manuscript. Please do not cite without the authors' permission (contact jae.moon@mail.utoronto.ca)

1 Introduction

Covert speech (CS), the silent production of words in one's mind, is a fundamental trait in mental cognition (Alderson-Day and Fernyhough, 2012; Perrone-Bertolotti et al., 2014). It is referred to as a linguistic form of thought and linked to a wide range of neurocognitive functions, such as reading, writing, planning, and memory (Alderson-Day et al., 2018; Morin et al., 2011, 2018). Due to its ubiquity, many researchers in the realm of brain-computer interfaces (BCIs) have been assessing this task to restore some form of communication and control in motor-impaired individuals (DaSalla et al., 2009; Idrees and Farooq, 2016; Deng et al., 2010). However, training a CS BCI requires that individuals mentally rehearse each speech item numerous times (Rezazadeh Sereshkeh et al., 2019; Sereshkeh et al., 2017b). This repetitive and cognitively demanding task easily leads to user fatigue, a well-documented performance-degrading issue in BCI (Talukdar et al., 2019; Myrden and Chau, 2015). Fortuitously, functional magnetic resonance imaging (fMRI) studies have revealed that CS and SP activate common brain regions along the linguistic processing pathway (Okada and Hickok, 2006; Shergill et al., 2002; Skipper et al., 2005; van de Ven et al., 2009; Venezia et al., 2016), and elicit similar patterns of regional activation (Tian and Poeppel, 2010, 2012). An implication of these findings is that CS signals could potentially be modeled using SP signals and thus a CS BCI could be trained using the brain signals corresponding to the less taxing task of passive perception of speech. This type of machine transfer learning may help to overcome the user fatigue barrier in CS BCIs and thus enhance their translational potential. However, one must first understand the manifestation of neural oscillations in CS and SP tasks; these oscillations constitute an important mechanism of information transmission (Buzsáki and Draguhn, 2004; Buzsáki et al., 2004; Morillon and Schroeder, 2015; Luo and Poeppel, 2007; Ding et al., 2017) and a prominent class of signal features for BCIs (Moghimi et al., 2013).

1.1 Evidence of shared processing between speech perception and production

Numerous studies support a bi-directional linkage between perception and production systems of speech (Buchsbaum et al., 2001; Hickok and Poeppel, 2007, 2004; Poeppel, 2014; Tian and Poeppel, 2010; Okada and Hickok, 2006; Shergill et al., 2002; Skipper et al., 2005; van de Ven et al., 2009; Venezia et al., 2016). It is thought that SP initiates in the auditory regions for direct

processing of ongoing speech and ultimately maps the speech units into an articulatory network via a sensorimotor interface (Hickok, 2014; Hickok and Poeppel, 2004). Speech production, on the other hand, is thought to initiate as an articulatory motor expression, which, through the same sensorimotor interface, becomes transformed into auditory sensory targets in the temporal lobe (Hickok, 2014; Tian and Poeppel, 2012). Although the directionality between the two tasks may be opposed, they have been consistently shown to elicit similar activations in phonological networks where the fundamental contrastive speech units (phonemes) are represented (Tian and Poeppel, 2010; Hickok et al., 2011; Hickok and Poeppel, 2004; Hickok et al., 2009; Okada and Hickok, 2006; Okada et al., 2018).

These commonalities between speech production and perception invite the question of whether oscillations in CS and SP exhibit similar characteristics and serve comparable functions. In the brain, information transmission is characterized through neural oscillations in the δ-(1-2.5Hz), θ-(4-7Hz), α-(8-11Hz), β-(13-30Hz), and γ-(30-60Hz) bands (Gross et al., 2013). The most important function of these oscillations is to allow the brain to operate at multiple temporal and spatial scales such that information can be integrated into a holistic percept (Buzsáki and Draguhn, 2004; Buzsáki et al., 2004; Morillon and Schroeder, 2015). Indeed, oscillations are thought to provide the most energy-efficient physical mechanism for synchrony and temporal coordination (Mirollo and Strogatz, 1990). In the study of language processing, such oscillations (e.g. δ, θ, α, β, γ) contribute to synchronized activities across spatially segregated neuronal assemblies for the coordinated processing of speech units at varying time scales (Bastiaansen and Hagoort, 2006; Weiss and Mueller, 2003).

1.2 Neural oscillations in speech perception

In SP, it has been observed that lower frequency oscillations (e.g. θ) relate to syllabic content and speech envelope tracking (Luo and Poeppel, 2007; Di Liberto et al., 2015), whereas higher frequency (e.g. γ) parses temporally fine units of speech such as phonemes (Giraud et al., 2007; Giraud and Poeppel, 2012; Poeppel and Assaneo, 2020). In addition, the fluctuating θ phase has been found to modulate the amplitude bursts of the γ-band by imbuing a rhythmicity to the signal through phase-amplitude coupling (PAC), which enables the coordinated sampling of syllabic and phonological speech items (Hyafil et al., 2015; Assaneo and Poeppel, 2018; Hermes et al., 2014).

Active sensing, a theorized predictive processing mechanism via a motor sampling of sensory

faculties, suggests that overt speech and its variants use neural oscillations similarly to SP since acts of speech production effectively produce self-generated speech noises (Morillon and Schroeder, 2015). For instance, differential γ-band augmentations have been observed in phonological processing regions during overt and covert phoneme repetition tasks, suggesting a possible common role of γ activity to that of SP (Fukuda et al., 2010; Toyoda et al., 2014). More generally, dorsal stream motor areas have been found to exhibit locally preferred rhythms of speech production (Restle et al., 2012; Assaneo and Poeppel, 2018; Poeppel and Assaneo, 2020), suggesting that the quasi-rhythmicity of the vocal tract articulators generates the cadence of the speech envelope which, in turn, improves speech intelligibility during comprehension (Giraud et al., 2000; Boemio et al., 2005; Trouvain, 2007).

A commonly synthesized interpretation from literature is that SP invokes *multiplexed* neuronal oscillations; that is, the dynamic sampling of incoming acoustic information at multiple time scales simultaneously (Gross et al., 2013; Luo and Poeppel, 2007; Ding et al., 2017; Giraud et al., 2007; Poeppel and Assaneo, 2020). Generally, it is thought that the higher the frequency of oscillation, the finer the detail to which speech information is sampled. For instance, δ oscillations (1-2.5Hz) have been associated with the processing of words, phrases, and sentences (Giraud et al., 2007; Morillon et al., 2012; Doelling et al., 2014; Ding et al., 2015) while θ activity (4-7Hz) is critically sensitive to syllabic modulations, namely, tracking the ongoing speech envelope at the 4.5Hz syllable rate (Luo and Poeppel, 2007; Giraud et al., 2007; Ghitza, 2013; Doelling et al., 2014). In contrast, high frequency γ activity (30-60Hz) indexes processing at the phonemic level, as intracortical studies have located a 'phonotopic' map of phonemes in regions of the superior temporal gyrus producing differential γ-band augmentations to phonemes (Chang et al., 2010; Moses et al., 2016; Pasley et al., 2012). Finally, β (13-30Hz) in concert with θ and γ activities, is thought to conjoin phonological units into a broader syllabary by binding the activity of temporally segregated neuronal assemblies (Bastiaansen et al., 2010; Weiss and Mueller, 2003, 2012).

Of these oscillations, the most crucial seem to be the θ and γ bands, whose phase and amplitude are coupled in order to synchronize the detection of syllabic boundaries and the parsing of phonemes (Lizarazu et al., 2019; Mai et al., 2016; Gross et al., 2013; Morillon et al., 2012). In other words, θ activity samples the input spike trains (induced by speech waveform) to generate basic units and time references of speech for subsequent, finer-detailed processing associated with γ activity (Giraud and Poeppel, 2012). The purpose behind this phase-amplitude coupling (PAC) may be to temporally localize γ processing power to more descriptive parts of syllabic sound patterns that

constitute reference time frames (Hyafil et al., 2015). Thus, SP reputedly entails the multiplexing of relevant frequency bands in order to detect and parse incoming speech for comprehension (Gross et al., 2013).

1.3 Neural oscillations in covert speech

Although the main oscillatory contributions during SP have been fleshed out, there are fewer studies investigating the role of oscillations in CS. The main feature that distinguishes CS from SP is corollary discharge. Corollary discharge is a neural sensory prediction of the consequences of self-generated movements (Wolpert and Ghahramani, 2000; Cullen, 2004) and, in the case of speech, it is regarded as an auditory prediction of self-generated speech noises (Ford and Mathalon, 2005, 2019; Jack et al., 2019; Scott, 2013). The sequential estimation mechanism by Tian and Poeppel (2012) theorizes that the principal reason why CS produces similar activation patterns as SP in temporal regions of the brain (Tian and Poeppel, 2010) is due to the auditory prediction of imagined articulation. Indeed, this corollary discharge during CS has been shown to be temporally precise and content-specific (Jack et al., 2019), sensory in nature (Scott, 2012, 2013), and cancels out self-generated sounds (Okada et al., 2018). Therefore, CS and SP can be considered as sharing a common sensory goal in the auditory domain.

The corollary discharge during speech production has been linked to a fronto-temporal γ-band synchrony (Chen et al., 2011). Moreover, investigations into auditory verbal hallucinations (AVH) have revealed that schizophrenic individuals exhibit significantly suppressed fronto-temporal γ synchrony, suggesting that an aberrant corollary discharge is responsible for thoughts manifesting as phantom perceptions (Uhlhaas et al., 2006; Uhlhaas and Singer, 2010; Gallinat et al., 2004; Ford and Mathalon, 2005, 2019; Mathalon and Ford, 2008). More relevant to neurolinguistics, intracranial recording studies of overt and covert phoneme repetition tasks have observed differential γ band augmentations in temporal brain regions putatively responsible for phonological processing (Fukuda et al., 2010; Toyoda et al., 2014). The corollary discharge affords a prediction of the sensory consequences of covert speech (Scott, 2013). As such, the corollary discharge produced during CS, as reflected in its γ-band response, may be phonological in nature, similar to that of SP.

Although these studies provide indirect evidence for a common γ-band function across tasks, the same may not be the case for θ. Hermes et al. (2014) showed that θ-γ PAC is suppressed during CS, with θ power being anti-correlated to high frequency activity. In contrast to findings of increased

PAC in SP studies, the authors surmised that when there is no external input, brain areas may need to downregulate θ activity in order to allow local neuronal processing (Schroeder and Lakatos, 2009). Therefore, it is possible that the θ-band in CS may play an alternative role to syllabic chunking seen in SP.

1.4 Oscillatory characteristics: speech perception vs. covert speech

While there is ample evidence supporting that SP and CS share functional networks and invoke activity in common brain structures, there seems to be a paucity of direct investigations comparing their respective signal characteristics. Thus, the present study assessed neurophysiological signals associated with SP and CS, with a focus on identifying shared signal characteristics and those that are task-specific. First, continuous wavelet transform was performed on epoched signals and subsequently two-tailed t-tests between two classes were conducted to determine statistical differences in frequency and time (`t`-CWT). To substantiate the validity of these time-frequency differences, features were also extracted using `t`-CWT and subsequently classified using a support vector machine. Furthermore, we assessed intra- and cross-task PAC to understand how lower and higher frequencies are coordinated. Considering the literature reviewed above, and that CS lacks overt vocalization and thus salient self-stimulation, we hypothesized that high frequency activity, such as β and γ, would uniquely dominate in CS and that the coupling of γ- and θ-band activities would differ between SP and CS.

2 Results

2.1 Task-dependent oscillations

Binary SVM classification accuracies, precision, recall, and F1-scores exceeded chance (Fig. 1) in all participants. The distribution of significant time-frequency differences across frequency bands (cumulative over all 62 channels and participants) revealed task-related oscillatory characteristics (Fig. 2). Consistent with previous studies (Gross et al., 2013; Luo and Poeppel, 2007), classification problems involving SP invoked multiple frequency bands while those involving CS, evoked activity predominantly in the higher frequencies like β and γ. Distinguishing between SP classes (i.e. listening to articulations of 'blue' or 'orange') preferentially engaged lower frequency θ and α ($p<$

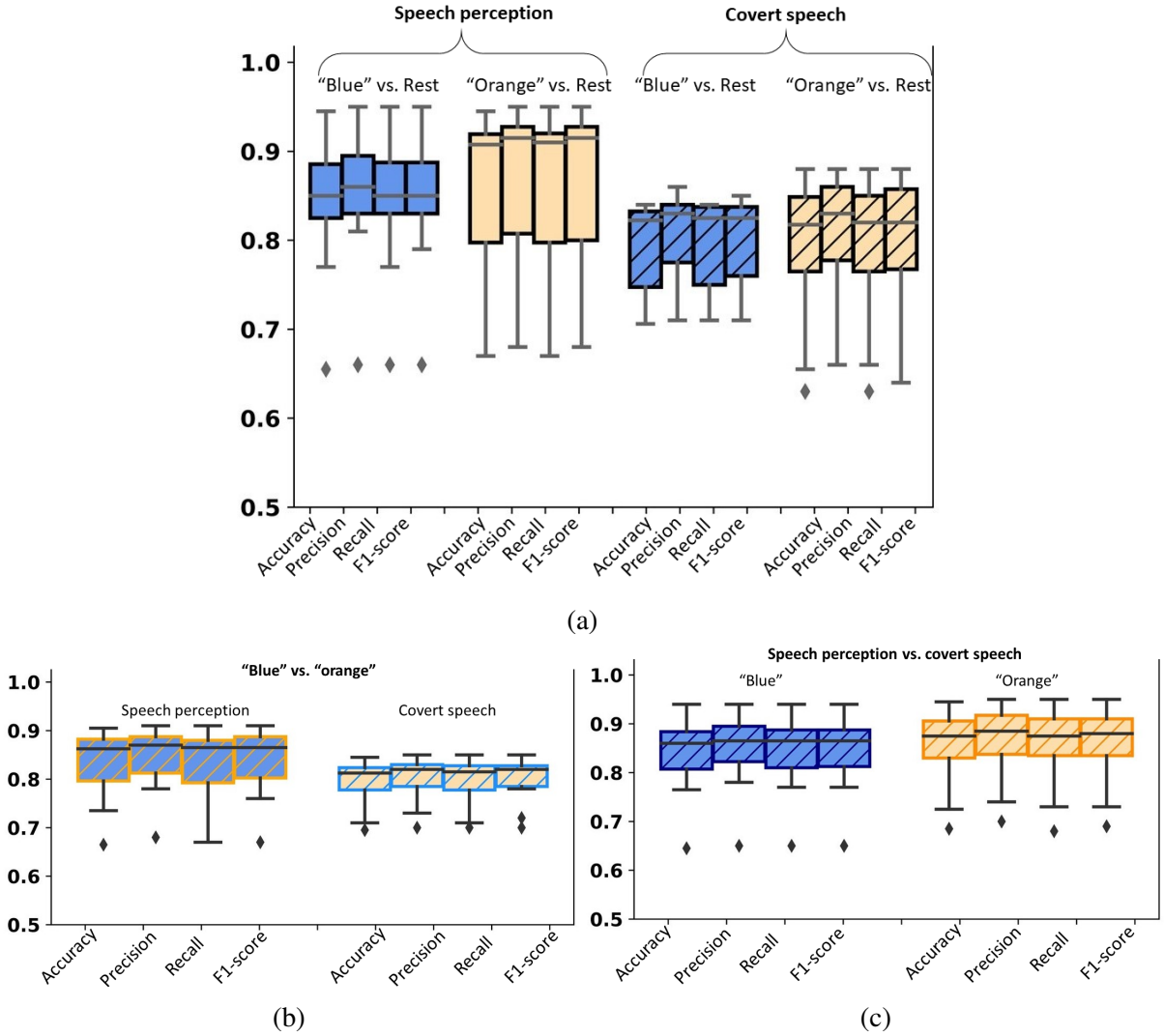


Figure 1: **Across-participant average performance for all binary classification problems.**2-column figure Top panel (a) shows the binary classification performances between rest and either speech perception or covert speech. (b) and (c) show the performance between words (blue vs. orange) and task type (speech perception vs. covert speech), respectively.

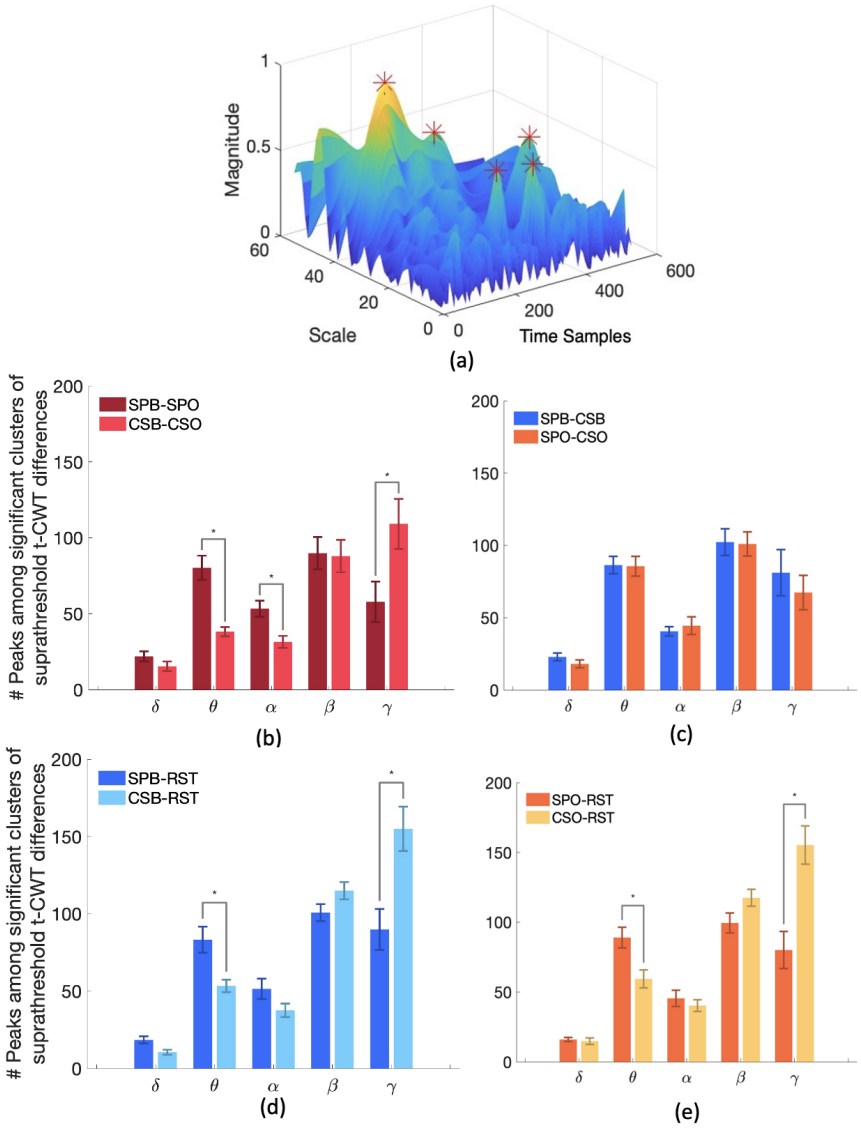


Figure 2: **Differential frequency characteristics in covert speech (CS) and speech perception (SP).** 2-column figure Average and standard deviation of counts of peaks among significant clusters across participants, per frequency bin are shown. (a) depicts the time-scale location of peaks for participant 1, for the comparison of SP Blue and SP Orange. 'B' and 'O' in SPB and SPO denote SP Blue and Orange, respectively. The same notation is used for CS. The vertical axis represents the number of differences detected through t-CWT while the horizontal axis denotes the frequency bands. b) t-CWT differences when comparing SP and CS of different words. c) t-CWT differences between SP and CS of the same word. d, e) t-CWT differences when comparing word and rest in SP and CS conditions. (*=p<0.05 in at least 6 out of 10 participants).

0.05), whereas distinguishing CS (e.g., covert rehearsal of 'blue' or 'orange') involved more γ (p<0.05) (Fig. 2b).

No significant differences in t-CWT in the δ and β bands were observed (p>0.05). t-CWT differences between perception and covert rehearsal of a given speech token (i.e. SPB vs CSB, SPO vs CSO) were consistent across words (p>0.05) (Fig. 2c). When distinguishing task from rest, the θ band dominated in SP (SPB vs RST, p<0.05; SPO vs RST, p<0.05), while in CS, γ was the key differentiator (p<0.05) (Fig.2d, e).

2.2 Task-dependent coupling of θ phase to γ amplitude

Inter-trial coherence (ITC) revealed strong transient synchronization in the θ band for CS and SP between 0-600ms (Fig. 3a, b). Rest activity exhibited weak and sparse θ-band synchrony. ITC was not correlated to θ power. In channel TP8, where significant θ phase coherence was observed, the PAC of both SP classes significantly departed from the surrogate PAC between 100-500ms (p<0.01), confirming that SP γ amplitude produces a rhythm in keeping with the cadence of fluctuating θ phase (Fig. 4) (Giraud and Poeppel, 2012; Hyafil et al., 2015). CS and rest were found to suppress or lack such a stable PAC, producing sparse distributions of significant departures from surrogates.

However, it was possible that CS θ activity may have served a separate encoding process unlike that putatively observed in SP (Restle et al., 2012; Albouy et al., 2017), especially in the absence of salient stimulation. Therefore, the PAC between SP's θ and CS's γ was assessed to determine whether the γ amplitude of CS contains a task-specific rhythm (Fig. 4a, b). Significant cross-task PAC (p<0.01) occurred again between 100-500ms for both words, confirming that CS's γ-band produced a rhythmic fluctuation specific to the time course of SP's θ synchronization (Fig. 3a). No significant PAC was observed between SP and rest tasks. Furthermore, relatively lower but nevertheless significant cross-task couplings were observed across words (e.g SPB-CSO). This lack of specificity between SP θ and CS γ was found to be due to significantly correlated θ phase patterns in SP of the two words between 200-400ms (circular correlation test, p<0.05) (Fig. 5). Finally, comparing the θ phase across tasks revealed a distinct lack of correlation.

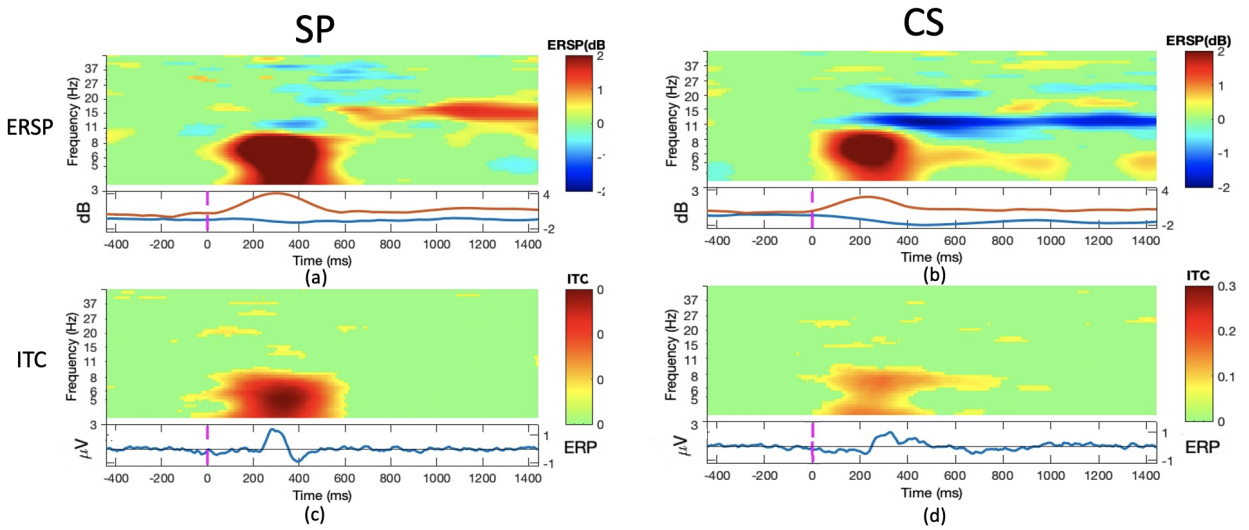


Figure 3: **Event related spectral perturbation (ERSP; top row) and inter-trial coherence (ITC; bottom row) in SP (first column) and CS (2nd column) in channel TP8.** 1-column figure Figures were generated after pooling data from all participants. The red and blue lines under ERSP time-frequency plots (a) and (b) depict the upper and lower envelopes respectively of the corresponding ERSP. The line graphs under the ITC time-frequency plots (c) and (d) are the corresponding grand averaged (over all participants) event-related potentials (ERPs). The ITC for SP (c) and CS (d) both show early latency synchronization in the θ band between 0-600ms. Interestingly, SP was associated with enhanced beta band power from 600-1400 ms, as seen in (a) following theta synchronization, shown in (c), while the opposite was observed in CS, as indicated by the blue strip in (b). No significant phase synchronization was observed in the rest task (not shown).

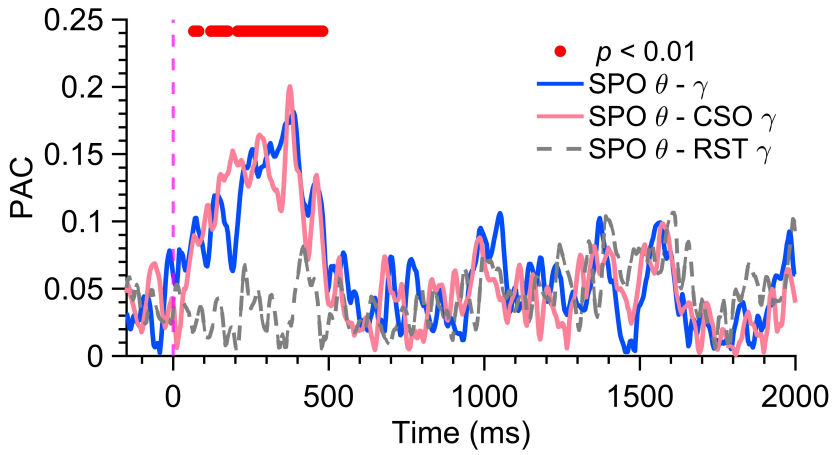


Figure 4: **Speech perception shows within-task coupling and its θ phase is correlated to covert speech low γ amplitude.** 1-column figure SPO and CSO refer to SP Orange and CS Orange, respectively. PACs were calculated by pooling trials from all participants. PAC calculations on the participant level yielded greater values in significant regions. SPO produced significant PAC relationships between 100-500ms (blue trace). SPB also exhibited significant within-task coupling within the same time frame (data not shown). Furthermore, the figure shows that SP's θ phase produced significant cross-task coupling to CS's low γ amplitude compared to surrogate PACs between 100-500ms, with a similar coupling morphology to SP PAC (pink trace). SPB-CSB exhibited the same cross-task coupling. Finally, there was no coupling between SP θ and rest γ (great dotted trace). Red dotted lines indicate significance difference (p <0.01) against surrogate population.

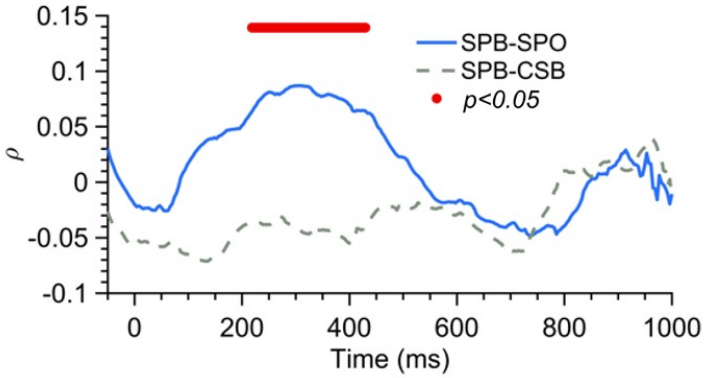


Figure 5: θ **activity is task dependent and may subserve different encoding processes in speech perception and covert speech.** 1-column figure 'B' and 'O' in SPB and SPO for instance denote the words, Blue and Orange, respectively. Circular correlation tests were conducted after pooling data from all participants. Significant (p<0.05) θ phase correlations were observed between 200-400ms for SPB-SPO, revealing that the SP of the two words did not lead to divergent θ patterns. Importantly, no phase correlations were observed across tasks (SPB-CSB, grey; SPO-CSO, not shown).

3 Discussion

Our findings support the hypothesis that CS largely utilizes higher frequency oscillations relative to SP. Crucially, we observed that the γ-bands of CS and SP have common rhythms coupled to the same SP θ phase, suggesting that they subserve similar encoding processes. Like SP, the γ activity of CS was found to contain a processing rhythm time-locked to the cadence of event-related SP θ phase. However, the lack of within-task coupling within CS suggests that CS's θ activity likely serves an alternative encoding process to syllabic chunking seen in SP, possibly in preparatory motor or memory-related activity (Restle et al., 2012; Albouy et al., 2017). The findings reported here lend credence to the notion of deriving CS classification models on the basis of SP signals. In other words, our observations suggest that a CS BCI could be trained on signals from the passive perception of speech, although the effectiveness of this transfer learning would have to be investigated in future research.

3.1 Oscillations in speech perception and covert speech

The main goal of the study was to assess the relative contribution of neural oscillations in CS relative to SP. Thus, the oscillatory references of SP will be discussed first. A commonly synthesized interpretation from studies on SP investigating the role of oscillations is that each frequency band contributes to a dynamic sampling of speech items at varying temporal scales, referred to as multiplexing (Gross et al., 2013). The greater the frequency, the greater the resolution and detail at which a speech item is sampled. In the present study, we observed that SP indeed utilizes the θ, α, β, and γ frequency bands, with significantly more prominent differences in the θ- and α bands than CS. The prominence of the α-band may be due to its role in auditory gating (Sadaghiani et al., 2012), which is outside of the scope of this paper . However, the observed θ prominence, in part, can be attributed to the importance of tracking the speech envelope of salient percepts (Luo and Poeppel, 2007; Giraud and Poeppel, 2012) and syllabic chunking (Ghitza, 2012, 2013; Doelling et al., 2014), features which seem to be common across languages (Ding et al., 2017; Varnet et al., 2017).

The comparably high count of γ-band differences suggests that the one of the driving forces of delineating words in SP lies in phonemic processes (Chang et al., 2010). Moreover, the significant θ-γ PAC between 100-500ms likely depicts a linkage between syllabic and phonemic processes (Giraud and Poeppel, 2012), suggesting that the two oscillations may be inherently coordinated to focus γ activity to specific times within syllabic time references upheld by θ (Hyafil et al., 2015). This PAC was likely related to high frequency formant transitions, which tend to be nested within the lower frequency envelope (Schroeder et al., 2008). The timing of the observed PAC was likely linked to strong θ-band inter-trial phase coherence occurring in this time period, as observed in Figure 3.

CS, on the other hand, did not exhibit a diverse oscillatory relationship akin to SP, but rather favoured high frequency activity, namely β and γ, for the distinction of words. The relatively low count of θ differences compared to SP was likely due to lack of salient percepts in CS, suggesting that, unlike SP, θ activity in CS is focal and may serve a different encoding process (discussed in Section 5.2). On the other hand, the high amount of differences in the γ-band suggested that this frequency is highly specific to word identity. Indeed, studies of overt and covert phoneme/word repetition tasks have shown differential γ-band augmentations in the temporal and temporo-parietal lobe (Fukuda et al., 2010; Pei et al., 2011; Toyoda et al., 2014).

It is possible that the γ-band activity during CS corresponded to a corollary discharge, as

enhanced fronto-temporal γ synchrony has been observed during speech production tasks relative to perception conditions (Chen et al., 2011; Ford and Mathalon, 2005). Furthermore, in investigations of auditory verbal hallucinations (AVH), fronto-temporal γ synchrony was found to be significantly suppressed in schizophrenic individuals (Uhlhaas et al., 2006; Uhlhaas and Singer, 2010; Gallinat et al., 2004), suggesting an improper transmission of corollary discharge leads to phantom perceptions (Mathalon and Ford, 2008; van Lutterveld et al., 2011). Although such inter-regional synchrony was not investigated here, the fact that a suppression of this auditory prediction produces phantom perceptions of internal thoughts suggests that the pattern of activity in corollary discharge reflects the potentials during SP of the same words. This invites the hypothesis that CS's γ activity serves a 'mirrored phonological' function to SP's γ activity.

3.2 θ and γ encoding in speech perception and covert speech

It was observed that CS and SP engage their oscillations differentially during speech processing. Specifically, CS produced the greatest number of significant t-CWT differences in the γ-band. As previously mentioned, a multitude of studies describe a phonological processing function of γ activity during SP (Chang et al., 2010; Pasley et al., 2012). Although the methods of the present study were not sensitive to determining phonological cognitive load, we speculated that CS's γ activity contained a processing rhythm specific to θ activity. SP's γ activity has been shown to maintain a rhythm with respect to the rise and fall (periodicity) of its θ phase, which, when coupled, putatively allows individual phonemes to be processed in the context of larger syllabic units (Giraud and Poeppel, 2012; Hyafil et al., 2015). The existence of such a rhythm specific to a time frame would lend support to the hypothesis that γ activity in CS serves a similar encoding process to that in SP. Therefore, understanding this γ rhythm would facilitate modeling CS from SP.

However, within-task θ-γ PAC was not observed within the CS condition. This was likely due to θ activity in speech production (and its variants) being responsible for non-linguistic portions of the task such as motor (Restle et al., 2012) and memory-related processes (Albouy et al., 2017). Indeed, θ phase correlations were not observed between SP and CS (Fig. 5). Thus, we asked if CS γ would produce a rhythm that corresponds to the event-related θ phase of SP. Indeed, the results of the current study support this hypothesis with significant SP θ-CS γ PAC also occurring at 100-500ms. This coupling was found to be temporally sensitive and specific to this particular time period, as an otherwise random relationship would portend a sparsely distributed coupling pattern. This finding

indicates that the two tasks produce similar γ rhythmic bursts collectively modulated by the putative tracking of syllabic quantities by SP θ (Luo and Poeppel, 2007). Therefore, the observed cross-task coupling supports the idea that the γ-bands of SP and CS served a common encoding process.

In systems exhibiting coupling, the higher frequencies are often modulated according to a preferred phase of lower frequency oscillations (Thatcher et al., 2008; Onslow et al., 2011). In this manner, a circularly varying phase value periodically modulates the bursting of high frequency amplitudes, thereby influencing neural coding (Giraud and Poeppel, 2012). Therefore, the coupling of CS γ to SP θ is likely explained by an aligned preference of θ phase. Furthermore, combining this result together with evidence from literature would suggest that SP and CS produce similar γ activation patterns. Thus, it may be possible to predict the CS γ bursts of novel speech items from the corresponding SP θ phase. Furthermore, no significant differences were observed in the `t`-CWT analyses between SPB-CSB and SPO-CSO (Fig. 2), which may point to differences between CS and SP that are independent of the words being modeled.

It should be noted that significant cross-task couplings were observed across tasks between dissimilar words (e.g. SPB-CSO). This does not necessarily portend that the γ activities themselves are general, but rather that the current corpus does not include a large variety of θ phase patterns. Indeed, circular correlation tests revealed that θ phases of the SP signals of the two words were significantly correlated (Fig. 5). This finding suggests that the current speech items produced non-divergent theta patterns leading to cross-word couplings likely caused by a common preferred phase of γ. Therefore, it is suggested that future studies investigate PAC and preferred phase in a large variety of speech tokens larger differences in linguistic properties such as number of syllables and phonemes. With such a study, a more comprehensive model of phase and amplitude may be constructed via clustering methods, so as to enable cross-condition classification.

The temporally co-localized γ rhythms of SP and CS both seemed to correspond to transient θ-band synchronizations in the same time period, likely caused by a phase resetting prior to synchronization. In SP, θ phase has been found to reset to the temporal edges of the speech envelope (Gross et al., 2013) in order to initiate the coordination of processing at syllabic and phonemic levels (Assaneo and Poeppel, 2018). Similarly, θ phase has been found to reset to the onset of CS, resulting in strong phase-locking between 250-500ms that represents a temporal marker of CS processing (Yao et al., 2020). Furthermore, supplementary analyses in this study showed that the observed θ-band phase coherence was not significantly related to evoked oscillatory power, suggesting that the observed ITC and PAC were not simply due to task demands or protocol-induced evoked potentials.

From a neural architectural perspective (i.e. neural circuits), such phase-resetting has been proposed to underlie information transmission such as communication through coherence (Roberts et al., 2013) and the phase-dependent coordination of large scale neural networks for encoding and decoding during attention and goal-directed behaviours (Canavier, 2015; Voloh and Womelsdorf, 2016). It is thought that phase alignment through resetting forms predictable windows for integration which aids the coordinated parsing of segments (Fries, 2009). Hence, the transient θ-band synchronizations observed here may have demarcated the points of processing in the tasks which were common for both SP and CS.

If there is significant θ synchronizations in CS but no within-task coupling, what is its role? Considering that stimulation of a major dorsal stream area (posterior inferior frontal gyrus) with θ burst stimulation facilitates speech repetition accuracy (Restle et al., 2012), it may be reasoned that θ activity during CS correlates to motor planning and activity. Indeed, 4-7Hz also corresponds to the mandibular movement rate during articulation (Giraud et al., 2007), which is also demonstrated by an enhanced coupling between motor and auditory areas during syllable presentation at 4.5Hz (Assaneo and Poeppel, 2018; Poeppel and Assaneo, 2020). These studies indicate that the θ-band represents a preferred articulatory rhythm and thus has motor origins in speech production. If this interpretation holds, it would suggest that relationships between phonological and articulatory processing in CS do not manifest in PAC, but potentially some other measure. For instance, it may be possible that θ-based articulatory expressions induces γ-based corollary discharge that contains the sensory predictions outlined by the motor code. Indeed, this is the main prediction put forth by the active sensing framework, whereby low frequency oscillations arising in motor areas provide the temporal schema in which to process ongoing speech (Morillon and Schroeder, 2015; Morillon et al., 2019).

To further study the roles of γ and θ in CS and SP, future studies ought to investigate their phase coding, a neural framework in which information is coded through the phase of neural oscillations (Sauseng et al., 2010). Measurements in this domain include inter-regional synchrony such as the weighted phase lag index (Assaneo and Poeppel, 2018) as well as inter-trial coherence and inter-regional PAC. If the present interpretations hold, γ activity should be localized to temporal channels and its synchrony would reveal a motor/somatosensory to auditory synchronization indicating corollary discharge. Furthermore, if θ activity is indeed motoric in origin, its activity and synchrony should be targeted to frontal sensors.

4 Conclusion

The present study investigated oscillatory engagement during CS and SP. We found that CS favours higher frequency activity such as β and γ. Crucially, we found that CS's γ-band response has a similar rhythmic pattern to that of SP by coupling to SP θ, which suggests that this high frequency oscillation performs similar encoding processes across tasks. Contrarily, we suspect that θ activity in CS subserves different encoding processes to SP possibly in motor-related processes under the active sensing framework (Morillon and Schroeder, 2015; Morillon et al., 2019).

Understanding the relative oscillatory engagements and their encoding processes in the two tasks are elemental to the modelling of CS based on SP signals. Therefore, the present work may eventually lead to the development of CS BCIs through the passive perception of speech. In order to achieve this modelling, we direct future studies to investigate the relationship between SP's θ activity and CS's γ activity on a large variety of words.

5 Methods

5.1 Participants

Ten adults between the ages of 20 and 40 without disabilities or known health conditions were recruited for this study (8 female, aged 29.3 +/-7.0 years; 2 male, aged 25.3 +/-2.8 years). All participants were right-handed to ensure consistency in the hemispheric dominance of neurolinguistic processing. Furthermore, all participants were native English speakers (i.e. first language). The research ethics boards of Holland Bloorview and the University of Toronto approved this study. Participants provided informed written consent.

5.2 Instrumentation

Participants wore a 128 electrode ActiCap EEG cap (Brain Products GmbH). Of these 128 channels, 64 were utilized (Fig. 6), with the ground electrode at AFz and the reference electrode at FCz. Channels Fp1 and Fp2 were used as ocular artifact detectors. Electrode coverage included the frontal, temporal, and temporo-parietal areas on both hemispheres, including midline components such as Fz, Cz, CPz, and Pz). Data were sampled at 1000 Hz and collected through BrainVision

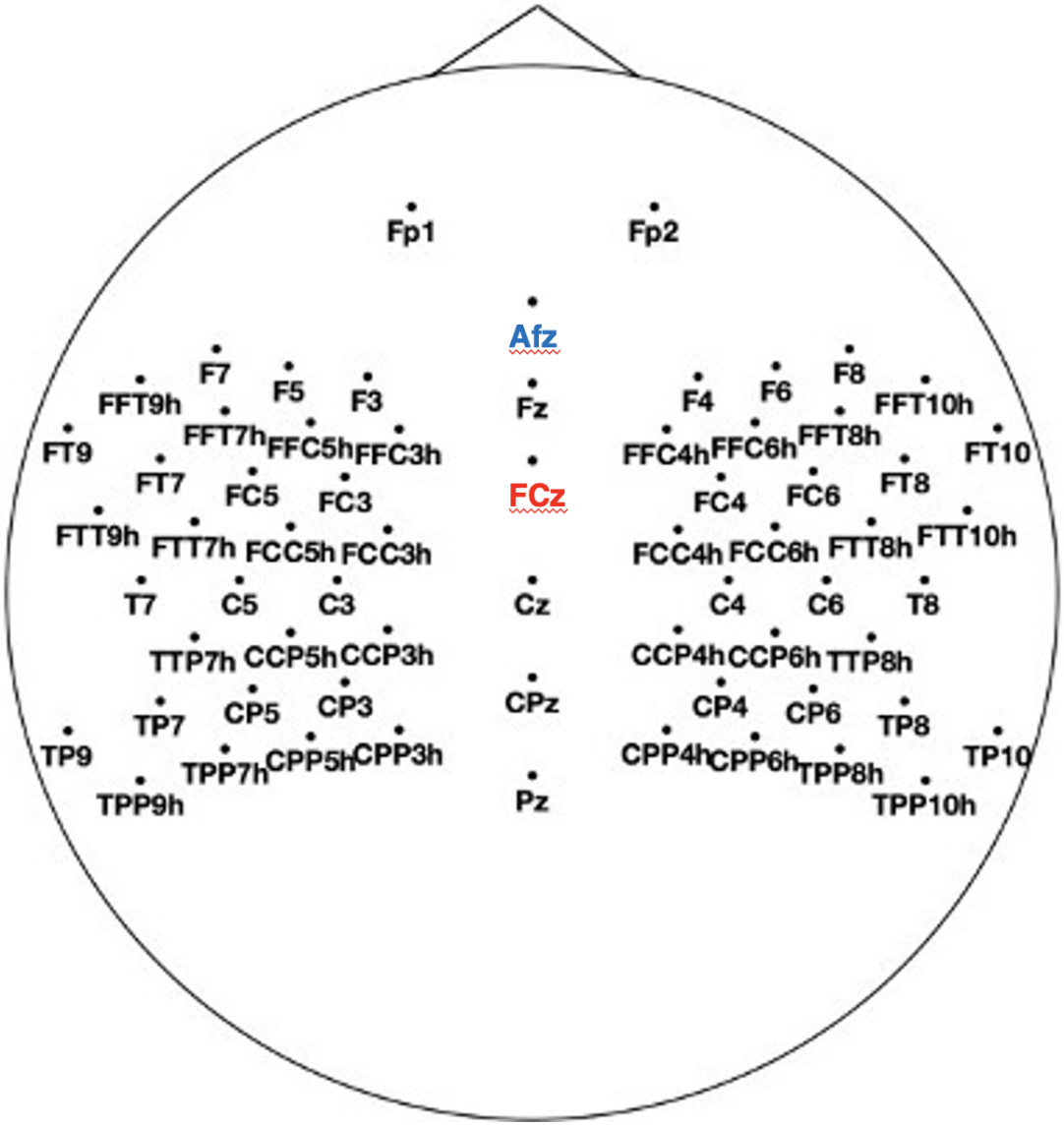


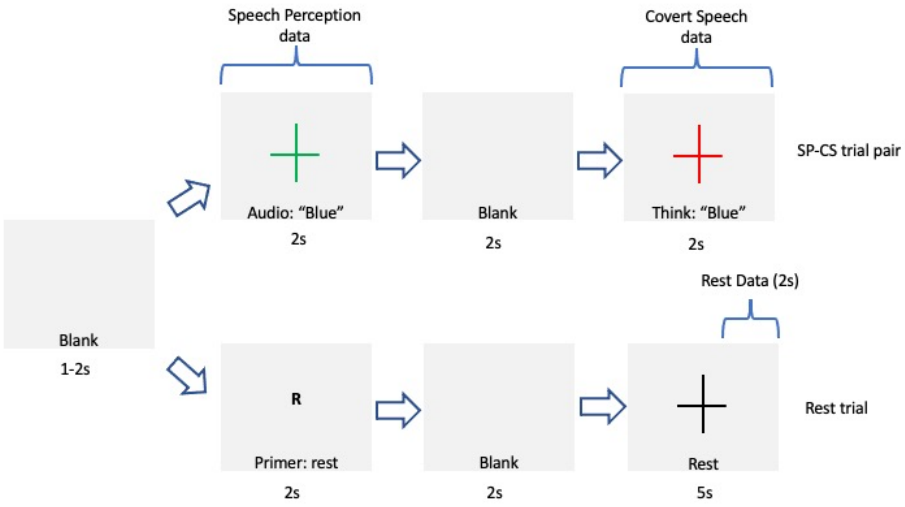
Figure 6: **64 channel ActiCap wet EEG montage.**2 column if 1 column leads to small text

Recorder (Brain Products GmbH).

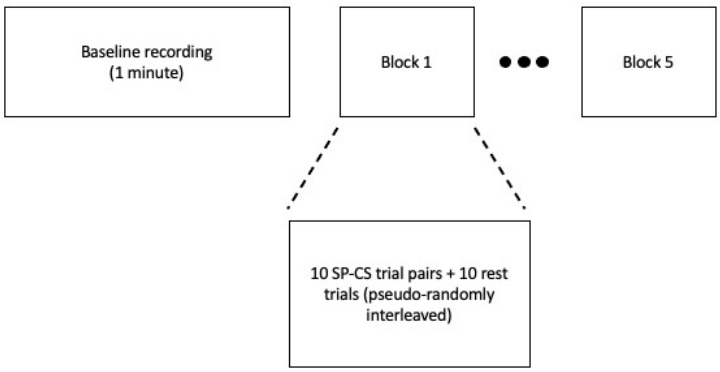
As we were primarily interested in studying the time-evolving characteristics of brain oscillations, EEG was invoked as the measurement modality given its millisecond temporal resolution. In speech processing studies, EEG has been frequently used to characterize the temporal dynamics of oscillations in phase entrainment (Zoefel and VanRullen, 2016), processing asymmetry (Morillon et al., 2012), speech intelligibility (Onojima et al., 2017), semantic evaluation of speech (Shahin et al., 2009), categorical processing (Bidelman, 2015), and oscillatory abnormalities in schizophrenic individuals (Uhlhaas and Singer, 2010; Ford and Mathalon, 2005).

5.3 Experimental procedure

Each participant was seated comfortably approximately 50 cm from a LCD computer screen with a refresh rate of 75 Hz. The screen was positioned in the central field of vision and light in the data collection room was turned off prior to beginning the computer task to minimize peripheral vision distractions. Prior to the experiment, prompted by a constant green cross, baseline signals (i.e. neural activity at rest) were recorded for a minute. During the session, SP and CS trial pairs were presented sequentially (Fig. 7). First, a blank screen with a durational jitter between 1-2 seconds was presented. Next, a green cross appeared for 2 seconds, during which the computer articulated the speech token ('Blue' or 'Orange') once in a synthesized female voice. Subsequently, the blank screen reappeared for 2 seconds, during which the participant remained idle. Finally, a red cross appeared for 2 seconds, during which participants covertly rehearsed the speech token that they had just heard. Therefore, every SP trial was followed by a CS trial with the same word. Participants were instructed prior to the start of the experiment that the green and red crosses signify perception and covert speech, respectively. For the rest trial, the letter R cued the participants to refrain from the speech task and fixate on a black cross on the screen for 5 seconds. The additional time for the rest task was intended to give participants a break. Each block consisted of 10 SP-CS trial pairs and 10 rest trials. Each session consisted of 5 blocks, yielding 50 trials/class. Participants completed two sessions each at least 2 days apart and at roughly the same time of day. Each session was no more than 50 minutes.



(a)



(b)

Figure 7: **The trial (a) and sessional (b) protocols.** 2-column figure In the top branch of (a), speech perception (SP) and rest were preceded with a jitter of 1-2 seconds and followed by computer articulation of "blue" or "orange" (speech perception task) or a visual presentation of the letter "R" (rest task). Speech perception was followed by a 2-second blank screen and a 2-second window for covert speech (CS), demarcated by a red cross. In the bottom branch of (a), a 2-second blank screen was followed by a black cross signaling 5 seconds of rest. The last 2-seconds of the rest data were subsequently used for analyses. As depicted in (b), the session consisted of a 1-minute baseline recording and five blocks each consisting of 10 SP-CS trial pairs pseudo-randomly interleaved with rest trials.

5.4 Speech items

The two speech items ('Blue', 'Orange') were chosen as they differ in the number of syllables and phonemes and differ in their place and manner of articulation. As such, they were expected to engender substantially different, machine-discernible patterns of neural activity. The rationale was that these variations would encourage the activation of different motor, somatosensory, and auditory neural representations, thereby enhancing signal discriminability. However, these words were deliberately selected from the same semantic category.

Speech stimuli were generated by Google Cloud Text-to-Speech platform (female voice) and presented at an approximate rate of 150 words per minute, which is within the range of the natural speech (Giraud et al., 2007; Luo and Poeppel, 2007). Phoneme models were generated through the Montreal Forced Aligner (McAuliffe et al., 2017). Total word length was between 600-700 ms.

5.5 Preprocessing

Raw data were analyzed in EEGLAB (Delorme and Makeig, 2004). A 4^{th} order zero-phase high pass Butterworth filter with a cutoff frequency of 1Hz was applied to remove baseline drift. Subsequently, a 4^{th} order zero-phase low pass Butterworth filter with a cutoff frequency of 60Hz was applied to remove high frequency noise, as well as the high-γ band. Subsequently, the PREP pipeline (Bigdely-shamlo et al., 2015) was applied to remove line noise, detect noisy or outlier channels, and to interpolate bad channels. Ocular and myogenic artifacts were removed in two separate steps using blind source separation (EEGLAB) and the SOBI algorithm to extract optimal spatial filters (Gomez-Herrero, 2007). To minimize volume conduction effects, we transformed the measured potentials using a surface Laplacian estimated from a smoothing spherical spline function (Babiloni et al., 2001; Kayser and Tenke, 2006, 2015). This transformation emphasizes local characteristics in the topography, while attenuating broadly distributed spatial effects (Kayser and Tenke, 2006, 2015).

After performing the above steps, signals were downsampled to 256Hz prior to epoching. Data from sessions 1 and 2 were combined for offline analyses. There were a total of 5 classes: SP Blue (SPB), SP Orange (SPO), CS Blue (CSB), CS Orange (CSO), and rest (RST). All SP and CS trials were epoched between 0 to 2 seconds. The mean of the corresponding 1 second pre-stimulus baseline was subtracted from each epoch. Rest trials were initially epoched between 0 to 5 seconds. After removing the mean of the 1 second baseline prior to the presentation of the black fixation cross, rest data were selected between 3 to 5 seconds, as the presentation of the visual stimulus (i.e.

black cross) may have elicited event-related potentials (Huang et al., 2017; Guérin-Dugué et al., 2018). For a similar approach to rest epoching, see Lotte and Jeunet, 2018. The preprocessing pipeline is depicted in Figure 8.

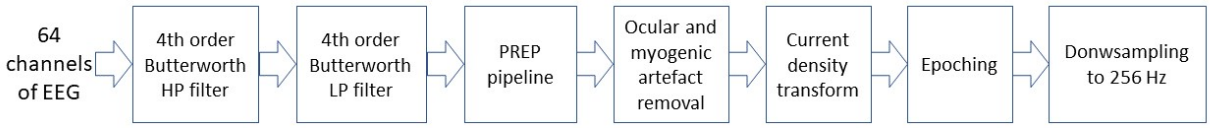


Figure 8: **Preprocessing pipeline.**2 column

5.6 Overall analytical approach

To hone in on shared and task-specific oscillatory characteristics, we pursued a machine learning based analytical approach, whereby a series of classifiers were developed for the binary problems depicted in Figure 9, facilitating the automatic identification of discriminatory signal features between tasks. The following sections describe the key computational steps of this approach, i.e., feature extraction, selection and classification. Similar machine learning-based approaches have facilitated the study of the role of gamma and theta frequencies in working memory in patients with schizophrenia (Johannesen et al., 2016), the left-lateralized increase in alpha power in depression (Hosseinifard et al., 2013), and resting state spectral, biocoherence and complexity EEG characteristics in children with ADHD (Chen et al., 2019).

5.7 Feature extraction: Localizing differences between signals in the time-frequency plane

To investigate the contributions of different oscillations to CS and SP, we first localized time-frequency differences between task-specific signals using a studentized continuous wavelet transform (t-CWT; Bostanov, 2004). The t-CWT method has been previously deployed in the classification of motor imagery (Darvishi and Al-Ani, 2007; Hsu et al., 2007) and extraction of ERPs in real and simulated EEG data (Real and Kotchoubey, 2014). Kimata et al. (2018) contends that the CWT produces pseudo-frequency estimates with greater temporal stability than either the discrete wavelet or Fourier transforms.

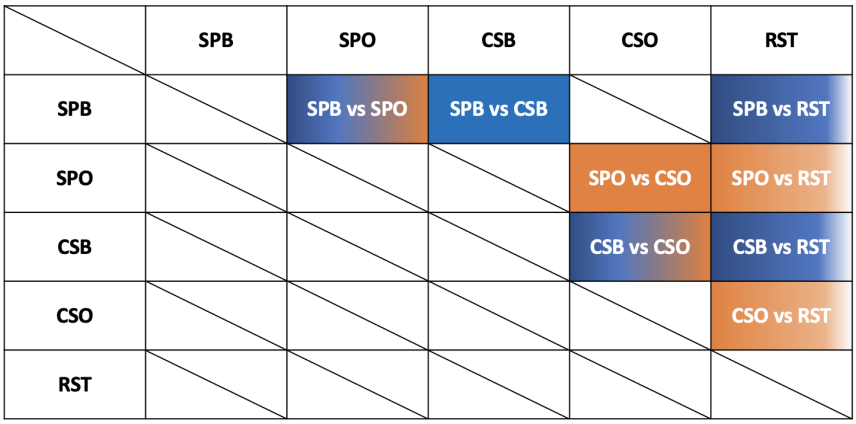


Figure 9: **Binary classification problems.** The binary comparisons are indicated by colored boxes. SPB - speech perception blue; SPO - speech perception orange; CSB - covert speech blue; CSO - covert speech orange.

First, each epoched trial was zero-padded with 12 samples each in the beginning and end to limit edge effects. A continuous wavelet transform was computed for each epoched signal over a linear sequence of scales ranging from 1 to 55, yielding a matrix of coefficients, $W(s,t)$, where $s = 1, 2...55$, $t = 1, 2..536$ denote scale and time samples, respectively (Wavelet Toolbox, MATLAB, MathWorks) (Lilly and Olhede, 2009; Lilly, 2017). Consistent with the choice of wavelet in other EEG oscillation studies, a Morlet mother wavelet (center frequency = 0.8125Hz, spectral bandwidth = 0.6094Hz) was selected (Ende et al., 1998; Senkowski and Herrmann, 2002). Compared to Bump, Haar, and Morse alternatives, the Morlet wavelet maximized the cross-correlation between original and reconstructed signals after discarding coefficients representing frequencies in excess of 60Hz.

Suppose we would like to compare oscillatory contributions of the epoched signals from two different task conditions, A and B, where A vs. B can be any of the binary problems in Figure 9. For each channel, $k = 1, \ldots, 62$ (excluding frontal eye channels), for each scale $s = 1, \ldots, 55$ and time $t = 1, \ldots, 536$, a two-sample t-test was computed between the wavelet coefficients across trials between conditions. Thus, a 55×536 matrix of sample t-values was generated for each channel between tasks A and B. Namely, for channel, $k = 1, \ldots, 62$, we obtained a matrix whose elements were specified by:

$$\mathrm{t}^{k}_{\Delta AB}(s,t) = [\overline{W^{k}_{A}(s,t)} - \overline{W^{k}_{B}(s,t)}] \Big/ \sqrt{\sigma^{2}_{A}/N_{A} + \sigma^{2}_{B}/N_{b}} \qquad (1)$$

where $\overline{W_A^k(s,t)}$, $\overline{W_B^k(s,t)}$, $\sigma_A^2(s,t)$, and $\sigma_B^2(s,t)$ denote sample means and standard deviations of wavelet coefficients across trials at each scale s and time t for classes A and B, and where number of examples in each class are equal, i.e., $N_A = N_B$. The absolute value t-statistic matrices served as time-frequency scalograms with greater magnitudes denoting greater differences.

To compare t-statistic matrices between classes while mitigating false discoveries due to multiple comparisons, cluster-based permutation tests (Groppe et al., 2011; Real and Kotchoubey, 2014) were conducted. The null hypothesis was that the signals were sampled from the same population. A cluster was defined minimally as comprising 4-connected components, i.e., 4 adjacent suprathreshold t-statistics in the time-scale plane. The choice of a 4-connected neighborhood maximized the detection of potentially significant regions of the time-frequency plane while avoiding the selection of edge regions. The threshold t-value for cluster determination corresponded to that for $p < 0.05$. Cluster-level statistics were obtained by summing the within-cluster t-values. The maximum cluster-level statistic across all clusters was identified. The probability of observing a t-value this extreme under the null hypothesis was estimated via Monte Carlo simulation. Specifically, for each time-frequency comparison, 1000 permutations of class labels were generated under the null hypothesis. We confirmed that the estimated p-values corresponding to the observed cluster-level statistics stabilized (i.e., changed by less than 2×10^{-3}) from 1000 permutations onward. If $p < 0.05$ for observing the cluster-level statistic under question, the corresponding cluster was considered significant. The corresponding wavelet coefficients for the time-frequency points within the significant cluster constituted candidate features for differentiating between the 2 classes of signals in question. The above procedure was repeated for the second, third, and up to the tenth highest cluster-level statistic. Figure 10 graphically summarizes the feature extraction process.

5.8 Feature selection and classification

For each binary classification task (i.e. SPB vs SPO, CSB vs CSO, SPB vs CSB, SPO vs CSO, SPB vs RST, SPO vs RST, CSB vs RST, CSO vs RST), the features were a subset of the wavelet coefficients obtained by the procedure described above. The real and imaginary parts of the complex wavelet coefficients for each of the 62 channels were appended to form a matrix of trials × features. For each participant, the original feature matrix was 200 trials x ~500 features. The number of features varied across participants due to variations in the number and topology of the clusters. Twenty (20) features were selected via the Minimally Redundant Maximally Relevant (mRmR)

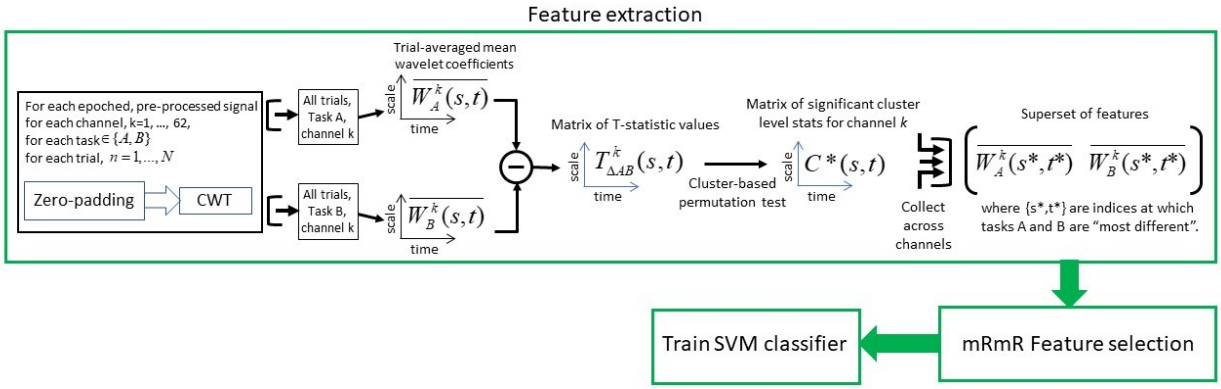


Figure 10: **Computational steps within a cross-validation training fold.** 1 column figure The continuous wavelet transform was invoked on zero-padded signals and cluster-based permutation tests were conducted for each channel, for trials from two classes (denoted A and B) to extract wavelet coefficients with the most significant differences between classes.

algorithm (Ding and Peng, 2005). Subsequently, a 10-fold cross validation was conducted on the 200 trials x 20 features matrices and subsequently classified through a support vector machine (SVM) with a radial basis function kernel (Hastie et al., 2009). Average classification accuracies, their standard deviations, precision, recall and F1-score were obtained via the 10-fold cross-validation (Sereshkeh et al., 2017a; Zeid and Chau, 2015).

In order to determine the differences in oscillatory engagement between speech tasks, we collected for each task, the 20 most discriminatory wavelet coefficients for each channel, for each participant. The coefficients from all channels and participants were then categorized by their corresponding pseudo-frequency value into one of the 5 EEG frequency bands. The number of discriminatory wavelet coefficients within each band were then compared between classification problems using a χ^2 test of homogenous proportions.

5.9 Measuring phase-amplitude coupling (PAC)

The existence of PAC in speech processing is well documented (Giraud and Poeppel, 2012; Hyafil et al., 2015; Assaneo and Poeppel, 2018; Voytek et al., 2013; Hermes et al., 2014). To investigate θ-γ coupling in SP and CS, the event-related phase-amplitude coupling (ERPAC) toolbox was utilized (Voytek et al., 2013). A 4^{th}-order Butterworth filter was applied between 4-7Hz to obtain θ-band signals, after which the angle of its Hilbert transform was taken as the corresponding phase

signal, ϕ_θ. The γ band signals were derived by Butterworth filtering (order 26) the pre-processed EEG signals between 30-60Hz (Voytek et al., 2013). Subsequently, the γ-band signal envelope, e_γ, was approximated by taking the absolute value of the Hilbert transform of the γ band signals. In a specific channel, the instantaneous PAC for each phase-amplitude signal pair was estimated according to Berens (2009), as the circular-linear correlation between θ phase and γ envelope at each time sample across all trials from all participants:

$$PAC_{\theta\gamma} = \sqrt{\frac{r_{\theta\gamma}^2 + q_{\theta\gamma}^2 - 2r_{\theta\gamma}q_{\theta\gamma}s_{\theta\theta}}{1 - s_{\theta\theta}^2}} \tag{2}$$

where $r_{\theta\gamma} = corr(\cos\phi_\theta, e_\gamma)$, $q_{\theta\gamma} = corr(\sin\phi_\theta, e_\gamma)$, and $s_{\theta\theta} = corr(\cos\phi_\theta, \sin\phi_\theta)$, are Pearson correlation coefficients, and ϕ_θ and e_γ are 1000×1 column vectors (i.e., 50 samples $\times$ 2 sessions $\times$ 10 participants) representing θ phase and γ envelope, respectively. Note that for readability, time dependence is suppressed in Equation 2.

To determine whether the observed PAC values exceeded that expected by chance, the observed PAC values were compared against those derived from 1000 surrogate runs. For each surrogate run for a given phase-amplitude comparison, the trials of the envelop data (e_γ) were randomly permuted and the correlation of Equation 2 was recomputed across trials at each time point. PAC values were standardized using a Fisher's z-transform:

$$\widehat{PAC} = \frac{1}{2}\ln\left(\frac{1 + PAC}{1 - PAC}\right) \tag{3}$$

where $\widehat{PAC} \sim N(0, 1)$. The standardized observed ($\widehat{PAC_{obs}}$) and surrogate ($\widehat{PAC_{surg}}$) phase-amplitude coupling values at each instance were then compared using a z-test:

$$z = \frac{\widehat{PAC}_{obs} - \widehat{PAC}_{surg}}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} \tag{4}$$

where $n_1 = n_2 = 1000$ in this analysis and for simplicity, the $\theta\gamma$ subscript has been omitted. The time points where the z-test yielded p<0.05 identified significant θ-γ PAC within a specific channel

across trials.

5.10 Task-specific effects

To determine the phase consistency across trials within a given task, inter-trial coherence (Delorme and Makeig, 2004) was computed after pooling data from all participants. Likewise, to visualize task-specific, mean event-related changes in spectral power over time within the frequency bands of interest, we computed the event-related spectral perturbations (Delorme and Makeig, 2004) by averaging spectrograms of task-specific responses across trials.

5.11 Intertask coupling

For cross-task PAC (e.g. SP θ-CS γ), an analysis identical to the above was conducted but replacing the envelope values with those of the corresponding CS class. To test for significant correlations between the θ phases of SP and CS classes for a given speech token, the circular correlation test between two circular random variables was invoked as in Berens (2009).

$$c(\Psi_\theta, \Omega_\theta) = \frac{\sum_i \sin(\psi_i - \overline{\Psi_\theta}) \sin(\omega_i - \overline{\Omega_\theta})}{\sqrt{\sum_i \sin^2(\psi_i - \overline{\Psi_\theta}) \sin^2(\omega_i - \overline{\Omega_\theta})}} \tag{5}$$

where $\Psi_\theta = \{\psi_1, \ldots, \psi_N\}$ and $\Omega_\theta = \{\omega_1, \ldots, \omega_N\}$ denote respectively, the θ phases of speech perception and covert speech, $\overline{\Psi_\theta}$ and $\overline{\Omega_\theta}$ denote their means, and $N = 1000$ as in Equation 2. Under the null hypothesis of no significant correlations, the p-value for this correlation was computed by a normally distributed test statistic (Berens, 2009).

6 Acknowledgements

We thank Christine Horner and Sarah Holman for their help in data collection, and Ka Lun Tam and Pierre Duez for assisting in developing the protocol.

7 Funding

This research was funded by the Toronto Rehabilitation Institute and the Margaret and Howard GAMBLE Research Grant.

There are 39070 characters and approximately 7186 spaces. That makes approximately 46256 characters total.

References

P. Albouy, A. Weiss, S. Baillet, and R. J. Zatorre. Selective Entrainment of Theta Oscillations in the Dorsal Stream Causally Enhances Auditory Working Memory Performance. *Neuron*, 94(1): 193–206.e5, 2017. ISSN 10974199. doi: 10.1016/j.neuron.2017.03.015.

B. Alderson-Day and C. Fernyhough. Inner Speech: Development, Cognitive Functions, Phenomenology, and Neurobiology Ben. *Cirugia Espanola*, 90(9):545–547, 2012. ISSN 0009739X. doi: 10.1016/j.ciresp.2012.05.015.

B. Alderson-Day, K. Mitrenga, S. Wilkinson, S. McCarthy-Jones, and C. Fernyhough. The varieties of inner speech questionnaire – Revised (VISQ-R): Replicating and refining links between inner speech and psychopathology. *Consciousness and Cognition*, 65(July):48–58, 2018. ISSN 10902376. doi: 10.1016/j.concog.2018.07.001. URL https://doi.org/10.1016/j.concog.2018.07.001.

M. F. Assaneo and D. Poeppel. The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. *Science Advances*, 4(2):1–10, 2018. ISSN 23752548. doi: 10.1126/sciadv.aao3842.

F. Babiloni, F. Cincotti, F. Carducci, P. M. Rossini, and C. Babiloni. Spatial enhancement of EEG data by surface Laplacian estimation: The use of magnetic resonance imaging-based head models. *Clinical Neurophysiology*, 112(5):724–727, 2001. ISSN 13882457. doi: 10.1016/S1388-2457(01)00494-1.

M. Bastiaansen and P. Hagoort. Chapter 12 Oscillatory neuronal dynamics during language comprehension. *Progress in Brain Research*, 159(06):179–196, 2006. ISSN 00796123. doi: 10.1016/S0079-6123(06)59012-0.

M. Bastiaansen, L. Magyari, and P. Hagoort. Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *Journal of Cognitive Neuroscience*, 22(7):1333–1347, 2010. ISSN 0898929X. doi: 10.1162/jocn.2009.21283.

P. Berens. CircStat: a MATLAB toolbox for circular statistics. *Journal of Statistical Software*, 31 (10), 2009. ISSN 00085286.

G. M. Bidelman. Induced neural beta oscillations predict categorical speech perception abilities. *Brain and Language*, 141:62–69, 2015. ISSN 10902155. doi: 10.1016/j.bandl.2014.11.003. URL http://dx.doi.org/10.1016/j.bandl.2014.11.003.

N. Bigdely-shamlo, T. Mullen, C. Kothe, K.-m. Su, and A. Widmann. The PREP pipeline : standardized preprocessing for large-scale EEG analysis. *Frontiers in neuroinformatics*, 9(June): 1–20, 2015. doi: 10.3389/fninf.2015.00016.

A. Boemio, S. Fromm, A. Braun, and D. Poeppel. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, 8(3):389–395, 2005. ISSN 10976256. doi: 10.1038/nn1409.

V. Bostanov. BCI competition 2003 - Data sets Ib and IIb: Feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. *IEEE Transactions on Biomedical Engineering*, 51(6):1057–1061, 2004. ISSN 00189294. doi: 10.1109/TBME.2004.826702.

B. R. Buchsbaum, G. Hickok, and C. Humphries. Cognitive Science : A Multidisciplinary Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science*, 25(784375790):663–678, 2001. doi: 10.1207/s15516709cog2505.

G. Buzsáki and A. Draguhn. Neuronal olscillations in cortical networks. *Science*, 304(5679): 1926–1929, 2004. ISSN 00368075. doi: 10.1126/science.1099745.

G. Buzsáki, C. Geisler, D. A. Henze, and X. J. Wang. Interneuron Diversity series: Circuit complexity and axon wiring economy of cortical interneurons. *Trends in Neurosciences*, 27(4): 186–193, 2004. ISSN 01662236. doi: 10.1016/j.tins.2004.02.007.

C. C. Canavier. Phase-resetting as a tool of information transmission. *Current Opinion in Neurobiology*, 31:206–213, 2015. ISSN 18736882. doi: 10.1016/j.conb.2014.12.003. URL http://dx.doi.org/10.1016/j.conb.2014.12.003.

E. F. Chang, J. W. Rieger, K. Johnson, M. S. Berger, N. M. Barbaro, and R. T. Knight. Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, 13(11):1428–1432, 2010. ISSN 10976256. doi: 10.1038/nn.2641.

C. M. A. Chen, D. H. Mathalon, B. J. Roach, I. Cavus, D. D. Spencer, and J. M. Ford. The corollary discharge in humans is related to synchronous neural oscillations. *Journal of Cognitive Neuroscience*, 23(10):2892–2904, 2011. ISSN 0898929X. doi: 10.1162/jocn.2010.21589.

H. Chen, W. Chen, Y. Song, L. Sun, and X. Li. EEG characteristics of children with attention-deficit/hyperactivity disorder. *Neuroscience*, 406:444–456, 2019. ISSN 18737544. doi: 10.1016/j.neuroscience.2019.03.048. URL https://doi.org/10.1016/j.neuroscience.2019.03.048.

K. E. Cullen. Sensory signals during active versus passive movement. *Current Opinion in Neurobiology*, 14(6):698–706, 2004. ISSN 09594388. doi: 10.1016/j.conb.2004.10.002.

S. Darvishi and A. Al-Ani. Brain-computer interface analysis using continuous wavelet transform and adaptive neuro-fuzzy classifier. *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, pages 3220–3223, 2007. ISSN 05891019. doi: 10.1109/IEMBS.2007.4353015.

C. DaSalla, H. Kambara, Y. Koike, and M. Sato. Spatial filtering and single-trial classification of EEG during vowel speech imagery. *International Convention on Rehabilitation Engineering and Assistive Technology (ICREAT)*, 5:1–4, 2009. doi: 10.1145/1592700.1592731. URL http://dl.acm.org/citation.cfm?id=1592731.

A. Delorme and S. Makeig. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1): 9–21, 2004. ISSN 01650270. doi: 10.1016/j.jneumeth.2003.10.009.

S. Deng, R. Srinivasan, T. Lappas, and M. D'Zmura. EEG classification of imagined syllable rhythm using Hilbert spectrum methods. *Journal of Neural Engineering*, 7(4), 2010. ISSN 17412560. doi: 10.1088/1741-2560/7/4/046006.

G. M. Di Liberto, J. A. O'Sullivan, and E. C. Lalor. Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology*, 25(19):2457–2465, 2015. ISSN 09609822. doi: 10.1016/j.cub.2015.08.030. URL http://dx.doi.org/10.1016/j.cub.2015.08.030.

C. Ding and H. Peng. Minimum redundancy feature selection from microarray gene expression data. *Journal of Bioinformatics and Computational Biology*, 3(2):185–205, 2005. ISSN 02197200. doi: 10.1142/S0219720005001004.

N. Ding, L. Melloni, H. Zhang, X. Tian, and D. Poeppel. Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, 19(1):158–164, 2015. ISSN 15461726. doi: 10.1038/nn.4186.

N. Ding, L. Melloni, A. Yang, Y. Wang, W. Zhang, and D. Poeppel. Characterizing Neural Entrainment to Hierarchical Linguistic Units using Electroencephalography (EEG). *Frontiers in Human Neuroscience*, 11(September):1–9, 2017. doi: 10.3389/fnhum.2017.00481.

K. B. Doelling, L. H. Arnal, O. Ghitza, and D. Poeppel. Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, 85:761–768, 2014. ISSN 10959572. doi: 10.1016/j.neuroimage.2013.06.035. URL http://dx.doi.org/10.1016/j.neuroimage.2013.06.035.

M. Ende, A. K. Louis, P. Maass, and G. Mayer-Kress. EEG Signal Analysis by Continuous Wavelet Transform Techniques. In *Nonlinear Analysis of Physiological Data*, number 1, pages 213–219. Springer, 1998. doi: 10.1007/978-3-642-71949-3_12.

J. M. Ford and D. H. Mathalon. Corollary discharge dysfunction in schizophrenia: Can it explain auditory hallucinations? *International Journal of Psychophysiology*, 58(2-3 SPEC. ISS.):179–189, 2005. ISSN 01678760. doi: 10.1016/j.ijpsycho.2005.01.014.

J. M. Ford and D. H. Mathalon. Efference Copy, Corollary Discharge, Predictive Coding, and Psychosis. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 4(9):764–767, 2019. ISSN 24519030. doi: 10.1016/j.bpsc.2019.07.005. URL https://doi.org/10.1016/j.bpsc.2019.07.005.

P. Fries. Neuronal Gamma-Band Synchronization as a Fundamental Process in Cortical Computation. *Annual Review of Neuroscience*, 32(1):209–224, 2009. ISSN 0147-006X. doi: 10.1146/annurev.neuro.051508.135603.

M. Fukuda, R. Rothermel, C. Juhász, M. Nishida, S. Sood, and E. Asano. Cortical gamma-oscillations modulated by listening and overt repetition of phonemes. *NeuroImage*, 49(3):2735–2745, 2010. ISSN 10538119. doi: 10.1016/j.neuroimage.2009.10.047. URL http://dx.doi.org/10.1016/j.neuroimage.2009.10.047.

J. Gallinat, G. Winterer, C. S. Herrmann, and D. Senkowski. Reduced oscillatory gamma-band responses in unmedicated schizophrenic patients indicate impaired frontal network processing. *Clinical Neurophysiology*, 115(8):1863–1874, 2004. ISSN 13882457. doi: 10.1016/j.clinph.2004.03.013.

O. Ghitza. On the role of theta-driven syllabic parsing in decoding speech: Intelligibility of speech with a manipulated modulation spectrum. *Frontiers in Psychology*, 3(JUL):1–12, 2012. ISSN 16641078. doi: 10.3389/fpsyg.2012.00238.

O. Ghitza. The theta-syllable: A unit of speech information defined by cortical function. *Frontiers in Psychology*, 4(MAR):1–5, 2013. ISSN 16641078. doi: 10.3389/fpsyg.2013.00138.

A. L. Giraud and D. Poeppel. Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15(4):511–517, 2012. ISSN 10976256. doi: 10.1038/nn.3063. URL http://dx.doi.org/10.1038/nn.3063.

A. L. Giraud, C. Lorenzi, J. Ashburner, J. Wable, I. Johnsrude, R. Frackowiak, and A. Kleinschmidt. Representation of the temporal envelope of sounds in the human brain. *Journal of Neurophysiology*, 84(3):1588–1598, 2000. ISSN 00223077. doi: 10.1152/jn.2000.84.3.1588.

A. L. Giraud, A. Kleinschmidt, D. Poeppel, T. E. Lund, R. S. J. Frackowiak, and H. Laufs. Endogenous Cortical Rhythms Determine Cerebral Specialization for Speech Perception and Production. *Neuron*, 56(6):1127–1134, 2007. ISSN 08966273. doi: 10.1016/j.neuron.2007.09.038.

G. Gomez-Herrero. Automatic Artifact Removal (AAR) toolbox Germ ´. *Tampere University of Technology*, 3(January 2007):1–23, 2007.

D. M. Groppe, T. P. Urbach, and M. Kutas. Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, 48(12):1711–1725, 2011. ISSN 14698986. doi: 10.1111/j.1469-8986.2011.01273.x.

J. Gross, N. Hoogenboom, G. Thut, P. Schyns, S. Panzeri, P. Belin, and S. Garrod. Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain. *PLoS Biology*, 11(12), 2013. ISSN 15449173. doi: 10.1371/journal.pbio.1001752.

A. Guérin-Dugué, R. N. Roy, E. Kristensen, B. Rivet, L. Vercueil, and A. Tcherkassof. Temporal dynamics of natural static emotional facial expressions decoding: A study using event- and eye fixation-related potentials. *Frontiers in Psychology*, 9(JUN):1–19, 2018. ISSN 16641078. doi: 10.3389/fpsyg.2018.01190.

T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning. *Springer Series in Statistics*, 27(2), 2009. ISSN 03436993. URL http://www.springerlink.com/index/D7X7KX6772HQ2135.pdf.

D. Hermes, K. J. Miller, M. J. Vansteensel, E. Edwards, C. H. Ferrier, M. G. Bleichner, P. C. van Rijen, E. J. Aarnoutse, and N. F. Ramsey. Cortical theta wanes for language. *NeuroImage*, 85:738–748, 2014. ISSN 10959572. doi: 10.1016/j.neuroimage.2013.07.029. URL http://dx.doi.org/10.1016/j.neuroimage.2013.07.029.

G. Hickok. The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience*, 29(1):2–20, 2014. ISSN 23273801. doi: 10.1080/01690965.2013.834370.

G. Hickok and D. Poeppel. Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1-2):67–99, 2004. ISSN 00100277. doi: 10.1016/j.cognition.2003.10.011.

G. Hickok and D. Poeppel. The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5):393–402, 2007. ISSN 1471-003X. doi: 10.1038/nrn2113. URL http://www.nature.com/doifinder/10.1038/nrn2113.

G. Hickok, K. Okada, and J. T. Serences. Area Spt in the Human Planum Temporale Supports Sensory-Motor Integration for Speech Processing. *Journal of Neurophysiology*, 101(5):2725–2732, 2009. ISSN 0022-3077. doi: 10.1152/jn.91099.2008.

G. Hickok, J. Houde, and F. Rong. Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. *Neuron*, 69(3):407–422, 2011. ISSN 08966273. doi: 10.1016/j.neuron.2011.01.019. URL http://dx.doi.org/10.1016/j.neuron.2011.01.019.

B. Hosseinifard, M. H. Moradi, and R. Rostami. Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal. *Comput. Methods Programs Biomed.*, 109(3):339–345, 2013. ISSN 01692607. doi: 10.1016/j.cmpb.2012.10.008. URL http://dx.doi.org/10.1016/j.cmpb.2012.10.008.

W. Y. Hsu, C. C. Lin, M. S. Ju, and Y. N. Sun. Wavelet-based fractal features with active segment selection: Application to single-trial EEG data. *Journal of Neuroscience Methods*, 163(1): 145–160, 2007. ISSN 01650270. doi: 10.1016/j.jneumeth.2007.02.004.

J. Huang, T. Hensch, C. Ulke, C. Sander, J. Spada, P. Jawinski, and U. Hegerl. Evoked potentials and behavioral performance during different states of brain arousal. *BMC Neuroscience*, 18(1): 1–11, 2017. ISSN 14712202. doi: 10.1186/s12868-017-0340-9.

A. Hyafil, L. Fontolan, C. Kabdebon, B. Gutkin, and A. L. Giraud. Speech encoding by coupled cortical theta and gamma oscillations. *eLife*, 4(MAY):1–45, 2015. ISSN 2050084X. doi: 10.7554/eLife.06213.

B. M. Idrees and O. Farooq. Vowel classification using wavelet decomposition during speech imagery. *3rd International Conference on Signal Processing and Integrated Networks, SPIN 2016*, pages 636–640, 2016. doi: 10.1109/SPIN.2016.7566774.

B. N. Jack, M. E. Le Pelley, N. Han, A. W. Harris, K. M. Spencer, and T. J. Whitford. Inner speech is accompanied by a temporally-precise and content-specific corollary discharge. *NeuroImage*, 198(March):170–180, 2019. ISSN 10959572. doi: 10.1016/j.neuroimage.2019.04.038. URL https://doi.org/10.1016/j.neuroimage.2019.04.038.

J. K. Johannesen, J. Bi, R. Jiang, J. G. Kenney, and C.-M. A. Chen. Machine learning identification of EEG features predicting working memory performance in schizophrenia and healthy adults. *Neuropsychiatr. Electrophysiol.*, 2(1):1–21, 2016. ISSN 2055-4788. doi: 10.1186/s40810-016-0017-0. URL http://dx.doi.org/10.1186/s40810-016-0017-0.

J. Kayser and C. E. Tenke. Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: I. Evaluation with auditory oddball tasks. *Clinical Neurophysiology*, 117(2):348–368, 2006. ISSN 13882457. doi: 10.1016/j.clinph.2005.08.034.

J. Kayser and C. E. Tenke. Issues and considerations for using the scalp surface Laplacian in EEG/ERP research: A tutorial review. *Int. J. Psychophysiol.*, 97(3):189–209, 2015. ISSN 18727697. doi: 10.1016/j.ijpsycho.2015.04.012. URL http://dx.doi.org/10.1016/j.ijpsycho.2015.04.012.

A. Kimata, Y. Yokoyama, S. Aita, H. Nakamura, K. Higuchi, Y. Tanaka, A. Nogami, K. Hirao, and K. Aonuma. Temporally stable frequency mapping using continuous wavelet transform analysis

in patients with persistent atrial fibrillation. *Journal of Cardiovascular Electrophysiology*, 29(4): 514–522, 2018. ISSN 15408167. doi: 10.1111/jce.13440.

J. M. Lilly. Element analysis: a wavelet-based method for analyzing time-localized events in noisy time series. *arXiv*, 2017. ISSN 23318422.

J. M. Lilly and S. C. Olhede. Higher-order properties of analytic wavelets. *IEEE Transactions on Signal Processing*, 57(1):146–160, 2009. ISSN 1053587X. doi: 10.1109/TSP.2008.2007607.

M. Lizarazu, M. Lallier, and N. Molinaro. Phase amplitude coupling between theta and gamma oscillations adapts to speech rate. *Annals of the New York Academy of Sciences*, 1493(April): 140–152, 2019. ISSN 17496632. doi: 10.1111/nyas.14099.

F. Lotte and C. Jeunet. Defining and quantifying users' mental imagery-based BCI skills: A first step. *Journal of Neural Engineering*, 15(4), 2018. ISSN 17412552. doi: 10.1088/1741-2552/aac577.

H. Luo and D. Poeppel. Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron*, 54(6):1001–1010, 2007. ISSN 08966273. doi: 10.1016/j.neuron.2007.06.004.

G. Mai, J. W. Minett, and W. S. Wang. Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing. *NeuroImage*, 133, 2016. ISSN 10959572. doi: 10.1016/j.neuroimage.2016.02.064.

D. H. Mathalon and J. M. Ford. Corollary discharge dysfunction in schizophrenia: Evidence for an elemental deficit. *Clinical EEG and Neuroscience*, 39(2):82–86, 2008. ISSN 15500594. doi: 10.1177/155005940803900212.

M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger. Montreal forced aligner: Trainable text-speech alignment using kaldi. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2017-Augus:498–502, 2017. ISSN 19909772. doi: 10.21437/Interspeech.2017-1386.

R. E. Mirollo and S. H. Strogatz. Synchronization of pulse-coupled biological oscillators. *SIAM Journal on Applied Mathematics*, 50(6):1645–1662, 1990. ISSN 00361399. doi: 10.1137/0150098.

S. Moghimi, A. Kushki, A. Marie Guerguerian, and T. Chau. A review of EEG-Based brain-computer interfaces as access pathways for individuals with severe disabilities. *Assistive Technology*, 25(2): 99–110, 2013. ISSN 10400435. doi: 10.1080/10400435.2012.723298.

B. Morillon and C. E. Schroeder. Neuronal oscillations as a mechanistic substrate of auditory temporal prediction. *Annals of the New York Academy of Sciences*, 1337(1):26–31, 2015. ISSN 17496632. doi: 10.1111/nyas.12629.

B. Morillon, C. Liégeois-Chauvel, L. H. Arnal, C. G. Bénar, and A. L. Giraud. Asymmetric function of theta and gamma activity in syllable processing: An intra-cortical study. *Frontiers in Psychology*, 3(JUL):1–9, 2012. ISSN 16641078. doi: 10.3389/fpsyg.2012.00248.

B. Morillon, L. H. Arnal, C. E. Schroeder, and A. Keitel. Prominence of delta oscillatory rhythms in the motor cortex and their relevance for auditory and speech perception. *Neuroscience and Biobehavioral Reviews*, 107(September):136–142, 2019. ISSN 18737528. doi: 10.1016/j.neubiorev.2019.09.012. URL https://doi.org/10.1016/j.neubiorev.2019.09.012.

A. Morin, B. Uttl, and B. Hamper. Self-reported frequency, content, and functions of inner speech. *Procedia - Social and Behavioral Sciences*, 30:1714–1718, 2011. ISSN 18770428. doi: 10.1016/j.sbspro.2011.10.331.

A. Morin, C. Duhnych, and F. Racy. Self-reported inner speech use in university students. *Applied Cognitive Psychology*, 32(3):376–382, 2018. ISSN 10990720. doi: 10.1002/acp.3404.

D. A. Moses, N. Mesgarani, M. K. Leonard, and E. F. Chang. Neural speech recognition: Continuous phoneme decoding using spatiotemporal representations of human cortical activity. *Journal of Neural Engineering*, 13(5):1–19, 2016. ISSN 17412552. doi: 10.1088/1741-2560/13/5/056004. URL http://dx.doi.org/10.1088/1741-2560/13/5/056004.

A. Myrden and T. Chau. Effects of user mental state on EEG-BCI performance. *Frontiers in Human Neuroscience*, 9(June):1–11, 2015. doi: 10.3389/fnhum.2015.00308.

K. Okada and G. Hickok. Left posterior auditory-related cortices participate both in speech perception and speech production: Neural overlap revealed by fMRI. *Brain and Language*, 98(1): 112–117, 2006. ISSN 0093934X. doi: 10.1016/j.bandl.2006.04.006.

K. Okada, W. Matchin, and G. Hickok. Neural evidence for predictive coding in auditory cortex during speech production. *Psychonomic Bulletin and Review*, 25(1):423–430, 2018. ISSN 15315320. doi: 10.3758/s13423-017-1284-x.

T. Onojima, K. Kitajo, and H. Mizuhara. Ongoing slow oscillatory phase modulates speech intelligibility in cooperation with motor cortical activity. *PLoS ONE*, 12(8):1–17, 2017. ISSN 19326203. doi: 10.1371/journal.pone.0183146.

A. C. Onslow, R. Bogacz, and M. W. Jones. Quantifying phase-amplitude coupling in neuronal network oscillations. *Progress in Biophysics and Molecular Biology*, 105(1-2):49–57, 2011. ISSN 00796107. doi: 10.1016/j.pbiomolbio.2010.09.007. URL http://dx.doi.org/10.1016/j.pbiomolbio.2010.09.007.

B. N. Pasley, S. V. David, N. Mesgarani, A. Flinker, S. A. Shamma, N. E. Crone, R. T. Knight, and E. F. Chang. Reconstructing speech from human auditory cortex. *PLoS Biology*, 10(1), 2012. ISSN 15449173. doi: 10.1371/journal.pbio.1001251.

X. Pei, E. C. Leuthardt, C. M. Gaona, P. Brunner, J. R. Wolpaw, and G. Schalk. Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition. *NeuroImage*, 54(4):2960–2972, 2011. ISSN 10538119. doi: 10.1016/j.neuroimage.2010.10.029. URL http://dx.doi.org/10.1016/j.neuroimage.2010.10.029.

M. Perrone-Bertolotti, L. Rapin, J. P. Lachaux, M. Baciu, and H. Lœvenbruck. What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behavioural Brain Research*, 261:220–239, 2014. ISSN 01664328. doi: 10.1016/j.bbr.2013.12.034. URL http://dx.doi.org/10.1016/j.bbr.2013.12.034.

D. Poeppel. The neuroanatomic and neurophysiological infrastructure for speech and language. *Current Opinion in Neurobiology*, 28:142–149, 2014. ISSN 18736882. doi: 10.1016/j.conb.2014.07.005. URL http://dx.doi.org/10.1016/j.conb.2014.07.005.

D. Poeppel and M. F. Assaneo. Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, 21(6):322–334, 2020. ISSN 14710048. doi: 10.1038/s41583-020-0304-4. URL http://dx.doi.org/10.1038/s41583-020-0304-4.

R. G. Real and B. Kotchoubey. Studentized continuous wavelet transform (t-CWT) in the analysis of individual ERPs: Real and simulated EEG data. *Frontiers in Neuroscience*, 8(SEP):1–9, 2014. ISSN 1662453X. doi: 10.3389/fnins.2014.00279.

J. Restle, T. Murakami, and U. Ziemann. Facilitation of speech repetition accuracy by theta burst stimulation of the left posterior inferior frontal gyrus. *Neuropsychologia*, 50(8):2026–2031, 2012. ISSN 00283932. doi: 10.1016/j.neuropsychologia.2012.05.001. URL http://dx.doi.org/10.1016/j.neuropsychologia.2012.05.001.

A. Rezazadeh Sereshkeh, R. Yousefi, A. T. Wong, and T. Chau. Online classification of imagined speech using functional near-infrared spectroscopy signals. *Journal of Neural Engineering*, 16(1):1–13, 2019. ISSN 17412552. doi: 10.1088/1741-2552/aae4b9.

M. J. Roberts, E. Lowet, N. M. Brunet, M. TerWal, P. Tiesinga, P. Fries, and P. DeWeerd. Robust gamma coherence between macaque V1 and V2 by dynamic frequency matching. *Neuron*, 78(3):523–536, 2013. ISSN 08966273. doi: 10.1016/j.neuron.2013.03.003. URL http://dx.doi.org/10.1016/j.neuron.2013.03.003.

S. Sadaghiani, R. Scheeringa, K. Lehongre, B. Morillon, A. L. Giraud, M. D'Esposito, and A. Kleinschmidt. Alpha-band phase synchrony is related to activity in the fronto-parietal adaptive control network. *J. Neurosci.*, 32(41):14305–14310, 2012. ISSN 02706474. doi: 10.1523/JNEUROSCI.1358-12.2012.

P. Sauseng, B. Griesmayr, R. Freunberger, and W. Klimesch. Control mechanisms in working memory: A possible function of EEG theta oscillations. *Neuroscience and Biobehavioral*

Reviews, 34(7):1015–1022, 2010. ISSN 01497634. doi: 10.1016/j.neubiorev.2009.12.006. URL http://dx.doi.org/10.1016/j.neubiorev.2009.12.006.

C. E. Schroeder and P. Lakatos. Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, 32(1):9–18, 2009. ISSN 01662236. doi: 10.1016/j.tins.2008.09.012.

C. E. Schroeder, P. Lakatos, Y. Kajikawa, S. Partan, and A. Puce. Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences*, 12(3):106–113, 2008. ISSN 13646613. doi: 10.1016/j.tics.2008.01.002.

M. Scott. Corollary Discharge Provides the Sensory Content of Inner Speech. *Psychological Science*, 24(9):1824–1830, 2013. ISSN 14679280. doi: 10.1177/0956797613478614.

S. K. Scott. The neurobiology of speech perception and production-Can functional imaging tell us anything we did not already know? *Journal of Communication Disorders*, 45(6):419–425, 2012. ISSN 00219924. doi: 10.1016/j.jcomdis.2012.06.007. URL http://dx.doi.org/10.1016/j.jcomdis.2012.06.007.

D. Senkowski and C. S. Herrmann. Effects of task difficulty on evoked gamma activity and ERPs in a visual discrimination task. *Clinical Neurophysiology*, 113(11):1742–1753, 2002. ISSN 13882457. doi: 10.1016/S1388-2457(02)00266-3.

A. R. Sereshkeh, R. Trott, A. Bricout, and T. Chau. Online EEG Classification of Covert Speech for Brain–Computer Interfacing. *International Journal of Neural Systems*, 27(08):1750033, 2017a. ISSN 0129-0657. doi: 10.1142/s0129065717500332.

A. R. Sereshkeh, R. Trott, A. Bricout, and T. Chau. Online EEG Classification of Covert Speech for Brain–Computer Interfacing. *International Journal of Neural Systems*, 27(0):1750033, 2017b. ISSN 0129-0657. doi: 10.1142/S0129065717500332. URL http://www.worldscientific.com/doi/abs/10.1142/S0129065717500332.

A. J. Shahin, T. W. Picton, and L. M. Miller. Brain oscillations during semantic evaluation of speech. *Brain and Cognition*, 70(3):259–266, 2009. ISSN 02782626. doi: 10.1016/j.bandc.2009.02.008. URL http://dx.doi.org/10.1016/j.bandc.2009.02.008.

S. S. Shergill, M. J. Brammer, R. Fukuda, E. Bullmore, E. Amaro, R. M. Murray, and P. K. McGuire. Modulation of activity in temporal cortex during generation of inner speech. *Human Brain Mapping*, 16(4):219–227, 2002. ISSN 10659471. doi: 10.1002/hbm.10046.

J. I. Skipper, H. C. Nusbaum, and S. L. Small. Listening to talking faces: Motor cortical activation during speech perception. *NeuroImage*, 25(1):76–89, 2005. ISSN 10538119. doi: 10.1016/j.neuroimage.2004.11.006.

U. Talukdar, S. M. Hazarika, and J. Q. Gan. Motor imagery and mental fatigue: inter-relationship and EEG based estimation. *J. Comput. Neurosci.*, 46(1):55–76, 2019. ISSN 15736873. doi: 10.1007/s10827-018-0701-0.

R. W. Thatcher, D. M. North, and C. J. Biver. Intelligence and EEG phase reset: A two compartmental model of phase shift and lock. *NeuroImage*, 42(4):1639–1653, 2008. ISSN 10538119. doi: 10.1016/j.neuroimage.2008.06.009.

X. Tian and D. Poeppel. Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology*, 1(OCT):1–23, 2010. ISSN 16641078. doi: 10.3389/fpsyg.2010.00166.

X. Tian and D. Poeppel. Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. *Frontiers in Human Neuroscience*, 6(November):1–11, 2012. ISSN 1662-5161. doi: 10.3389/fnhum.2012.00314. URL http://journal.frontiersin.org/article/10.3389/fnhum.2012.00314/abstract.

G. Toyoda, E. C. Brown, N. Matsuzaki, K. Kojima, M. Nishida, and E. Asano. Electrocorticographic correlates of overt articulation of 44 English phonemes: Intracranial recording in children with focal epilepsy. *Clinical Neurophysiology*, 125(6):1129–1137, 2014. ISSN 18728952. doi: 10.1016/j.clinph.2013.11.008. URL http://dx.doi.org/10.1016/j.clinph.2013.11.008.

J. Trouvain. On the comprehension of extremely fast synthetic speech. *Saarland Working Papers in Linguistics (SWPL)*, 1(1):5–13, 2007.

P. J. Uhlhaas and W. Singer. Abnormal neural oscillations and synchrony in schizophrenia. *Nature Reviews Neuroscience*, 11(2):100–113, 2010. ISSN 1471003X. doi: 10.1038/nrn2774.

P. J. Uhlhaas, D. E. Linden, W. Singer, C. Haenschel, M. Lindner, K. Maurer, and E. Rodriguez. Dysfunctional long-range coordination of neural activity during gestalt perception in schizophrenia. *Journal of Neuroscience*, 26(31):8168–8175, 2006. ISSN 02706474. doi: 10.1523/JNEUROSCI.2002-06.2006.

V. van de Ven, F. Esposito, and I. K. Christoffels. Neural network of speech monitoring overlaps with overt speech production and comprehension networks: A sequential spatial and temporal ICA study. *NeuroImage*, 47(4):1982–1991, 2009. ISSN 10538119. doi: 10.1016/j.neuroimage.2009.05.057. URL http://dx.doi.org/10.1016/j.neuroimage.2009.05.057.

R. van Lutterveld, I. E. C. Sommer, and J. M. Ford. The Neurophysiology of Auditory Hallucinations – A Historical and Contemporary Review. *Frontiers in Psychiatry*, 2(May):1–7, 2011. ISSN 1664-0640. doi: 10.3389/fpsyt.2011.00028. URL http://journal.frontiersin.org/article/10.3389/fpsyt.2011.00028/abstract.

L. Varnet, M. C. Ortiz-Barajas, R. G. Erra, J. Gervain, and C. Lorenzi. A cross-linguistic study of speech modulation spectra. *The Journal of the Acoustical Society of America*, 142(4):1976–1989, 2017. ISSN 0001-4966. doi: 10.1121/1.5006179. URL http://dx.doi.org/10.1121/1.5006179.

J. H. Venezia, P. Fillmore, W. Matchin, A. Lisette Isenberg, G. Hickok, and J. Fridriksson. Perception drives production across sensory modalities: A network for sensorimotor integration of visual speech. *NeuroImage*, 126:196–207, 2016. ISSN 10959572. doi: 10.1016/j.neuroimage.2015.11.038. URL http://dx.doi.org/10.1016/j.neuroimage.2015.11.038.

B. Voloh and T. Womelsdorf. A role of phase-resetting in coordinating large scale neural networks during attention and goal-directed behavior. *Frontiers in Systems Neuroscience*, 10(MAR):1–19, 2016. ISSN 16625137. doi: 10.3389/fnsys.2016.00018.

B. Voytek, M. D. Esposito, N. Crone, and R. T. Knight. A method for event-related phase / amplitude coupling. *NeuroImage*, 64:416–424, 2013. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2012.09.023. URL http://dx.doi.org/10.1016/j.neuroimage.2012.09.023.

S. Weiss and H. M. Mueller. The contribution of EEG coherence to the investigation of language. *Brain and Language*, 85(2):325–343, 2003. ISSN 0093934X. doi: 10.1016/S0093-934X(03)00067-1.

S. Weiss and H. M. Mueller. "Too many betas do not spoil the broth": The role of beta brain oscillations in language processing. *Frontiers in Psychology*, 3(JUN):1–15, 2012. ISSN 16641078. doi: 10.3389/fpsyg.2012.00201.

D. M. Wolpert and Z. Ghahramani. Computational principles of movement neuroscience. *Nature Neuroscience*, 3(11s):1212–1217, 2000. ISSN 15461726. doi: 10.1038/81497.

B. Yao, J. R. Taylor, B. Banks, and S. A. Kotz. Theta activity phase-locks to inner speech in silent reading. *PsyArXiv*, 44(0), 2020. doi: 10.31234/osf.io/vta7d. URL https://psyarxiv.com/vta7d/.

E. A. Zeid and T. Chau. Electrode Fusion for the Prediction of Self-Initiated Fine Movements from Single-Trial Readiness Potentials. 25(4):1–14, 2015. doi: 10.1142/S0129065715500148.

B. Zoefel and R. VanRullen. EEG oscillations entrain their phase to high-level features of speech sound. *NeuroImage*, 124:16–23, 2016. ISSN 10959572. doi: 10.1016/j.neuroimage.2015.08.054.