A COMPARISON OF OSCILLATORY CHARACTERISTICS IN COVERT SPEECH AND SPEECH PERCEPTION

A PREPRINT

Jae Moon
Institute of Biomedical Engineering
University of Toronto
Bloorview Research Institute
Holland Bloorview Kid’s Rehabilitation Hospital
jae.moon@mail.utoronto.ca

Silvia Orlandi
Bloorview Research Institute
Holland Bloorview Kid’s Rehabilitation Hospital
sorlandi@hollandbloorview.ca

Tom Chau
Institute of Biomedical Engineering
University of Toronto
Bloorview Research Institute
Holland Bloorview Kid’s Rehabilitation Hospital
tom.chau@utoronto.ca

September 8, 2020

ABSTRACT

Covert speech, the silent production of words in the mind, has been studied increasingly to understand and decode thoughts. This task has often been compared to speech perception as it brings about similar topographical activation patterns in common brain areas. In studies of speech comprehension, neural oscillations are thought to play a key role in the sampling of speech at varying temporal scales. However, very little is known about the role of oscillations in covert speech. In this study, we aimed to determine to what extent each oscillatory frequency band is used to process words in covert speech and speech perception tasks. Secondly, we asked whether the \( \theta \) and \( \gamma \) activity in the two tasks are related through phase-amplitude coupling (PAC). First, continuous wavelet transform was performed on epoched signals and subsequently two-tailed t-tests between two classes were conducted to determine statistical distinctions in frequency and time. While the perception task dynamically uses all frequencies with more prominent \( \theta \) and \( \gamma \) activity, the covert task favoured higher frequencies with significantly higher \( \gamma \) activity than perception. Moreover, the perception condition produced significant \( \theta \)-\( \gamma \) PAC suggesting a linkage of syllabic and phonological sampling. Although this was found to be suppressed in the covert condition, we found significant pseudo-coupling between perception \( \theta \) and covert speech \( \gamma \). We report that covert speech processing is largely conducted by higher frequencies, and that the \( \gamma \)- and \( \theta \)-bands may function similarly and differently across tasks, respectively. This study is the first to characterize covert speech in terms of neural oscillatory engagement. Future studies are directed to explore oscillatory characteristics and inter-task relationships with a more diverse vocabulary.

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 speech in motor-impaired individuals by decoding thoughts (DaSalla et al., 2009).
was first introduced by Bostanov (2004) with the intention of improving feature extraction methods for classification whereas higher frequency (e.g. $\gamma$) the additive oscillatory components together form the broader characteristics of the signal, and thus efforts to produce $\delta$-(1-2.5Hz), $\theta$-(4-7Hz), $\alpha$-(8-11Hz), $\beta$-(13-30Hz), and $\gamma$-(30-60Hz) bands (Gross et al., 2013; Luo and Poeppel, 2007). It is established that lower frequency activity (e.g. $\theta$) detects syllabic quantities through tracking of the speech envelope, whereas higher frequency (e.g. $\gamma$) parses temporally fine units of speech such as phonemes. In addition, the fluctuating $\theta$ phase has been found to modulate the amplitude bursts of the $\gamma$-band by imbuing a rhythmicity to the signal through phase-amplitude coupling (PAC); which enables the coordinated sampling of syllabic and phonological speech items (Hyaflil et al., 2015; Assanoe and Poeppel, 2018; Hermes et al., 2014). In speech production, active sensing - a theorized predictive processing mechanism via a motor sampling of sensory faculties - would suggest 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 $\gamma$-band augmentations have been observed in phonological processing regions during overt and covert phoneme repetition tasks, suggesting a possible common role of $\gamma$ activity to SP (Fukuda et al., 2010; Toyoda et al., 2014). More generally, dorsal stream motor areas have been found to have its own preferred rhythm of speech production (Restle et al., 2012; Assanoe and Poeppel, 2018; Poeppel and Assanoe, 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). Hence, it is increasingly possible that neural oscillations in SP and CS serve similar functions.

However, no studies to date have directly investigated the relative oscillatory contributions and differences in CS and SP. The additive oscillatory components together form the broader characteristics of the signal, and thus efforts to produce models of CS based on signals from SP must first delineate the oscillatory differences across tasks. One such method of understanding how frequency characteristics differ across classes is through a t-test of complex-valued coefficients of a time-frequency transform. This method is referred to as a studentized continuous wavelet transform (t-CWT) and was first introduced by Bostanov (2004) with the intention of improving feature extraction methods for classification in BCI paradigms. Since then, it has been used in the classification of motor imagery (Darvishi and Al-Ani, 2007; Hsu et al., 2007) and the analysis of ERPs in real and simulated EEG data (Real and Kochouhby, 2014). According to the latter study, distinguishing between ERPs with t-CWT produced high specificity and sensitivity under various signal-to-noise ratios in comparison to common peak detection methods. Studentizing time-frequency information in this manner allows for a direct statistical comparison of time-frequency information between two classes in order to detect frequency indices which are significantly different. Furthermore, compared to the discrete wavelet transform or fast Fourier transform, CWT is an ideal candidate for time series analysis due to its more fine-grained resolution and
temporal stability with respect to frequency (Kimata et al., 2018). It is therefore sensible to implement t-CWT with a recording modality with strong temporal resolution such as EEG.

In the present study, we asked whether t-CWT can identify the frequency bands which are used to distinguish words within CS and SP, comparatively. Considering that CS lacks overt vocalization and thus salient self-stimulation, we hypothesized that lower frequency elements would be overshadowed by high frequency activity, such as $\beta$ and $\gamma$. Subsequently, we asked whether the more pertinent oscillations of CS perform similar functions to those of SP, namely by testing for $\theta$-$\gamma$ PAC. However, in the very likely scenario that the $\theta$-band does not play a major linguistically-relevant role in CS, we tested whether CS’s $\gamma$ activity is ‘pseudo-coordinated’ (or pseudo-coupled) to SP’s $\theta$ activity. Such a coupling would indicate that CS’s $\gamma$-band response has a rhythmicity that is related to the putative tracking of syllabic quantities by SP’s $\theta$ activity, thereby asserting that the $\gamma$-band performs a similar function across tasks. Therefore, we hypothesized that the roles of oscillations in SP and CS would be similar and different in the $\gamma$- and $\theta$-bands, respectively. The remainder of the paper is organized as follows: in Section 2, we provide an overview of the previous works on neural oscillation in CS and SP. In Sections 3 and 4, we provide the details of the study methodology and describe the results. In section 5 we discuss the findings of the study and finally conclude our paper in section 6.

2 Neural oscillations in speech perception and covert speech

Over the past two decades, neural oscillations have proven to be a window to understanding a wide variety of cognitive processes. The most important function of neural 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). Indeed, oscillations are thought to provide the most energy-efficient physical mechanism for synchrony and temporal coordination (Miroslav and Strogatz, 1990). In the study of language processing, the induction of such oscillations (e.g. $\delta$, $\theta$, $\alpha$, $\beta$, $\gamma$) contributes to synchronized activities across spatially segregated neuronal assemblies for the coordinated processing of speech units at varying scales (Bastiaansen and Hagoort, 2006; Weiss and Mueller, 2003).

A multitude of studies have investigated the role of oscillations during SP, and a commonly synthesized interpretation from literature is that SP multiplexes neuronal oscillations; that is, SP dynamically samples incoming acoustic information at multiple time scales simultaneously (Gross et al., 2013; Lu and Poeppel, 2007; Ding et al., 2017a; Giraud et al., 2007; Poeppel and Assaneo, 2020). A general rule of thumb is that the higher the frequency of oscillation, the finer the detail to which speech information is sampled. For instance, $\delta$ oscillations (1-2.5Hz) have been implicated in the processing of words, phrases, and sentences (Giraud et al., 2007; Morillon et al., 2012; Doelling et al., 2014; Ding et al., 2013). $\theta$ activity (4-7Hz) has been found to be critically sensitive to speech modulations and to the ongoing speech envelope which contains a 4.5Hz syllabic speech rate (Luo and Poeppel, 2007; Giraud et al., 2007; Ghitza et al., 2013; Doelling et al., 2014). In contrast, high frequency $\gamma$ activity (30-60Hz) has been found to index processing at the phonemic level, as intracortical studies have located a ‘phonotopic’ map of phonemes in regions of the superior temporal gyrus producing differential $\gamma$-band augmentations to phonemes (Chang et al., 2010; Moses et al., 2016; Pasley et al., 2012). Although $\beta$ activity (13-30Hz) has been commonly associated with motor-related potentials in motor imagery studies, in language, this oscillation is thought to be involved in playing a simultaneous role alongside $\theta$ and $\gamma$ activity by conjoining phonological units into a broader syllable by binding the activity of temporally segregated neuronal assemblies (Bastiaansen et al., 2010; Weiss and Mueller, 2003; 2012).

Of these oscillations, the most crucial oscillations seem to be the $\theta$ and $\gamma$ band. Evidence for this comes from numerous observations that the phase and amplitude of these frequency bands 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, $\theta$ activity samples the input spike trains (induced by speech waveform) to generate basic syllable units and time references of speech for subsequent, finer-detailed processing by $\gamma$ activity (Giraud and Poeppel, 2012). The purpose behind this phase-amplitude coupling (PAC) may be to temporally localize $\gamma$ processing power to more descriptive parts of syllabic sound patterns that constitute reference time frames (Hyahl et al., 2015).

Thus, SP multiplexes in relevant frequency bands in order to detect incoming speech and parse them for necessary speech comprehension (Pickering and Garrod, 2013). Although the main oscillatory contributions during SP have been fleshed out, there are vastly fewer studies investigating the role of oscillations in CS. The main feature that distinguishes CS from SP is corollary discharge. Corollary discharge is, in essence, 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 an 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 cancel out self-generated sounds (Okada et al., 2018). Therefore, CS and SP are bridged by a common sensory goal in the auditory domain. This fact invites the question: does CS use oscillations in a similar manner to SP?

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 thought to be responsible for phonological processing (Fukuda et al., 2010; Toyoda et al., 2014). As the purpose of corollary discharge is to match the sensory consequences of self-generated actions, these results invite the hypothesis that the corollary discharge produced during CS, reflected in its γ-band response, may be phonological in nature similar to 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 increase 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 increasingly possible that the θ-band may play an alternative role to syllabic chunking seen in SP.

EEG is an appropriate modality in which to measure and analyze such brain oscillations due to its fast temporal resolution and ease of setup. In speech processing studies, this modality has been frequently used characterize 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 finally oscillatory abnormalities in schizophrenic individuals (Uhlhaas and Singer, 2010; Ford and Mathalon, 2005). Although the spatial resolution for EEG is poor due to volume conduction effects, its strong temporal resolution presents this modality as an optimal medium for tracking fast temporal dynamics of ongoing neural oscillations.

In summary, the findings outlined above demonstrate that certain oscillations may be functionally correlated or divergent in SP and CS. However, no studies have yet determined the relative oscillatory engagements across the two tasks. Hence, the present study used EEG to demonstrate how CS utilizes oscillations relative to SP and whether the most pertinent frequency bands (i.e. θ, γ) perform similar functions across tasks.

3 Methods

3.1 Participants

Ten adults between the ages of 20 and 40 without disabilities or known health conditions were recruited for this study (7 Female, age X +/-Y; 3 Male, age X +/-Y). All participants were right-handed to ensure a consistency in the hemispheric dominance of neurolinguistic processing. Furthermore, all participants were native English speakers (i.e. first language). The research ethics board of the Bloorview Research Institute approved this study. Participants provided informed written consent forms.

Participants donned a 128 electrode ActiCap EEG cap. Of these 128 channels, 64 were utilized (Fig. 1), 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 was sampled at 1000Hz and collected through BrainVision Recorder.

3.2 Experimental procedure

Each participant was seated comfortably approximately 50cm from the computer screen with a refresh rate of 75Hz. 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. 2). First, a blank screen with a duration jitter between 1-2 seconds was presented, followed by a green cross for 2 seconds. During this time, the audio of the speech token (‘Blue’ or ‘Orange’) was presented once. Succeeding this was another blank screen for 2 seconds, followed by a red cross for 2 seconds. Participants were instructed to covertly rehearse the speech token that they had just heard. Therefore, every SP trial was succeeded by a CS trial with the same word. For the rest trial, a letter R cued the participants to refrain from
Figure 1: 64 channel ActiCap wet EEG montage.

Figure 2: The experimental protocol for SP, CS, and rest. SP and rest were preceded with a jitter of 1-2 seconds and followed by either the audio or the symbol, R. In the former case, this was succeeded by a 2 second blank screen and a 2 second window for CS, demarcated by the red cross. In the latter case, the 2 second blank screen was followed by a black cross signalling rest for 5 seconds.
the speech task and fixate on a black cross held on the screen for 5 seconds. Each block consisted of 10 SP-CS trial pairs and 10 rest trials, and each session consisted of 5 blocks. Therefore, each session involved 50 trials/class, and participants underwent two sessions each at least 2 days apart and at roughly the same times of the day. Each session took approximately one hour.

3.3 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 patterns of neural activity that could be suitable for detecting substantial differences in the neural activity associated with each item. The rationale was that these variations would encourage the activation of different motor, somatosensory, and auditory neural representations, thereby enhancing signal discriminability. Speech stimuli were generated by Google Cloud Text-to-Speech platform and presented at an approximate rate of 150 words per minute, which is within the range of the natural speaking rate (Giraud et al., 2007; Luo and Poeppel, 2007). Phoneme models were generated through the Montreal Forced Aligner (McAuliffe et al., 2017).

3.4 Preprocessing

Raw data were analyzed in EEGLAB (Delorme and Makeig, 2004). A 4th order zero-phase high pass Butterworth filter with a cutoff frequency of 1Hz was applied to remove baseline drift. Subsequently, a 4th 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. Eye movement artifacts and muscular artifacts were removed in two separate steps using blind source separation through the EEGLAB plugin ‘Automatic Artifact Removal toolbox’. Following preprocessing, a spline Laplacian was applied to establish local relationships between surface potentials and the underlying source activity (Babiloni et al., 2001). Data were downsampled to 256Hz prior to epoching. Data from sessions 1 and 2 were combined. There were a total of 5 classes: SP Blue (SPB), SP Orange (SPO), CS Blue (CSB), CS Orange (CSO), and rest (RST).

3.5 EEG signal processing

To determine how oscillations drive the distinction of words in CS and SP, we employed a t-CWT routine in order to find the frequency and time indices at which two sets of wavelet coefficients were significantly different. First, each epoched trial was zero-padded with 12 samples in the begining and end. A CWT was conducted on each signal yielding a 55 frequency by 536 time sample matrix (frequencies above 60Hz were removed). CWT was performed using equation (1):

$$ W(s,t) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} f(\tau) \psi(\frac{\tau-t}{s}) d\tau $$

(1)

Where $W(s,t)$ represents the wavelet coefficients, $s$ denotes the scale or frequency, $t$ denotes the time shift, and $\psi$ is the wavelet function which has a zero mean. CWT is thus a sort of template matching computation whereby the cross-covariance between the signal and mother wavelet (here, a Morlet wavelet) is measured by shifting back and forth the latter at dilated and constricted scales. The local extrema of $W(s,t)$ signify the points in frequency and time that are best matched between the signal and template wavelet, and can be visualized in the form of a time-frequency plot, referred to as a scalogram.

To determine which mother wavelet suited the data best, each mother wavelet (Bump, Haar, Morse, Morlet) was used to create wavelet coefficients which were then used to reconstruct the signal via inverse CWT. The correlation between the original and reconstructed signals was calculated by conducting cross-correlation tests, divided by the auto-correlation of the original signal. This analysis yielded a metric that showed how similar the two sets of waveforms are related. It was found that the Morlet wavelet correlated best compared to all other mother wavelets. This is consistent with reports that this wavelet is useful for the detection of salient oscillations (Ende et al., 1998; Senkowski and Herrmann, 2002).

For each classification type and the 58 chosen channels, aggregated two-sample t-tests were conducted on the wavelet coefficients across all trials, yielding 55 frequency x 536 sample t-statistic and H (hypothesis test; 0 or 1) matrices. For each channel, the t-CWT was calculated by:
\[ t^k(s, t) = \frac{W^k_w(s, t) - W^k_y(s, t)}{\sqrt{s_x^2/n + s_y^2/m}} \]  

where \( k \) is channel and \( W^k_w(s, t), W^k_y(s, t), \sigma_x^2(s, t), \) and \( \sigma_y^2(s, t) \) denote sample means and standard deviations of wavelet coefficients across trials at each scale \( s \) and time \( t \), with sample sizes \( n \) and \( m \). Studentizing the wavelet coefficients in this manner enabled the statistical comparison of two classes, describing the frequency and time indices at which they are significantly different. As such, the complex-valued \( t \)-statistic matrices served as time-frequency scalograms with greater magnitudes denoting greater differences. Thus, the absolute values of the \( t \)-statistic matrices were calculated and subsequently normalized to determine the regional maxima, with the condition that the maxima must be located within the cone of influence, but importantly, where \( H=1 \); i.e. where there is a significant difference between classes (Fig. 3). Detecting maxima only within the cone of influence mitigated the risk of detecting artifactual maxima in the scalograms. Each channel produced a different amount of regional maxima.

![Figure 3: Magnitude scalogram of a \( t \)-statistic. Red stars denote the time and scale indices of the peaks of the magnitude scalogram. Peaks were chosen under the criteria that they would exist inside the cone of influence, and importantly, where \( H=1 \) (i.e. significant difference).](image)

### 3.6 Feature extraction and classification

Wavelet coefficients of the CWT output were extracted at the frequency and time indices obtained through t-CWT. These values were extracted for each of the 58 channels and appended into a complex feature matrix. Subsequently, the real and imaginary values of these complex wavelet coefficients were obtained and appended side by side to form a matrix of trials x features, with the last column representing the class identity. For each participant, the original feature matrix was 100 trials x 500 features approximately. However, Minimally Redundant Maximally Relevant (mRmR) feature selection was conducted to select the top 20 features, which was the approximate cutoff point for feature importance during mRmR. Subsequently, a 10-fold cross validation was conducted on the 100 trials x 20 features matrices and subsequently classified through a support vector machine (SVM) with a radial basis function kernel. All classification types were binary.

### 3.7 Performance evaluation

Average classification accuracies and their standard deviations were obtained using a 10 fold cross-validation. Precision, Recall, and F1-score were calculated for each of these folds and subsequently averaged.

\[ Precision = \frac{TruePositives}{TruePositives + FalsePositives} \]  

7
Recall = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalseNegatives}} \quad (4)

F1 - score = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)

3.8 Measuring phase-amplitude coupling (PAC)

The importance of PAC is well documented in speech processing studies (Giraud and Poeppel, 2012; Hyafil et al., 2015; Assaneo and Poeppel, 2018; Voytek et al., 2013; Hermes et al., 2014). To determine significant $\theta$-$\gamma$ coupling in CS, the event-related phase-amplitude coupling (ERPAC) toolbox was utilized (Voytek et al., 2013). A 4th-order Butterworth filter was applied between 4-7Hz to obtain $\theta$-band signals, after which the angle of the Hilbert transform was taken to obtain phase data. The $\gamma$-band was calculated by filtering the signals between 30-60Hz, with a Butterworth filter order of 3*r where r is the sampling rate divided by the low frequency cutoff of the filter, rounded. The $\gamma$-band amplitude was obtained by taking the absolute value of the Hilbert transform. In a specific channel, PAC for each phase-amplitude pair was calculated through the circ_corrcl.m function from the CircStat toolbox (Berens, 2009) at each timepoint and across all trials. This function calculates the correlation coefficient ($\rho$) between a circular/angular ($\phi, c, s$) and linear random variable (such as amplitude - a) by linearizing the phase variable into sin and cosine components:

$$\rho_{\phi a} = \sqrt{\frac{r_{ca}^2 + r_{sa}^2 - 2r_{ca}r_{sa}r_{cs}}{1 - r_{cs}^2}}$$ \quad (6)

where $r_{ca} = \text{corr}(\cos\phi[t], a[t]), r_{sa} = \text{corr}(\sin\phi[t], a[n]), r_{cs} = \text{corr}(\sin\phi[t], \cos\phi[t]),$ and $\text{corr}(x, y)$ is the Pearson correlation between $x$ and $y$ with the assumption that the distribution of $x$ and $y$ are Gaussian. $\phi[t]$ and $a[t]$ are the instantaneous phase and instantaneous analytic amplitude, respectively. Utilizing this function enabled the assessment of relationships between circular $\theta$ phase and linear $\gamma$ amplitude at each time point and across trials.

Subsequently, 1000 surrogate runs were conducted by shifting the trials of the amplitude data and testing correlation between the phase data and shifted amplitude data across trials at each time point. These PAC $\rho$ values were compared by first applying Fisher’s z-transform to normalize the correlation coefficients:

$$z_{rt} = \frac{1}{2} \ln \left( \frac{1 + \rho_t}{1 - \rho_t} \right)$$ \quad (7)

and calculating the difference between z-transformed coefficients:

$$\Delta \rho_z = z(\rho_{\text{true}}) - z(\rho_{\text{surrogate}})$$ \quad (8)

From this, the z-score can be calculated by:

$$z = \frac{\Delta \rho_z}{\sigma}$$ \quad (9)

where $\sigma$ is standard error. z-scores were then transformed into $p$-values via a normal cumulative distribution function with $\mu = 0$, $\sigma = 1$. The reported $p$-values denote the time points at which there are significant $\theta$-$\gamma$ PAC occurring against a surrogate population within a specific channel. For cross-task PAC (e.g. SP $\theta$-CS $\gamma$), the class indices used for amplitude calculation was simply switched to a CS class.

3.9 Statistical analysis

For the t-CWT results, the frequency indices were tabulated across all channels and participants and subsequently categorized into the five major bands. Shapiro-Wilks tests were conducted on each category to confirm non-normality. Subsequently, Wilcoxon Rank Sum tests were conducted for unequal medians.

To test for significant correlations between the $\theta$ phases of SP and CS classes, the circ_corrc.m function from the CircStat toolbox was invoked (Berens, 2009). This function assesses the correlation between two circular/angular random signals:
\[ \rho_{\alpha\beta} = \frac{\sum_i \sin(\alpha_i - \bar{\alpha})\sin(\beta_i - \bar{\beta})}{\sqrt{\sum_i \sin^2(\alpha_i - \bar{\alpha})\sin^2(\beta_i - \bar{\beta})}} \]  

(10)

where \( \alpha \) and \( \beta \) denote two samples of angular data and \( \bar{\alpha} \) and \( \bar{\beta} \) denote their means. Under the null hypothesis of no significant correlations, the \( p \)-value to this correlation was computed by a normally distributed test statistic. This enabled the testing of angular correlation between two sets of SP \( \theta \) phase across all trials of participants (1000 trials).

4 Results

4.1 Task-dependent utilization of oscillations for distinction of words

| Error Type | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |
|------------|----|----|----|----|----|----|----|----|----|----|
| SPO-CSO    | Ave.| 85.00(6.2) | 75.5(11.3) | 80.0(9.4) | 85.0(3.3) | 80.0(5.4) | 85.0(3.2) | 85.0(2.9) | 80.0(3.3) | 85.0(3.3) |
|            | Rst.| 85.00(6.2) | 75.5(11.3) | 80.0(9.4) | 85.0(3.3) | 80.0(5.4) | 85.0(3.2) | 85.0(2.9) | 80.0(3.3) | 85.0(3.3) |
| CSB-CSO    | Ave.| 85.00(6.2) | 75.5(11.3) | 80.0(9.4) | 85.0(3.3) | 80.0(5.4) | 85.0(3.2) | 85.0(2.9) | 80.0(3.3) | 85.0(3.3) |
|            | Rst.| 85.00(6.2) | 75.5(11.3) | 80.0(9.4) | 85.0(3.3) | 80.0(5.4) | 85.0(3.2) | 85.0(2.9) | 80.0(3.3) | 85.0(3.3) |
| SPB-CSB    | Ave.| 85.00(6.2) | 75.5(11.3) | 80.0(9.4) | 85.0(3.3) | 80.0(5.4) | 85.0(3.2) | 85.0(2.9) | 80.0(3.3) | 85.0(3.3) |
|            | Rst.| 85.00(6.2) | 75.5(11.3) | 80.0(9.4) | 85.0(3.3) | 80.0(5.4) | 85.0(3.2) | 85.0(2.9) | 80.0(3.3) | 85.0(3.3) |
| SPO-CSO    | Ave.| 85.00(6.2) | 75.5(11.3) | 80.0(9.4) | 85.0(3.3) | 80.0(5.4) | 85.0(3.2) | 85.0(2.9) | 80.0(3.3) | 85.0(3.3) |
|            | Rst.| 85.00(6.2) | 75.5(11.3) | 80.0(9.4) | 85.0(3.3) | 80.0(5.4) | 85.0(3.2) | 85.0(2.9) | 80.0(3.3) | 85.0(3.3) |
| SPB-RST    | Ave.| 85.00(6.2) | 75.5(11.3) | 80.0(9.4) | 85.0(3.3) | 80.0(5.4) | 85.0(3.2) | 85.0(2.9) | 80.0(3.3) | 85.0(3.3) |
|            | Rst.| 85.00(6.2) | 75.5(11.3) | 80.0(9.4) | 85.0(3.3) | 80.0(5.4) | 85.0(3.2) | 85.0(2.9) | 80.0(3.3) | 85.0(3.3) |
| CSB-RST    | Ave.| 85.00(6.2) | 75.5(11.3) | 80.0(9.4) | 85.0(3.3) | 80.0(5.4) | 85.0(3.2) | 85.0(2.9) | 80.0(3.3) | 85.0(3.3) |
|            | Rst.| 85.00(6.2) | 75.5(11.3) | 80.0(9.4) | 85.0(3.3) | 80.0(5.4) | 85.0(3.2) | 85.0(2.9) | 80.0(3.3) | 85.0(3.3) |

Table 1: Classification scores for each classification type and for each participant. Post-mRmR feature selection features were classified using a SVM with a radial basis function kernel. Classification accuracies were calculated by averaging across the 10 cross validation folds (standard deviation in brackets). Precision, recall, and F1-scores were calculated for each fold and subsequently averaged.

Extracting the real and imaginary values of imaginary coefficients at these indices produced significantly higher binary SVM classification accuracies than chance level (Table 1) in most participants with the exception of P2 and P7. Similarly, these features were found to have high precision, recall, and F1-score across most participants. The two participants who seemed to perform relatively poorly had greater standard deviations across classifications and lower performance scores.

Assessing the frequency indices at which opposing classes are significantly different (cumulative over all 58 channels and participants) revealed significant differences in oscillatory characteristics between SP and CS (Fig. 4). Consistent with previous studies (Gross et al. 2013, Luo and Poeppel 2007), SP multiplexed in all relevant frequency bands such as \( \delta \), \( \theta \), \( \beta \), and \( \gamma \). After confirming non-normality of distribution of frequencies through Shapiro-Wilk tests \((p<0.05)\), Wilcoxon Rank Sum tests revealed that distinguishing SP classes engages significantly more of lower frequency \( \delta \) and \( \theta \) \((p<0.01, p<0.0001)\), whereas distinguishing CS involves more usage of low \( \gamma \) \((p<0.05)\) (Fig. 4b). No significant differences in \( \alpha \) and \( \beta \) were observed \((p>0.01)\). Distinguishing between corresponding SP and CS classes (e.g. Blue) showed that SP and CS classes differ similarly across words \((p>0.05)\) (Fig. 2). The distinction of active classes from rest compared between CS and SP revealed significantly higher involvement of \( \delta \) and \( \theta \) bands in SP \((p<0.05, p<0.01)\), whereas CSO vs RST produced significantly higher \( \beta \) involvement than SPO vs RST \((p<0.05)\) (Fig. 4d).

To determine the topography of oscillatory differences, EEGLAB’s (Delorme and Makeig 2004) topoplot.m function was used on features selected through mRmR feature selection. Distinctions in channel locations were tabulated (cumulative across participants) for each frequency band (Fig. 5). SP showed relatively greater inter-participant consistency in the \( \theta \)-band, with widespread distinctions across temporal and temporoparietal regions and high counts in this frequency (Fig. 5b). SP’s \( \gamma \) distinctions were focally distributed with lower consistencies across participants (Fig. 5c).
Figure 4: Differential utilization of frequency bands in covert speech and speech perception. Each frequency and time index was determined first by conducting a two-sample t-test against two sets of wavelet coefficients. Maxima of the t-statistic were used as indices provided that the hypothesis test returned a statistically significant difference. Data were pooled across all channels and participants. A Shapiro-Wilks test was conducted between each binary classification pair (e.g. SPB vs SPO) to confirm non-normality. Subsequently, a Wilcoxon Rank Sum test was conducted to test for significantly different medians. The distinction of SP classes produced a greater engagement of low frequency δ and θ, whereas CS involved more low γ activity for distinction (a). No significant differences were observed when comparing binary classifications involving corresponding SP and CS classes (b). δ and θ activity contributed significantly more to distinguishing SP from rest than CS from rest (c, d). β activity contributed significantly more to distinction of CSO vs RST than SPO vs RST (d). (*p<0.05; **p<0.01; ***p<0.001; ns-no significance).
Figure 5: Topography of frequency indices (a-d) time indices (e) of speech perception and covert speech tasks. (a-d) Calculated from 20 features selected through mRmR feature selection as a tabulated sum across participants. Figures were generated through EEGLAB. (Delorme and Makeig, 2004). Scale shows the number of tabulated distinctions in the frequency band. Brighter colors mean higher count. The distinction of CS classes produces involves higher frequencies whereas the distinction of SP is largely based on lower frequencies. (e) Time indices were calculated by taking the median of time values across participants. Scales show the time in seconds. Brighter and darker colors mean high- and low-latency, respectively. SP and CS both produced low latency distinctions in the temporal and temporo-parietal regions, with high latency activations in the frontal/motor regions.

In contrast, CS produced focal θ-band distinctions with low count/consistency and widespread γ-band distinctions with high amount of consistency across participants in temporal and temporo-parietal regions. Both tasks produced comparable amount of β-band distinctions, but CS produced more counts of distinguishable β patterns in the right hemisphere (Fig. 5c). CS had minimal δ activity, whereas SP showed three foci of δ activity in the left hemisphere (Fig. 5a).

4.2 Task-dependent coupling of Theta phase to Gamma amplitude

Event-related spectral perturbations (ERSPs) revealed strong transient synchronization in the θ band for CS and SP between 200-500ms (Fig. 6a, b). For both tasks, this was succeeded by β-band synchronization starting at the offset of θ-band synchronization. γ-band desynchronizations were observed for both tasks between 200-500ms, but more scattered for CS. Rest showed scattered and unorderly synchronizations in the α and β bands (Fig. 6c).

Considering the putative coordination between θ and γ band activity (Giraud and Poeppel, 2012; Hyafil et al., 2015), PAC between these frequency bands were assessed across all classes (Fig. 7). Interestingly, single-channel PACs were observed specifically in only the right hemisphere temporal channels. Namely, in channel FT10, the PAC of both SP classes significantly departed from the surrogate PAC between 200-500ms (p<0.01), confirming that SP γ amplitude produces a rhythm in keeping with the cadence of fluctuating θ phase. 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 function 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
Figure 6: Event related spectral perturbation in perception (a), covert speech (b), and rest tasks (c) in channel T8. Figures were generated using EEGLAB’s `spectopo.m` function (Delorme and Makeig, 2004) and data aggregated from participants 1-5. Results are FDR-corrected at $p<0.05$. Frequency units are logarithmically spaced. The ERSP for SP (a) and CS (b) both show early latency synchronization in the $\theta$ band, followed by $\beta$-band synchronization. Rest showed a scattered synchronization of $\alpha$ and $\beta$-bands.

SP’s $\theta$ and CS’s $\gamma$ was assessed to determine whether the $\gamma$ amplitude of CS contains a task-specific rhythm (Fig. 8a, b). Significant pseudo-PAC ($p<0.01$) occurred again between 200-500ms for both words, confirming that CS’s $\gamma$-band produced a rhythmic fluctuation specific to the time course of SP’s $\theta$ synchronization (Fig. 6a). Furthermore, relatively less but nevertheless significant pseudo-PACs were observed across words (e.g. SPB-CSO), meaning that the SP-\text{-}CS $\gamma$ relationship contained both general and specific portions. This lack of specificity between SP $\theta$ and CS $\gamma$ was found to be due to significantly correlated $\theta$ phase patterns in SP of the two words between 200-500ms (circular correlation test $p<0.01$) (Fig. 9a). On the other hand, CS did not produce correlative $\theta$ phase patterns across words (Fig. 9b) and across tasks (Fig. 9b).

However, this did not portend that the $\gamma$-band responses of CS and SP words were general, as significant cross-trial $\gamma$ power correlations (Pearson correlation $p<0.05$) were observed across tasks. Such correlations were observed in left temporal and temporo-parietal regions, the right fronto-temporal edge, and along the right motor to somatosensory regions (Fig. 10), consistent to that seen in the topography of frequencies (Fig. 5). Importantly, significant $\gamma$-band power correlations were observed in channel FT10, where SP $\theta$-$\gamma$ and SP $\theta$-CS $\gamma$ PACs were observed.
Figure 7: Speech perception produces significant $\theta$ phase-$\gamma$ amplitude coupling between 200-500ms. Calculated through ERPAC toolbox (Voytek et al., 2013). To determine whether arising PAC relationships index an inter-trial relationship and not an artifact of stimulus-evoked responses, we conducted a resampling analysis (surrogate testing) that randomizes the phase-amplitude relationship across trials. This was done 1000 times per sample and resulted in a distribution of possible surrogate PACs. SP produced significant PAC relationships occurring between 200-500ms, while CS and rest produced relatively little and sparse PAC between $\theta$ phase and low $\gamma$ amplitude. Dotted lines indicate significant difference of true PAC from surrogates ($p<0.05$). Figures were generated from participant 5.
Figure 8: Speech perception $\theta$ phase predicts covert speech low $\gamma$ amplitude between 200-500ms. PACs were calculated across trials and per sample of the time series after a Hilbert transform using the ERPAC toolbox (Voytek et al., 2013). Phase information was extracted only on SP classes and amplitude information was extracted only on the corresponding CS class (a,b). Statistical significance was calculated through surrogate testing. a and b show that SP’s $\theta$ phase is significantly ‘pseudo-coupled’ to CS’s low $\gamma$ amplitude compared to surrogate PACs between 200-500ms, with a similar coupling morphology to SP PAC. Red dotted lines indicate significance difference ($p < 0.01$) for blue lines.

Figure 9: $\theta$ activity is task dependent and serves different functions in speech perception and covert speech. $\theta$ phase of each signal was determined by calculating the angle of the Hilbert transform of the signal after a Butterworth band pass filter between 4-7Hz. The circular mean of the phase angles were calculated across all channels. Subsequently, the circular correlation between SPB and SPO $\theta$ phase was calculated for each time point and across trials. Data were pooled across all participants. $\theta$ phase was correlated in the two SP classes (a). Significant ($p < 0.01$) correlations were observed between 200-500ms for SPB-SPO, whereas CSB-CSO produced no significant phase correlations. Furthermore, no relationship was observed between the theta phases of SP and CS (b). Grey line depicts significant circular correlations of theta phase between SPB and SPO. Red and blue lines depict the SP-CS pairs which produced no significant $\theta$ phase correlations.

Figure 10: Topography of cross-trial $\gamma$ power correlations between corresponding classes. Data were pooled across all participants and sessions. $\gamma$ amplitude was extracted first through a Butterworth filter band-passed between 30-60Hz on preprocessed signals and then taking the absolute value of the Hilbert transform of said signals. Cross-trial $\gamma$ power was calculated by squaring the amplitude at each point in time, summing across trials, and normalizing by number of trials to result in a single time vector of power. s the Pearson’s correlation coefficient was extracted for each channel between two opposing classes (e.g. SPB-CSB). High values (bright regions) correspond to high $\gamma$ power correlation across tasks (Pearson’s rho). Only correlations with $p < 0.05$ are depicted.
5 Discussion

Using a t-CWT method, the present study confirmed the hypothesis that CS largely utilizes higher frequency oscillations relative to SP. Crucially, we conclude that the $\gamma$-band likely functions similarly across tasks. Like SP, the $\gamma$ activity of CS was found to contain a processing rhythm time-locked to the cadence of event-related SP $\theta$ phase. Specifically, we suggest that CS’s $\gamma$ activity likely depicts a phonological processing function similar to SP reported here and in previous studies (Giraud and Poeppel, 2012; Pasley et al., 2012; Chang et al., 2010, 2016). However, the lack of $\theta$-$\gamma$ PAC within CS suggests that CS’s $\theta$ activity likely serves alternative roles to syllabic chunking seen in SP, possibly in preparatory motor or memory-related activity (Restle et al., 2012; Albouy et al., 2017). The present study represents the first investigation into the differences in oscillatory engagements between CS and SP and reports a relationship between SP’s $\theta$ and CS’s $\gamma$ activity. The findings reported here could enable the development of CS models based on SP signals, which can be used to train a CS BCI based on the passive perception of speech.

5.1 Oscillations in speech perception and covert speech

The main goal of the study was to assess the relative contribution and roles of major oscillatory dynamics 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 have identified that SP indeed utilizes the $\delta$, $\theta$, $\beta$, and $\gamma$ frequency bands, with significantly more prominent distinctions in the $\delta$ and $\theta$ bands than CS. This prominence 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., 2017a; Varnet et al., 2017). Moreover, $\theta$ activity may produce generalized patterns during SP, as the frequency topography showed widespread and consistent distinctions across participants in the temporal and temporo-parietal channels.

The relatively high count of $\gamma$-band distinctions suggests that the driving force of delineating words in SP lies primarily in phonological processes (Chang et al., 2010). These activations were found focally in left and right temporal and temporo-parietal regions, which may loosely correspond to phonological regions of interests as reported through intracranial recording studies (Chang et al., 2010, 2016; Pasley et al., 2012). Moreover, the significant $\theta$-$\gamma$ PAC between 200-500ms likely depicts a linkage between syllabic and phonological processes (Giraud and Poeppel, 2012), suggesting that the two oscillations may be inherently coordinated to focus $\gamma$ activity to specific times within syllabic time references upheld by $\theta$ (Hyafil et al., 2015).

The timing of this PAC was likely linked to strong $\theta$-band synchronization occurring in this time period, as observed through ERSP. Interestingly, $\beta$-band synchronization followed shortly after. The $\beta$-band, producing a comparable amount of distinctions, has been proposed to play a role in binding of semantic (Weiss and Mueller, 2003) and syntactic information (Bastaansen et al., 2010), being sensitive to the temporal alignment of ongoing speech (Rimmele et al., 2018). The observation of enhanced $\beta$-synchronization succeeding the $\theta$-$\gamma$ PAC can suggest that coordinated phonological and syllabic sampling becomes bound/conjoined into whole-word percepts via $\beta$ activity during SP.

CS, on the other hand, did not exhibit a diverse multiplexing relationship akin to SP, but rather favoured high frequency activity, namely $\beta$ and $\gamma$, for the distinction of words. The low count and sparse distribution of $\theta$ activity was likely due to lack of salient percepts in CS, suggesting that, unlike SP, $\theta$ activity in CS is focal and may serve a dissimilar function (discussed in Section 5.3). On the other hand, the high amount of distinctions in the $\gamma$-band suggested that this frequency is highly specific to word identity. Indeed, studies of overt and covert phoneme/word repetition tasks have shown differential $\gamma$-band augmentations in the temporal and temporo-parietal lobe (Fukuda et al., 2010; Peri et al., 2011; Toyoda et al., 2014), which may loosely correspond to the frequency topography reported here. What the frequency topography does firmly suggest is that the widespread and high count of $\beta$- and $\gamma$-band distinctions likely indicated a greater degree of inter-participant consistency and can, in turn, allude to the existence of generalized activation profiles in temporal and temporo-parietal channels during CS.

It is possible that the $\gamma$-band activity during CS corresponded to a corollary discharge, as enhanced fronto-temporal $\gamma$ 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 $\gamma$ 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 (discussed in Section 5.2). Similarly, it can follow that the observed β-band synchronization (occurring at the same time as SP) also signifies the temporal binding routines of phonological speech units enacted by the γ-band (Section 5.4).

5.2 Common role of Gamma activity

It was observed that CS and SP engage their oscillations differentially during speech processing. Interestingly, γ distinctions were found to be the highest within tasks. As previously mentioned, a multitude of studies describe a phonological processing function of γ activity during SP (Chang et al., 2010, 2016; Pasley et al., 2012). Although the methods of the present study were not sensitive to determining phonological cognitive load, we specifically asked whether CS’s γ activity contained a processing rhythm specific to θ activity. SP’s γ activity has been shown to keep 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, but not proof, for the hypothesis that γ activity in CS serves a similar function to that in SP. Therefore, confirming this γ rhythm would resolve an important step toward modelling CS from SP through a demonstration of possible functional equivalence.

However, a 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, task-dependent θ phase correlations were not observed within CS (Fig. 4). Thus, we asked if CS γ would produce a rhythm that corresponds to the task-event-related cadence of SP’s θ activity. Indeed, the results of the current study support this hypothesis with significant SP-γ CS PAC also occurring at 200-500ms. 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 result confirms that CS’s γ activity is rhythmic, and importantly, specific to the periodicity of SP’s θ phase, which putatively tracks syllabic quantities through the stimulus envelope (Luo and Poeppel, 2007). Therefore, the observed cross-task pseudo-coupling supports the idea that the γ-bands of SP and CS served similar functions.

The notion that this observed time-localized rhythmicity of CS γ activity corresponds to a phonological processing function like SP may be loosely entertained by the frequency topography, which shows that CS words are distinguished in the γ band more consistently in the temporal and temporo-parietal channels. Intracranial studies employing overt and covert phoneme repetition tasks also report γ activity in these regions, but specifically in the superior temporal gyrus and supramarginal gyrus (Fukuda et al., 2010; Toyoda et al., 2014), regions which have previously shown to play a role in phonological processing through fMRI investigations (Okada and Hickok, 2006; van de Ven et al., 2009; Venezia et al., 2016). Although it is tempting to connect the present topographical results to the source localization in these studies, a word of caution is warranted as EEG is known to have poor spatial resolution and activations portrayed by the scalp map may not project ideally to the putative sources of speech processing.

In contrast to the loose functional correspondence depicted by the topographical results, the finding of γ rhythmicity in SP and CS in the same time frame (200-500ms) substantiates the interpretation that γ activity served a similar function across tasks. 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 priorly. 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). It remains inconclusive whether the present θ-band synchronization in CS was a result of phase resetting (Luo and Poeppel, 2007) or greater evoked potentials (Obleser and Weisz, 2012). However, if like overt speech, CS tracks self-generated and temporally regular speech through neural oscillations, then the θ phase would necessarily reset to cause enhanced synchronization across trials (Luo and Poeppel, 2007).

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 likely demarcated the points of processing in the tasks, indicating that CS and SP process words at the same time. Naturally, it then follows that the common occurrences of γ rhythms, both of which are modulated and pseudo-modulated by SP’s θ phase, represented a similar function between SP and CS. Since CS is a variant of speech production (only lacking overt articulation and production of sounds), it must generate a timely internal auditory prediction to match the processing of self-generated speech sounds (Jack et al., 2019; Scott, 2013).
Therefore, under the view that CS is equivalent to self-generated SP without feedback, we propose that CS’s $\gamma$-band response may have represented a similar function, potentially relating to phonological processing.

While the above discussion supports the idea that the $\gamma$-bands subserved similar functions across tasks, it may be further reasoned that $\gamma$ activity during CS represents a ‘mirrored phonological’ activation pattern, as previously pondered. This hypothesis emerged out of studies of AVH whereby an aberrant corollary discharge (i.e. $\gamma$ synchrony) results in phantom perceptions (Mathalon and Ford, 2008, van Lutterveld et al., 2011, Ford and Mathalon, 2009). As the purpose of corollary discharge is to cancel out self-generated sounds, it follows that CS’s $\gamma$-band response must continuously predict the sound patterns of ongoing speech. This invites the hypothesis that CS’s $\gamma$ activity may produce similar activation patterns to that of SP’s $\gamma$-band response, upon some transformation. As the current study did not analyze any existing correlations between the $\gamma$ activities, future studies are directed to employ distance correlation measures to quantify the predictability and/or dependence between the $\gamma$-band amplitudes of SP and CS.

However, some caution is warranted to the above interpretations as the current study involved only two speech tokens. Hence, future studies are directed to design studies with more diverse arrays of speech tokens for understanding the relationship between SP and CS’s $\theta$- and $\gamma$-band responses.

5.3 Differential role of Theta activity

It was found that SP and CS produced rhythmic fluctuations of $\gamma$-band activity that correlated to the tracking of the speech envelope by SP’s $\theta$ activity. However, the lack of $\theta$-$\gamma$ PAC occurring within CS suggested that the $\gamma$ band response of CS retained its own processing rhythm in the absence of modulation by its $\theta$ phase. This observation is corroborated by a lack of $\theta$ phase correlations occurring between SP and CS (Fig. 8). Thus, it is possible that $\theta$ activity served a dissimilar function to syllabic chunking as seen in SP, or that the relationship between $\theta$ and $\gamma$ in CS cannot be described by a coupling of phase and amplitude. Similar to the current results, Hermes et al. (2014) showed that $\theta$-$\gamma$ PAC is suppressed during CS in Broca’s area, the tempo-parietal junction, and middle temporal gyrus, and that its $\theta$ power is anti-correlated to high frequency power. Perhaps more counter-intuitively, $\theta$-$\gamma$ PAC has been reported to increase in patients during AVH as thoughts manifest as phantom perception (Koutsoukos et al., 2013); the inverse of which suggests that a normal thought would suppress this $\theta$-$\gamma$ coupling. These results beg the question: if CS’s $\theta$ band synchronizes in the same time period (Fig. 6b) as the emergence of its $\gamma$ rhythm, what is the role of $\theta$ with respect to $\gamma$ in CS?

Considering that stimulation of a major dorsal stream area (posterior inferior frontal gyrus) with $\theta$ burst stimulation facilitates speech repetition accuracy (Restle et al., 2012), it may be reasoned that $\theta$ 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 $\theta$-band represents a preferred articulatory rhythm and thus has motor origins in speech production. If this interpretation holds, it would suggest that phonological and articulatory processing in CS are dependent in the context of PAC, but potentially related by some other measure. For instance, it may be possible that $\theta$-based articulatory expressions induces $\gamma$-based corollary discharge that pre-contains the sensory predictions outlined by the motor code and rules out the need for coupling between the two frequency bands. Indeed, $\theta$ coherence (Ford et al., 2002) and $\gamma$ synchrony (Uhlhaas et al., 2006, Uhlhaas and Singer, 2010, Gallinat et al., 2004) has both been found to be significantly reduced in schizophrenic patients with AVH, suggesting that the independent suppression of synchronization in the two oscillations each plays a role in an aberrant corollary discharge mechanism. Ameliorating the current results with the findings of these studies can lead to the theory that $\theta$-based motor discharges may inform $\gamma$ activity in CS, but not be linked through a PAC.

Alternatively, or perhaps in parallel, $\theta$ oscillations in the dorsal stream may also work by enhancing auditory working memory (Albouy et al., 2017), potentially in the form of access to lexical stores (Piat et al., 2014). This is consistent with the model proposed by Indefrey and Levelt (2004) where word production has been suggested to initiate with the lexical concept. More empirically, in a series of studies investigating oscillatory power during covert word reading (Bastiaansen et al., 2005) and lexical decision-making (Bastiaansen et al., 2008), $\theta$ power - local synchrony was found to be modulated as a function of lexicality, peaking between 300-500ms. Thus, $\theta$ activity as accessing the mental lexicon is a sensible interpretation, as instantiating a lexical memory can initiate subsequent unitary/phonological processing by the $\gamma$-band, but not be necessarily linked through phase and amplitude. Indeed, $\gamma$-band activity has been suggested to be necessary for the formation of both phonological and lexico-semantological representations of words through repetition and homophone priming tasks (Matsumoto and Iidaka, 2008). It is thus possible that $\theta$ activity registers the broader lexical framework for the auditory prediction and informs $\gamma$-band corollary discharge, which seemingly provides the sensory/phonological representation of the word. Although the singular role of $\theta$ activity in speech production is being debated, considering that slow oscillations can synchronize between widely distributed brain
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.

5.4 General discussion and Limitations

The processing of CS brought about a comparable amount of distinctions in the $\beta$-band as the $\gamma$-band. While $\beta$ activity for speech production has been reported to play a role in motor activity and motor preparation (Mersov et al., 2016; Piai et al., 2015), in language-related processes, it has importantly been proposed to serve as a top-down modulatory signal for the temporal management of ongoing speech (Rimmele et al., 2018). Indeed, in language processing, $\beta$ activity has been shown to play a role in the timely binding of syntactic (Bastaansen et al., 2010) and semantic (Weiss and Mueller, 2003) information. More broadly, Weiss and Mueller (2012) assert that synchronized $\beta$-band oscillations serve to bind the contents of distributed set of neuronal populations into one coherent memory unit. The role of $\beta$ activity in temporal binding routines is supported by studies revealing significant $\delta$-$\beta$ PAC (Arnal et al., 2015; Keitel et al., 2017; 2018; Morillon et al., 2019), for the registration of words, phrases, and sentences by $\delta$ must necessarily emerge from the temporal bindings along the hierarchy of speech units (phonemes to words). Although the current study did not investigate this coupling, the observation of temporally co-localized $\beta$-band synchronizations between CS and SP, succeeding the common $\theta$-band synchronization (discussed to signify the processing of SP and CS), invites the hypothesis that the $\beta$-band, too, served a similar function across tasks, namely in enacting binding routines of phonological/syllabic items into broader whole-word percepts.

Finally, it should be noted that the phase-amplitude relationships between SP’s $\theta$ and CS’s $\gamma$ were found to be general as weaker, but nevertheless existent, pseudo-couplings were observed also across words (e.g. SPB $\theta$-CSO $\gamma$). However, this does not necessarily portend that the $\gamma$ activities themselves are general, as no gamma-band correlations were observed. Indeed, CS and SP both produced a significant amount of distinctions in the $\gamma$-band, suggesting that $\gamma$ activity is likely specific to words. Instead, general nature of SP-CS PAC can be attributed to the lack of diversity in the current lexicon, which varied only between 1-2 syllables and spoken at the same rate. This evidently led to non-divergent $\theta$ patterns: average $\theta$ phases were found to be significantly correlated between 200-500 in the two SP classes, making this frequency band less pertinent to distinction of SP words than $\gamma$ activity (Fig. 9a). In contrast, studies using a larger vocabulary and sentential speech tokens have shown that $\theta$ phase adjusts to syllabic rate and the number of syllables (Assaneo and Poeppel, 2018; Lizarazu et al., 2019; Ding et al., 2017a,b). Therefore, future studies should experiment with a richer lexicon with more syllable counts, possibly embedded in sentential forms, in order to determine whether CS’s $\gamma$ activity forms a specific relationship to the putative syllabic tracking by SP’s $\theta$ activity. Such studies will assist in determining whether this pseudo-coupling between SP and CS simply reflects task demands or reflects common neurolinguistic processing. Furthermore, future studies are directed to determine whether correlations exist between the $\gamma$-band responses of SP and CS. These investigations should provide substantial supports for the hypothesis that CS $\gamma$ activity also reflects a phonological process.

6 Conclusion

The present study represents the first to investigate the similarities and differences with respect to oscillatory engagement during CS and SP. We found that CS favours higher frequency activity that likely reflects corollary discharge. Specifically, we found that CS’s $\gamma$-band response has a similar rhythmic pattern to SP, possibly representing a similar phonological process. These findings substantiate the results of Oppenheim and Dell (2008) who describe CS to contain robust phonological information, and further suggests that CS and SP may share a common function when it comes to $\gamma$ activity. Contrarily, we assert that $\theta$ activity in CS and SP play different roles possibly via differential processing through dorsal and ventral streams, respectively. Understanding the relative oscillatory engagements and their functional correlates in the two tasks are elemental to the modelling of CS based on SP signals. Therefore, the present work can lead to the development of CS BCIs through the passive perception of speech, which can help hurdle the difficulties of training by rendering the training process passive. In order to achieve this modelling, we direct future studies to investigate the details of the relationship between SP’s $\theta$ activity and CS’s $\gamma$ activity, as well as similarities in the $\gamma$-band responses for further confirmation of a common function of $\gamma$ activity.

7 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.
References

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

Alderson-Day, B. and Fernyhough, C. (2012). Inner Speech: Development, Cognitive Functions, Phenomenology, and Neurobiology Ben. *Cirugia Espanola*, 90(9):545–547.

Alderson-Day, B., Mitrenga, K., Wilkinson, S., McCarthy-Jones, S., and Fernyhough, C. (2018). 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.

Arnal, L. H., Doelling, K. B., and Poeppel, D. (2015). Delta-beta coupled oscillations underlie temporal prediction accuracy. *Cerebral Cortex*, 25(9):3077–3085.

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

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

Bastiaansen, M. and Hagoort, P. (2006). Chapter 12 Oscillatory neuronal dynamics during language comprehension. *Progress in Brain Research*, 159(06):179–196.

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

Bastiaansen, M. C., Oostenveld, R., Jensen, O., and Hagoort, P. (2008). I see what you mean: Theta power increases are involved in the retrieval of lexical semantic information. *Brain and Language*, 106(1):15–28.

Bastiaansen, M. C., Van Der Linden, M., Ter Keurs, M., Dijkstra, T., and Hagoort, P. (2005). Theta responses are involved in lexical-semantic retrieval during language processing. *Journal of Cognitive Neuroscience*, 17(3):530–541.

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

Bidelman, G. M. (2015). Induced neural beta oscillations predict categorical speech perception abilities. *Brain and Language*, 141:62–69.

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

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

Bostanov, V. (2004). 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.

Buchsbaum, B. R., Hickok, G., and Humphries, C. (2001). Cognitive Science : A Multidisciplinary Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science*, 25(78437590):663–678.

Buzsáki, G. and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304(5679):1926–1929.

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

Canavier, C. C. (2015). Phase-resetting as a tool of information transmission. *Current Opinion in Neurobiology*, 31:206–213.

Chang, C. K., Chiari, L., and Hutchison, D. (2016). *Inclusive Smart Cities*.

Chang, M. D., Sejadić, E., Wright, V., and Chau, T. (2010). Measures of dynamic stability: Detecting differences between walking overground and on a compliant surface. *Human Movement Science*, 29(6):977–986.

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

Cullen, K. E. (2004). Sensory signals during active versus passive movement. *Current Opinion in Neurobiology*, 14(6):698–706.

Darvishi, S. and Al-Ani, A. (2007). 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.
DaSalla, C., Kambara, H., Koike, Y., and Sato, M. (2009). Spatial filtering and single-trial classification of EEG during vowel speech imagery. *International Convention on Rehabilitation Engineering and Assistive Technology (ICREAT)*, 5:1–4.

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

Deng, S., Srinivasan, R., Lappas, T., and D’Zmura, M. (2010). EEG classification of imagined syllable rhythm using Hilbert spectrum methods. *Journal of Neural Engineering*, 7(4).

Di Liberto, G. M., O’Sullivan, J. A., and Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology*, 25(19):2457–2465.

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

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

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

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

Ende, M., Louis, A. K., Maass, P., and Mayer-Kress, G. (1998). EEG Signal Analysis by Continuous Wavelet Transform Techniques. In *Nonlinear Analysis of Physiological Data*, number 1, pages 213–219.

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

Ford, J. M. and Mathalon, D. H. (2019). Effference Copy, Corollary Discharge, Predictive Coding, and Psychosis. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 4(9):764–767.

Ford, J. M., Mathalon, D. H., Whitfield, S., Faustman, W. O., and Roth, W. T. (2002). Reduced communication between frontal and temporal lobes during talking in schizophrenia. *Biological Psychiatry*, 51(6):485–492.

Fries, P. (2009). Neuronal Gamma-Band Synchronization as a Fundamental Process in Cortical Computation. *Annual Review of Neuroscience*, 32(1):209–224.

Fukuda, M., Rothermel, R., Juhász, C., Nishida, M., Sood, S., and Asano, E. (2010). Cortical gamma-oscillations modulated by listening and overt repetition of phonemes. *NeuroImage*, 49(3):2735–2745.

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

Ghitza, O. (2012). 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.

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

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

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

Giraud, A. L., and Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15(4):511–517.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., and Garrod, S. (2013). Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain. *PLoS Biology*, 11(12).

Hermes, D., Miller, K. J., Vansteensel, M. J., Edwards, E., Ferrier, C. H., Bleichner, M. G., van Rijen, P. C., Aarnoutse, E. J., and Ramsey, N. F. (2014). Cortical theta wanes for language. *NeuroImage*, 85:738–748.

Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. *Language, Cognition and Neuroscience*, 29(1):2–20.
Hickok, G., Houde, J., and Rong, F. (2011). Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. *Neuron*, 69(3):407–422.

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

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

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

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

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

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

Indefrey, P. and Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1-2):101–144.

Jack, B. N., Le Pelley, M. E., Han, N., Harris, A. W., Spencer, K. M., and Whitford, T. J. (2019). Inner speech is accompanied by a temporally-precise and content-specific corollary discharge. *NeuroImage*, 198(March):170–180.

Keitel, A., Gross, J., and Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS Biology*, 16(3):1–19.

Keitel, A., Ince, R. A., Gross, J., and Kayser, C. (2017). Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. *NeuroImage*, 147(November 2016):32–42.

Kimata, A., Yokoyama, Y., Aita, S., Nakamura, H., Higuchi, K., Tanaka, Y., Nomami, A., Hirao, K., and Aonuma, K. (2018). Temporally stable frequency mapping using continuous wavelet transform analysis in patients with persistent atrial fibrillation. *Journal of Cardiovascular Electrophysiology*, 29(4):514–522.

Koutsoukos, E., Angelopoulos, E., Maillis, A., Papadimitriou, G. N., and Stefanis, C. (2013). Indication of increased phase coupling between theta and gamma EEG rhythms associated with the experience of auditory verbal hallucinations. *Neuroscience Letters*, 534(1):242–245.

Lizarazu, M., Lallier, M., and Molinaro, N. (2019). Phase amplitude coupling between theta and gamma oscillations adapts to speech rate. *Annals of the New York Academy of Sciences*, (April).

Luo, H. and Poeppel, D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron*, 54(6):1001–1010.

Mai, G., Minett, J. W., and Wang, W. S. (2016). Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing. *NeuroImage*, 133.

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

Matsumoto, A. and Idaka, T. (2008). Gamma band synchronization and the formation of representations in visual word processing: Evidence from repetition and homophone priming. *Journal of Cognitive Neuroscience*, 20(11):2088–2096.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., and Sonderegger, M. (2017). 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.

Mersov, A. M., Jobst, C., Cheyne, D. O., and De Nil, L. (2016). Sensorimotor oscillations prior to speech onset reflect altered motor networks in adults who stutter. *Frontiers in Human Neuroscience*, 10(SEP2016):1–16.

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

Morillon, B., Arnal, L. H., Schroeder, C. E., and Keitel, A. (2019). 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.

Morillon, B., Liégeois-Chauvel, C., Arnal, L. H., Bénar, C. G., and Giraud, A. L. (2012). Asymmetric function of theta and gamma activity in syllable processing: An intra-cortical study. *Frontiers in Psychology*, 3(JUL):1–9.
Morillon, B. and Schroeder, C. E. (2015). Neuronal oscillations as a mechanistic substrate of auditory temporal prediction. *Annals of the New York Academy of Sciences*, 1337(1):26–31.

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

Morin, A., Uttl, B., and Hamper, B. (2011). Self-reported frequency, content, and functions of inner speech. *Procedia - Social and Behavioral Sciences*, 30:1714–1718.

Moses, D. A., Mesgarani, N., Leonard, M. K., and Chang, E. F. (2016). Neural speech recognition: Continuous phoneme decoding using spatiotemporal representations of human cortical activity. *Journal of Neural Engineering*, 13(5):1–19.

Obleser, J. and Weisz, N. (2012). Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. *Cerebral Cortex*, 22(11):2466–2477.

Okada, K. and Hickok, G. (2006). 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.

Okada, K., Matchin, W., and Hickok, G. (2018). Neural evidence for predictive coding in auditory cortex during speech production. *Psychonomic Bulletin and Review*, 25(1):423–430.

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

Oppenheim, G. M. and Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, 106(1):528–537.

Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., Knight, R. T., and Chang, E. F. (2012). Reconstructing speech from human auditory cortex. *PLoS Biology*, 10(1).

Pei, X., Leuthardt, E. C., Gaona, C. M., Brunner, P., Wolpaw, J. R., and Schalk, G. (2011). Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition. *NeuroImage*, 54(4):2960–2972.

Perrone-Bertolotti, M., Rapin, L., Lachaux, J. P., Baciu, M., and Levenbruck, H. (2014). 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.

Piai, V., Dahlslettt, K., and Maris, E. (2015). Statistically comparing EEG/MEG waveforms through successive significant univariate tests: How bad can it be? *Psychophysiology*, 52(3):440–443.

Piai, V., Roelofs, A., and Maris, E. (2014). Oscillatory brain responses in spoken word production reflect lexical frequency and sentential constraint. *Neuropsychologia*, 53(1):146–156.

Pickering, M. J. and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4):329–347.

Poeppel, D. (2014). The neuroanatomic and neurophysiological infrastructure for speech and language. *Current Opinion in Neurobiology*, 28:142–149.

Poeppel, D. and Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, 21(6):322–334.

Real, R. G. and Kotchoubey, B. (2014). Studentized continuous wavelet transform (t-CWT) in the analysis of individual ERPs: Real and simulated EEG data. *Frontiers in Neuroscience*, 8(Sep):1–9.

Rrestle, J., Murakami, T., and Ziemann, U. (2012). Facilitation of speech repetition accuracy by theta burst stimulation of the left posterior inferior frontal gyrus. *Neuropsychologia*, 50(8):2026–2031.

Rimmele, J. M., Morillon, B., Poeppel, D., and Arnal, L. H. (2018). Proactive Sensing of Periodic and Aperiodic Auditory Patterns. *Trends in Cognitive Sciences*, 22(10):870–882.

Roberts, M. J., Lowet, E., Brunet, N. M., TerWal, M., Tiesinga, P., Fries, P., and DeWeerd, P. (2013). Robust gamma coherence between macaque V1 and V2 by dynamic frequency matching. *Neuron*, 78(3):523–536.

Schroeder, C. E. and Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, 32(1):9–18.

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

Scott, S. K. (2012). 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.
Senkowski, D. and Herrmann, C. S. (2002). Effects of task difficulty on evoked gamma activity and ERPs in a visual discrimination task. *Clinical Neurophysiology*, 113(11):1742–1753.

Shahin, A. J., Picton, T. W., and Miller, L. M. (2009). Brain oscillations during semantic evaluation of speech. *Brain and Cognition*, 70(3):259–266.

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

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

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

Tian, X. and Poeppel, D. (2012). Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. *Frontiers in Human Neuroscience*, 6(November):1–11.

Toyoda, G., Brown, E. C., Matsuzaki, N., Kojima, K., Nishida, M., and Asano, E. (2014). Electrocorticographic correlates of overt articulation of 44 English phonemes: Intracranial recording in children with focal epilepsy. *Clinical Neurophysiology*, 125(6):1129–1137.

Trouvain, J. (2007). On the comprehension of extremely fast synthetic speech. pages 5–13.

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

Uhlhaas, P. J. and Singer, W. (2010). Abnormal neural oscillations and synchrony in schizophrenia. *Nature Reviews Neuroscience*, 11(2):100–113.

Van de Ven, V., Esposito, F., and Christoffels, I. K. (2009). 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.

Van Lutterveld, R., Sommer, I. E. C., and Ford, J. M. (2011). The Neurophysiology of Auditory Hallucinations – A Historical and Contemporary Review. *Frontiers in Psychiatry*, 2(May):1–7.

Varnet, L., Ortiz-Barajas, M. C., Erra, R. G., Gervain, J., and Lorenzi, C. (2017). A cross-linguistic study of speech modulation spectra. *The Journal of the Acoustical Society of America*, 142(4):1976–1989.

Venezia, J. H., Fillmore, P., Matchin, W., Lisette Isenberg, A., Hickok, G., and Fridriksson, J. (2016). Perception drives production across sensory modalities: A network for sensorimotor integration of visual speech. *NeuroImage*, 126:196–207.

Volo, B. and Womelsdorf, T. (2016). 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.

Voytek, B., Esposito, M. D., Crone, N., and Knight, R. T. (2013). A method for event-related phase / amplitude coupling. *NeuroImage*, 64:416–424.

Weiss, S. and Mueller, H. M. (2003). The contribution of EEG coherence to the investigation of language. *Brain and Language*, 85(2):325–343.

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

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

Yao, B., Taylor, J. R., Banks, B., and Kotz, S. A. (2020). Theta activity phase-locks to inner speech in silent reading. *PsyArXiv*, 44(0).

Zoefel, B. and VanRullen, R. (2016). EEG oscillations entrain their phase to high-level features of speech sound. *NeuroImage*, 124:16–23.