2021

MULTIVARIATE ANALYSIS FOR UNDERSTANDING COGNITIVE SPEECH PROCESSING

MD SULTAN MAHMUD

MULTIVARIATE ANALYSIS FOR UNDERSTANDING COGNITIVE SPEECH PROCESSING

by

Md Sultan Mahmud

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Engineering

The University of Memphis

August 2021

To Mom and Dad for your true, unconditional, and never-ending love and support.

Acknowledgments

First, I am extremely grateful to my adviser, Dr. Mohammed Yeasin, for his continuous support and invaluable mentoring. I would like to express my deepest appreciation to my committee members Dr. Gavin M. Bidelman, Dr. Madhusudhanan Balasubramanian, and Dr. Eugene C. Eckstein for their valuable suggestions and directions. The completion of my dissertation would not have been possible without the support and encouragement of my parents and family members.

Special thanks to funding organizations specifically the Electrical and Computer Engineering Department; The University of Memphis; National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under award number NIH/NIDCD R01DC016267.

I thank my lab mates at CVPIA for being such good friends and helping me whenever I needed it specially Iftekhar Anam, Faruk Ahmed, Rakib Al-Fahad, Shahinur Alam, Kazi Ashraf Moinuddin, Felix Havugimana, Sultana Razia Akhter, Gaurav Singh, Adel Bany Muhammad, Haitham Alnajdwi, and Fatin Ishrak. And lastly, I want to express my gratitude to Auditory Cognitive and Neuroscience Laboratory (ACNL) and 50 anonymous volunteers, participating in our study, that without them none of this could have happened.

Abstract

Categorical perception (CP) of audio is critical to understand how the human brain perceives speech sounds despite widespread variability in acoustic properties. Most studies that examine cognitive speech processing have applied methodological approaches that focus on specific or a set of brain regions. Departing from hypothesis-driven approaches, in this dissertation, we proposed multivariate data-driven approaches to identify the spatiotemporal (i.e., *when in time* and *where in the brain*) and spectral (i.e., frequency-band power levels) characteristics of auditory neural activity that reflects CP for speech (i.e., differentiates phonetic prototypes from ambiguous speech sounds). We recorded 64-channel EEG as listeners rapidly classified vowel sounds along an acoustic-phonetic continuum. We used parameter optimized support vector machine (SVM), k-nearest neighbors (KNN) classifiers, and stability selection to determine spatiotemporal and spectral characteristics of neural activity that decode CP best via source-level neural activity. Using event-related potentials (ERPs), we found that early (120 ms) whole-brain data decoded speech categories (i.e., prototypical vs. ambiguous speech tokens) with 95.16% accuracy [area under the curve (AUC) 95.14%; F1-score 95.00%]. Separate analyses on the left hemisphere (LH) and right hemisphere (RH) responses showed that LH decoding was more accurate and earlier than RH (89.03% vs. 86.45% accuracy; 140 ms vs. 200 ms). Stability (feature) selection identified 13 regions of interest (ROIs) out of 68 brain regions (including auditory cortex, supramarginal gyrus, and inferior frontal gyrus (IFG)) that showed categorical representation during stimulus encoding (0-260 ms). In contrast, 15 ROIs (including fronto-parietal regions, IFG, motor cortex) were necessary to describe later decision stages (later 300 to 800 ms) of categorization but these areas were highly associated with the strength of listeners' categorical hearing (i.e., slope of behavioral identification functions).

Moreover, our induced vs. evoked mode analysis shows that whole-brain evoked β-band activity decoded prototypical from ambiguous speech sounds with ~70% accuracy. However, induced γ-band oscillations showed better decoding of speech categories with ~95% accuracy compared to evoked β-band activity (~70% accuracy). Induced high frequency (γ-band) oscillations dominated CP decoding in the LH, whereas lower frequency (θ-band) dominated decoding in the RH. In addition, feature selection identified 14 brain regions carrying induced activity and 22 regions of evoked activity that were most salient in describing category-level speech representations. Among the areas and neural regimes explored, we found that induced γ-band modulations were most strongly associated with listeners' behavioral CP.

In sum, our data-driven multivariate models demonstrate that abstract categories emerge surprisingly early (~120 ms) in the time course of speech processing and are dominated by engagement of a relatively compact fronto-temporal-parietal brain network. In addition, the category-level organization of speech is dominated by relatively high frequency induced brain rhythms.

Table of Contents

List of Tables

List of Figures

Chapter 1 - Introduction

The human brain can map an incredibly large number of stimulus features into a smaller set of groups (Chang et al., 2010; Holt & Lotto, 2010), a process known as categorical perception (CP). CP provides information about how speech sounds are perceived by humans despite the wide variability of acoustics properties. It plays a critical role in speech perception and language processing. For understanding speech processing, it is crucial to know the spatiotemporal (e.g., *where in the brain, when in time*) and spectral (e.g., which *mode of brain oscillations and frequency bands*) characteristics of the neural activities.

Electroencephalography (EEG) and magnetoencephalography (MEG) are commonly used non-invasive modalities for recording neural activities. EEG has high temporal resolution and low cost. EEG activity can be divided into "*evoked*" and "*induced*" responses.

In this dissertation, our main goal was to develop data-driven multivariate frameworks for understanding the neural mechanism underlying speech categorization. Particularly, we examined the spatiotemporal and spectral characteristics of neural activity while younger-adult categorized pure vowels (true phonetic categories) vs. ambiguous speech sounds (lacking a clear phonetic identity). Machine learning (ML) is a branch of artificial intelligence that *"learns a model"* from the past data to predict future data (Cruz & Wishart, 2006). Moreover, data mining approaches in ML identify important properties in neural activity with high accuracy without intervention from human observers. Furthermore, ML can predict and identify subtle changes in neural activity very accurately and quickly, without intervention from human observers. It would be meaningful if speech categorization could be decoded from neural data without or with minimal *a priori* assumptions. By extending prior hypothesis-driven work on the speech categorization, here, we have conducted an entirely different, comprehensive data-driven approach to test whether prototypical vowel vs. ambiguous speech can be decoded from full-brain activity (e.g., event-related potentials (ERPs) and spectral features). We aimed to identify the most probable global set of brain

regions that are associated with speech categorization using ML. To our knowledge, this is the first study to apply decoding and ML techniques to map spatiotemporal and spectral analysis in speech categorization in younger-adults listeners at the full-brain level.

1.1 Research aims

Aim 1: Decoding categorical speech perception (CP) in spatiotemporal neural processing from evoked brain responses (e.g., ERPs)

We have endeavored to explore the spatiotemporal analysis of CP inspired by our previous spatiotemporal analysis on aging data (Mahmud et al., 2020). We have hypothesized that speech-evoked responses (e.g., ERPs) would differ with regards to time and spatial regions that are recruited during the categorization of speech stimuli (e.g., *pure vowels* vs. *ambiguous speech*). We have investigated how well *when in time* and *where in the brain* ERP features could decode prototypical vowels from ambiguous speech sounds of an acoustic-phonetic continuum. In addition, we explored which brain hemisphere (e.g., left hemisphere (LH) or right hemisphere (RH)) is dominant in speech categorization. For instance, how speech categorization varies in latency using hemisphere data. We have proposed the development of a robust framework to decode prototypical vowel speech versus ambiguous speech sounds from source-level neural data (e.g., ERPs).

We further hypothesized that the core speech network for "encoding" and "decoding" via ML vary as a function of speech processing. We have aimed to explore what are the brain ROIs that engage in CP [e.g., prototypical vowel speech (Tk1/Tk5) vs. ambiguous speech (Tk3)]. Here we have used a similar approach in our previous study (Mahmud et al., 2020) on normal hearing vs. mild hearing-impaired older adults' brain network associated with age-related hearing loss but delved deeper into the analysis. We have explored brain networks that involve *early sensory* (i.e., speech encoding ~250 ms) and *post perceptual* (i.e., decision-process > 300 ms) time windows.

Aim 2: Speech categorization from induced and evoked neural oscillations.

Evoked activity is time and phase-locked and more related to stimuli. On the other hand, induced oscillations reflect important information about cognitive processing (David et al., 2006). Some researchers use induced activity for brain function and dysfunction studies. For instance, magnetoencephalography (MEG) studies demonstrated that oscillatory brain activity differs while perceived language vs. non-language stimuli (Eulitz et al., 1996); the segmentation and coding of continuous speech processing rely on cortical oscillations (Gross et al., 2013). Certain frequency bands of EEGs have been linked with specific neurocognitive and language processing (Giraud & Poeppel, 2012; Von Stein & Sarnthein, 2000). For example, studies (Youssofzadeh et al., 2020) showed that beta power decrement within the language processing areas and dominance in LH during auditory task processing. Our recent study (Mahmud et al., 2021) demonstrated that different brain regions are associated with the encoding vs. decision stages of processing while categorizing speech. These studies demonstrate that temporal dynamics of evoked activity provide a neural correlate of the different processes underlying speech categorization. However, ERP studies do not reveal how induced brain activity (so-called neural oscillations) might contribute to this process.

Here, we were interested in exploring how does "*induced*" or "*evoked*" oscillation relate to CP during the categorization of speech sounds [e.g., prototype vowel *(Tk1 /u/* or *Tk 5 /a/*) vs. ambiguous speech (e.g., *Tk3*)] by using spectral features (e.g., *evoked* vs. *induced activity's* power spectral density (PSD) of different frequency bands). The following questions we have aimed to examine from *evoked* vs. *induced* neural activities analysis:

- How does the categorization of speech affect on "*evoked*" or "*induced*" modes of brain processing during prototypical vowels vs. ambiguous speech sound perception?
- Which spectral profiles are associated with CP and correlate with behavioral measures (e.g., behavioral slope) from whole-brain data?

- How does the individual frequency bands of each hemisphere (i.e., LH or RH) dominate during the categorization of speech sounds?

1.2 Main results

Our spatiotemporal temporal analysis shows that early (120 ms) ERPs of whole-brain data decoded speech categories (i.e., prototypical vs. ambiguous speech tokens) with 95.16% accuracy [area under the curve (AUC) 95.14%; F1-score 95.00%]. We also found the following results.

1. The whole-brain ERPs data showed better speech categorization than single hemisphere data.
2. Individual hemispheres (e.g., LH and RH) analysis using ERP feature showed that LH data yielded more accurate and earlier decoding than RH (89.03% vs. 86.45% accuracy: 140 ms vs. 200 ms).
3. A smaller brain network is involved during encoding as compared to the decision process.
4. Brain areas associated with the decision process are highly linked with the strength of listeners' categorical hearing (i.e., slope of behavioral identification functions).

Our spectral analysis using "*induced*" vs. "*evoked*" responses demonstrates that induced brain oscillation could categorize the speech sound better than evoked activity. Particularly, the induced gamma frequency band is the best decoding ability of CP among all other frequency bands. Our results corroborate previous theoretical studies by supporting that induced activity can better predict speech categorization. Additionally, we also inferred the following.

1. Induced gamma activity could categorize speech best among all other frequency bands.

2. The evoked activity of the beta frequency band could categorize the speech sound the best among all evoked frequency bands.
3. Six brain regions' gamma activity predict the strength of listeners' categorical hearing 91.1%.

1.3 Broader impacts and novelty

The studies reported here reflect an interdisciplinary blend of engineering and neuroscience. The outcomes of this study identify the spatiotemporal and spectral characteristics of neural data in speech categorization ability (i.e., behavioral slope). In spatiotemporal analysis, we developed a comprehensive data-driven computational framework for decoding speech categorization using ERPs. In particular, we developed network descriptions of speech encoding and the decision process.

Furthermore, the spectral analysis showed that the induced mode of brain oscillation decodes speech categorization better than the evoked activity. Particularly, the induced gamma activity predicts the behavioral slope 91.1%. Apart from clinical implications, better speech categorization (i.e., classification) of pure vowels vs. ambiguous speech is likely to provide a breakthrough in the understanding of neural mechanisms that underlie speech categorization.

Chapter 2 - Research Context

Most of the studies on cognitive speech processing have applied methodological approaches that focus on specific or selected sets of brain regions. However, some complex cognitive processing emerges in a large-scale brain-network that supports multiple functions, including cognitive/language, attention, and motor control. How can this large-scale brain–network be identified from whole-brain data rather than specific hypothesis-driven. We proposed comprehensive data drive approaches to identify spatiotemporal and spectral characteristics of neural data while younger adults categorize the speech sounds (i.e., prototypical vowel vs. ambiguous vowel).

2.1 Spatiotemporal characteristics of speech categorization

Categories allow listeners to extract, manipulate, and precisely respond to sounds (C. T. Miller & Cohen, 2010; E. K. Miller et al., 2002, 2003; Russ et al., 2007; Tsunada & Cohen, 2014) despite wide variability in their acoustic properties. CP emerges in early life (Eimas et al., 1971) but is further modified by native language experience (Bidelman & Lee, 2015; Kuhl et al., 1992; Xu et al., 2006). As such, CP plays an important role in understanding receptive communication and the building blocks of speech perception and language processing across the lifespan.

ERPs are particularly useful for examining the brain mechanisms of phoneme and speech perception (Celsis et al., 1999; Molfese et al., 2005) given their excellent temporal resolution and the rapid time course required to process speech signals. Indeed, several neuroimaging studies have documented neural correlates to CP via ERPs (Bidelman, 2015; Chang et al., 2010; Shen & Froud, 2019). In particular, several studies have shown that the efficiency of listeners' speech categorization varies in accordance with their underlying brain activity (Bidelman et al., 2013; Bidelman & Alain, 2015b; Bidelman & Lee, 2015; Perlovsky, 2011). For example, Bidelman et al. demonstrated that brain responses in the time frame of 180-320 ms were more robust for phonetic prototypes vs. ambiguous speech tokens, thereby

reflecting category-level processing (Bidelman et al., 2020). Other studies have shown links between N1-P2 amplitudes of the auditory cortical ERPs and the strength of listeners' speech identification (Bidelman & Walker, 2017) and labeling speeds (Al-Fahad et al., 2020) in speech categorization tasks (Bidelman et al., 2014; Bidelman & Alain, 2015b). These findings are consistent with the notion that the early N1 and P2 waves of the ERPs are highly sensitive to speech processing and auditory object formation that is necessary to map sounds to meaning (Alain, 2007; Bidelman et al., 2013; Wood et al., 1971).

The neural organization of speech categories also varies spatially, recruiting a widely distributed system across a number of brain regions. Neural responses are elicited by prototypical speech sounds (i.e., those heard with a strong phonetic category) differentially engage Heschl's gyrus (HG) and inferior frontal gyrus (IFG) compared to ambiguous speech depending on a listeners perceptual skill level (Bidelman et al., 2013; Bidelman & Lee, 2015; Bidelman & Walker, 2017; Mankel et al., 2020). This suggests emergent categorical representations within the early auditory-linguistic pathways. Similarly, Alho et al. found that category-specific representations were activated in left IFG (Alho et al., 2016) at an early-latency (115-140 ms). Collectively, in terms of the time course of processing, M/EEG studies agree that the neural underpinnings of speech categories emerge within the first few hundred milliseconds after stimulus onset and reflect abstract "category level-effects" (Toscano et al., 2018) and "phonemic categorization" (Liebenthal et al., 2010).

Beyond conventional auditory-linguistic brain regions, neuroimaging also demonstrates a variety of additional areas important to speech perception and language processing (Hickok et al., 2011; Lee et al., 2012; Novick et al., 2010). Among them, the superior parietal lobe is associated with writing (Menon & Desmond, 2001) and supramarginal gyrus with phonological processing (Deschamps et al., 2014; Oberhuber et al., 2016) during speech and verbal working memory tasks. Relevant to CP, several studies have found that the left inferior parietal lobe is more activated during auditory phoneme sound

categorization (Desai et al., 2008; Dufor et al., 2007; Husain et al., 2006). Indeed, auditory categorical processing has been shown to recruit superior temporal gyrus/sulcus, middle temporal gyrus, premotor cortex, inferior parietal cortex, planum temporal, and inferior frontal gyrus (Bidelman & Walker, 2019; Guenther et al., 2004). Some other neuroimaging and electrocorticography studies have however shown that rostral anterior cingulate cortex is associated with speech control (Paus et al., 1993; Sahin et al., 2009; Tankus et al., 2012) and the orbitofrontal cortex in speech comprehension (Sabri et al., 2008). Under some circumstances (e.g., highly skilled listeners), speech categories can emerge as early as auditory cortex (Bidelman & Lee, 2015; Bidelman & Walker, 2019; Chang et al., 2010).

While category representations seem to emerge early in the time course of speech perception, the task of categorizing sounds can be further separated into pre- and post-perceptual stages of processing (i.e., stimulus encoding vs. decision mechanisms). "Early" vs. "late" stage models of category formation have long been discussed in the literature (Fox, 1984; McClelland & Elman, 1986; Noe & Fischer-Baum, 2020; Norris et al., 2000). However, few empirical studies have actually separately examined the encoding and decision stages of CP. The human brain encodes speech stimuli within ~250 ms after stimulus onset (Masmoudi et al., 2012) and decodes ~300 ms after stimulus onset (Domenech & Dreher, 2010; Mostert et al., 2015). Previous studies have largely focused on these specific time windows (e.g., ERP waves) and brain regions when attempting to describe the neural basis of CP. While informative, such hypothesis-based testing can be restrictive and potentially may miss the broader and distributed networks associated with speech-language processing that unfold on different time scales (Du et al., 2016; Rauschecker & Scott, 2009).

In this regard, ML techniques are increasingly used to "decode" high-dimensional neuroimaging data and better understand different states of brain functionality as measured via EEG. It would be meaningful if brain functioning that has been linked with speech processing (e.g., CP) could be decoded from neural data (e.g., ERPs) without, or at least with

minimal, *a priori* assumptions on when and where those representations emerge. Indeed, laying the groundwork for the present work, we have recently shown that the speed of listeners' identification in speech categorization tasks can be directly decoded from their full-brain EEGs using an entirely data-drive approach (Al-Fahad et al., 2020). We have also shown that ML can decode age-related changes in speech processing that occur in older adults (Mahmud et al., 2020).

Departing from previous hypothesis-driven studies (Bidelman & Alain, 2015b; Bidelman & Walker, 2019, 2017), the current work used a comprehensive, data-driven approach to examine the neural mechanisms of speech categorization during encoding and decision stages of processing using whole-brain, electrophysiological data. We analyzed cortical speech-evoked ERPs from 64-channel scalp EEG recorded during a rapid speech categorization task in young, normal hearing listeners. Our approach applied state-of-the-art ML techniques including neural classifiers and feature selection methods (i.e., stability selection) to source-level ERPs to investigate the spatiotemporal dynamics of speech categorization. We aimed to determine when (i.e., in time) and where (i.e., brain ROIs) neural activity from full-brain EEGs differentiated phonetic from phonetically ambiguous speech sounds, and thus showed the strongest evidence of categorical processing using an entirely data-driven, machine learning approach.

2.2 Evoked vs. induced analysis for speech categorization

The electroencephalogram (EEG) can be divided into evoked (i.e., phase-locked) and induced (i.e., non-phase locked) responses that vary in a frequency-specific manner (Shahin et al., 2009). Evoked responses are largely related to the stimulus, whereas induced responses are additionally linked to different perceptual and cognitive processes that emerge during task engagement. These later brain oscillations (neural rhythms) play an important role in perceptual and cognitive processes and reflect different aspects of speech perception. For example, low frequency [e.g., θ (4-8 Hz)] bands are associated with syllable segmentation

(Luo & Poeppel, 2012) whereas α (9-13 Hz) band has been linked with attention (Klimesch, 2012) and speech intelligibility (Dimitrijevic et al., 2017). Several studies report listeners' speech categorization efficiency varies in accordance with their underlying induced and evoked neural activity (Bidelman et al., 2013; Bidelman & Alain, 2015a; Bidelman & Lee, 2015). For instance, Bidelman assessed correlations between ongoing neural activity (e.g., induced activity) and the slopes of listeners' identification functions, reflecting the strength of their CP (Bidelman, 2017). Listeners were slower and varied in their classification of more category-ambiguous speech sounds, which covaried with increases in induced γ activity (Bidelman, 2017) in memory (Bashivan et al., 2014), whereas the higher γ frequency range (>30 Hz) is associated with auditory object construction (Tallon-Baudry & Bertrand, 1999) and local network synchronization (Giraud & Poeppel, 2012; Haenschel et al., 2000; Si et al., 2017).

Studies also demonstrate hemispheric asymmetries in neural oscillations. During syllable processing, there is a dominance of γ frequency activity in LH and θ frequency activity in RH (Giraud et al., 2007; Morillon et al., 2012). Other studies show that during speech perception and production, lower frequency bands (3-6 Hz) better correlate with behavioral reaction times than higher frequencies (20-50 Hz) (Yellamsetty & Bidelman, 2018). Moreover, induced γ-band correlates with speech discrimination and perceptual computations during acoustic encoding (Ou & Law, 2018), further suggesting it reflects a neural representation of speech above and beyond evoked activity alone.

Still, given the high dimensionality of EEG data, it remains unclear which frequency bands, brain regions, and "modes" of neural function (i.e., evoked vs. induced signaling) are most conducive to describing the neurobiology of speech categorization. To this end, the recent application of ML to neuroscience data might prove useful in identifying the most salient spectral features of brain activity that predict human behaviors.

Our goals were to evaluate which neural regime [i.e., evoked (phase-synchronized ERP) vs. induced oscillations], frequency bands, and brain regions are most associated with CP using whole-brain activity via a data-driven approach (i.e., SVM, KNN classifiers, and stability selection). Based on prior work, we hypothesized that evoked and induced brain responses would both differentiate the degree to which speech sounds carry category-level information (i.e., prototypical vs. ambiguous sounds from an acoustic-phonetic continuum). However, we predicted induced activity would best distinguish category-level speech representations, suggesting a dominance of endogenous brain rhythms in describing the neural underpinnings of CP.

Chapter 3 - Methods and Materials

In this chapter, we discuss data collection, methods, and materials that are used in this dissertation. Our multivariate data-driven frameworks robustly decode speech sounds (e.g., prototypical vowel vs. ambiguous) from the neural data. This study demonstrates that how the neural mechanism underlying speech categorization.

3.1 Participants

Forty-nine young adults (15 male, 34 female; aged 18 to 33 years) were recruited as participants from the University of Memphis student body to participate into our ongoing studies on the neural basis of speech perception and auditory categorization (Bidelman et al., 2020; Bidelman & Walker, 2017; Mankel et al., 2020). All participants had normal hearing sensitivity (i.e., <25 dB HL between 500-8000 Hz). Listeners were-right handed, native English speakers, and had achieved a collegiate level of education (Oldfield, 1971). None reported any history of neurological disease. All participants were paid for their time and gave informed written consent in accordance with the declaration of Helsinki and a protocol approved by the Institutional Review Board at the University of Memphis.

3.2 Stimulus & task

We used a synthetic five-step vowel token continuum to assess the most discriminating spatiotemporal features while categorizing prototypical vowel speech from ambiguous speech (Bidelman et al., 2013, 2014). Speech spectrograms are represented in Figure 1A. Each token of the continuum was separated by equidistant steps acoustically based on the first formant frequency (F1) and perceived categorically from /u/ to /a/. Each speech token was 100 ms, including 10 ms rise/fall to minimize the spectral splatter in the stimuli. Each speech token contained an identical voice fundamental frequency (F0), second (F2), and third formant (F3) frequencies (F0:150 Hz, F2: 1090 Hz, and F3:2350 Hz). To create a phonetic continuum that varied in percept from /u/ to /a/, F1 frequency was parameterized over five equal steps from 430 Hz to 730 Hz (Bidelman et al., 2013).

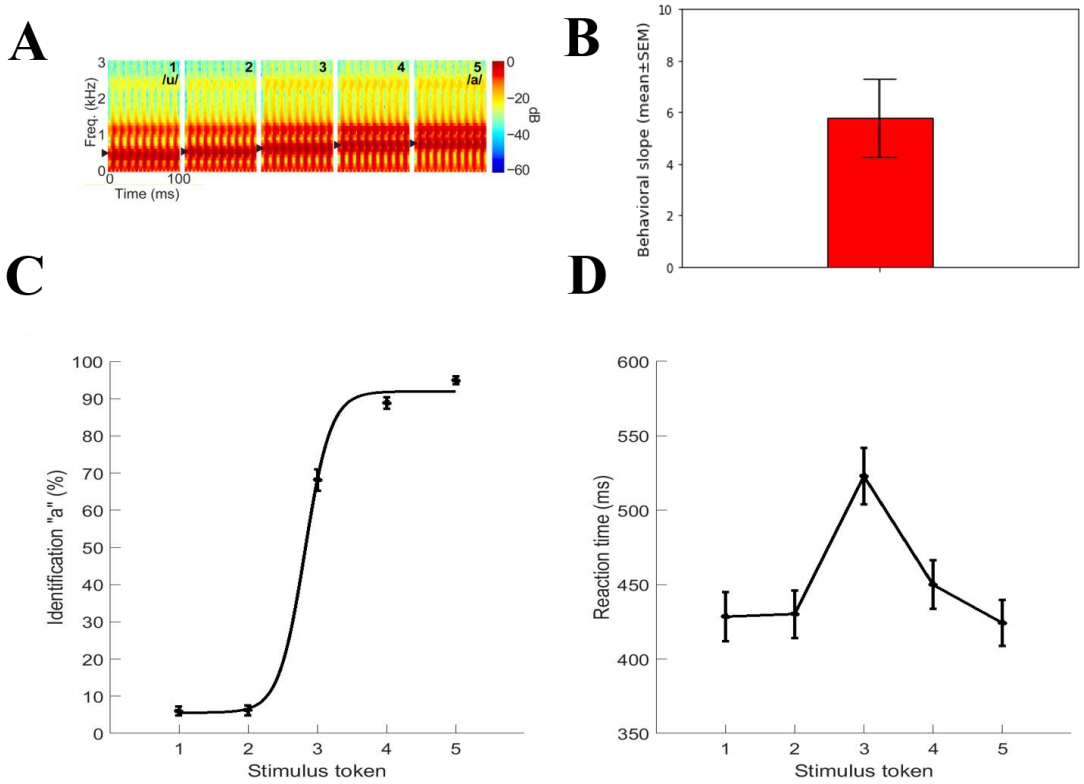

Figure 1: Speech stimuli and behavioral results. **A)** Acoustic spectrograms of the speech continuum from */u/* and */a/* ; Arrows: first formant frequency. **B)** Behavioral slope. **C)** Psychometric functions showing % "a" identification of each token. Listeners' perception abruptly shifts near the continuum midpoint, reflecting a flip in perceived phonetic category (i.e., "u" to "a"). **D)** Reaction time (RT) for identifying each token. RTs are fastest for category prototypes (i.e., Tk1/5) and slow when classifying ambiguous tokens at the continuum midpoint (i.e., Tk3). Errorbars = ±1 s.e.m.

Stimuli were presented binaurally at an intensity of 83 dB SPL through earphones (ER 2; Etymotic Research). Participants heard each token 150-200 times presented in random order. They were asked to label each sound in a binary identification task ("/u/" or "/a/") as fast and accurately as possible. Their response and reaction time were logged and reported in Figure 1C and D. The interstimulus interval (ISI) was jittered randomly between 400 and 600 ms (20 ms step and rectangular distribution) following listeners' behavioral responses to avoid anticipating the next trial (Luck, 2005).

3.3 EEG recordings and data pre-procedures

During the behavioral task, EEG was recorded from 64 channels at standard 10-10 electrode locations on the scalp (Oostenveld and Praamstra 2001). Continuous EEGs were

digitized using Neuroscan SynAmps RT amplifiers at a sampling rate of 500 Hz. Subsequent preprocessing was conducted in the Curry 7 neuroimaging software suite, and customized routines coded in MATLAB. Ocular artifacts (e.g., eye-blinks) were corrected in the continuous EEG using principal component analysis (PCA) (Picton et al., 2000) and then filtered (1-100 Hz bandpass; notched filtered 60 Hz). Trials with voltage ≥125 μV were discarded. Cleaned EEGs were then epoched into single trials (-200 to 800 ms, where $t = 0$ was stimulus onset). For details see in (Bidelman et al., 2020; Bidelman & Walker, 2017).

3.4 EEG source localization

To disentangle the sources of CP-related EEG activity, we reconstructed the scalp-recorded responses by performing a distributed source analysis in the Brainstorm software package (Tadel et al., 2011). All analyses were performed on single-trial data. We used a realistic boundary element head model (BEM) volume conductor and standard low-resolution brain electromagnetic tomography (sLORETA) as the inverse solution within Brainstorm (Tadel et al., 2011). A BEM model has less spatial errors than other existing head models (e.g., concentric spherical head model). We used Brainstorm's default parameter settings (SNR=3.00, regularization noise covariance = 0.1). From each single-trial sLORETA volume, we extracted the time-courses within 68 functional regions of interest (ROIs) across the left and right hemispheres defined by the Desikan-Killiany (DK) atlas (Desikan et al., 2006) (LH: 34 ROIs and RH: 34 ROIs). Single-trial data were then baseline corrected to the epoch's pre-stimulus interval (-200-0 ms).

Since we were interested to decode prototypical (Tk1/5) from ambiguous speech (Tk3)—a marker of categorical processing (Bidelman, 2015; Bidelman & Walker, 2019; Liebenthal et al., 2010)—we merged Tk1 and Tk5 responses since they reflect prototypical vowel categories ("u" vs. "a'). In contrast, Tk3 reflects a bistable percept—an category-ambiguous sound listeners sometimes label as "u" or "a" (Bidelman et al., 2020; Bidelman &

Walker, 2017; Mankel et al., 2020). To ensure an equal number of trials and signal to noise ratio (SNR) for prototypical and ambiguous stimuli, we considered only 50% of the data from the merged (Tk1/5) samples.

3.5 SVM classification to identify temporal dynamics and spectral bands of CP

Parameter optimized Support Vector Machine (SVM) classifiers provide better performance with small sample sizes data which is common in human neuroimaging studies. Classifier performance is greatly affected by tunable parameters in the SVM model (e.g., kernel, C, γ)[1] (Hsu et al., 2003). To lessen bias in parameter selection, we used a grid search approach during the training phase to find optimal kernel, *C,* and γ values. We randomly split the data into training (80%) and test (20%) sets (Park et al., 2011). During the training phase (e.g., using 80% data), we fine-tuned the *C* and γ parameters using grid search to find the optimal values such that the resulting classifier accurately distinguished prototypical vs. ambiguous speech in the test data (the remaining 20%) that models never seen. The grid search process was conducted with five-fold cross validation, kernels = 'RBF', fine-tune 20 different values of (*C* and γ) in the following range *C* = [1e-2 to 1e3], and γ = [1e-4 to 1e2] (Mahmud et al., 2020). The SVM learned the support vectors from the training data that comprised the attributes (e.g., ERP/frequency bands features) and class labels (e.g., Tk1/5 vs. Tk3). Then we selected the best model that has maximum margin with the optimal value of *C* and γ for predicting the unseen test data (only by providing the attributes but no class labels).

[1] Parameters γ and *C* in the SVM used in this study give measures of the influence of training data points on decision boundary and a measure of miss-classification tolerance. The first parameter γ comes from the radial basis function kernel (e.g., $K(x,x') = \exp\left(-\frac{||x-x'||^2}{2\sigma^2}\right)$ or equivalently $K(x,x') = \exp(-\gamma||x-x'||^2)$ with a parameter γ) where $\gamma = \frac{1}{2\sigma^2}$. In this study, the radial basis kernel is used as a transformation function. A larger value of γ implies smaller σ, which means that the classifier takes into account the effect of samples closer to the decision boundary. On the other hand, smaller γ means that the classifier considers the effect of samples farther from the decision boundary. The *C* is a parameter of SVM that acts as regularization. It provides the classifier a trade-off between the margin of decision boundary and miss- classification. A larger value of *C* produces a narrower (smaller-margin) hyperplane if that obtains less or no miss-classification. Whereas the smaller value of *C* allows drawing a wider (bigger-margin) hyperplane even if there are some miss-classifications. The optimal values of γ and *C* depend on data, which is why we used a grid search to tune these parameters in our classification model.

The classification performance metrics (accuracy, F1-score, precision, and recall) are calculated from standard formulas (Saito & Rehmsmeier, 2015).

3.6 Stability selection to identify spatial dynamics of CP

Our data included a large number of ERP/PSD of different frequency bands measurements for each stimulus condition of interest (e.g., Tk1/5 vs. Tk3). Larger numbers of variable/features can lead to overfitting and weak generalization in classification problems since the majority of features from brain activity (i.e., different ROIs, time segments) do not provide discriminative power for decoding CP. Consequently, we aimed to select a limited set of the most salient discriminating features. Stability selection is a state-of-the art feature selection method that works well in high dimensional or sparse data based on the Lasso (least absolute shrinkage and selection operator) (Meinshausen & Bühlmann, 2010; Yin et al., 2017). Stability selection can identify the most stable (relevant) features out of a large number of features over a range of model parameters, even if the necessary conditions required for the original Lasso method are violated (Meinshausen & Bühlmann, 2010).

In stability selection, a feature is considered to be more stable if it is more frequently selected over repeated subsampling of the data (Nogueira et al., 2017). Basically, the Randomized Lasso randomly subsamples the training data and fits a L1 penalized logistic regression model to optimize the error. Over many iterations, feature scores are (re)calculated. The features are shrunk to zero by multiplying the features' co-efficient by zero while the stability score is lower. Surviving non-zero features are considered important variables for classification. Detailed interpretation and mathematical equations of stability selection are explained in (Meinshausen & Bühlmann, 2010). The stability selection solution is less affected by the choice of the initial regularization parameters. Consequently, it is extremely general and widely used in high dimensional data even when the noise level is unknown (Meinshausen & Bühlmann, 2010).

In our implementation of stability selection, we used a sample fraction = 0.75, number of resamples = 1000, and tolerance = 0.01 (Meinshausen & Bühlmann, 2010). In the Lasso algorithm, the feature scores were scaled between 0 to 1, where 0 is the lowest score (i.e., irrelevant feature) and 1 is the highest score (i.e., most salient or stable feature). We estimated the regularization parameter from the data using the least angle regression (LARs) algorithm (Efron et al., 2004; Friedman et al., 2010). Over 1000 iterations, Randomized Lasso provided the overall feature scores (0~1) based on the number of times a variable was selected. We ranked stability scores to identify the most important, consistent, stable, and invariant features that could decode speech categories via the EEG (i.e., correctly classify Tk1/5 vs. Tk3). We submitted these ranked features and corresponding class labels to an SVM classifier with different stability thresholds and observed the model performance.

Chapter 4 - Spatiotemporal analysis of speech categorization

In this chapter, we demonstrate the spatio-temporal analysis to identify when in time and where in the brain CP could be decoded best from neural data (e.g., ERPs). We observed the classifier accuracy, as a sliding window decoder basis over the epoch to investigate the temporal characteristics of the neural signal that could categorize the speech sound best. In addition, we applied the stability selection to the whole epoch data for identifying the stable/invariant features or brain ROIs that associate with speech categorization.

4.1 Feature extraction

Previous computational studies have found that ERPs averaged over 100 trials provided the best classification of data while maintaining reasonable signal SNR and computational efficiency (Al-Fahad et al., 2020; Mahmud et al., 2020). We quantified source-level ERPs with a mean bootstrapping approach (James et al., 2013) by randomly averaging over 100 trials (with replacement) 30 times (Al-Fahad et al., 2020) for each stimulus condition per participant. For each resample and ROI time course, we measured the mean amplitude within a 20 ms sliding window (without overlapping) in the post-stimulus interval (i.e., 0 to 800 ms). In post hoc analysis, we parsed the epoch into "encoding" (0-260 ms) and "decoding/decision process" intervals

[1] (>300 ms) to investigate neural decoding related to pre- and post-perceptual processing, respectively. The sliding window resulted in 40 (800ms/20ms) ERP features (i.e., mean amplitude per window) for each ROI waveform, yielding a total of 68*40=2720 features per token (e.g., Tk1/5 vs. Tk3) from each listeners' data. Thus, the encoding and decision

[1]There is no clear division between "encoding" and "decision/post-processing" stages of perceptual chronometry. The choice of the ~300 ms mark was motivated by our previous demonstrating categorical coding within the time-frame of the N1-P2 waves of the ERP (< 250 ms) (Bidelman et al., 2013). We chose to include a subsequent time buffer between the two intervals so as to minimize overlap and therefore what was being decoded in each segment.

windows contained 13*68=884 (encoding) and 25*68=1700 (decision) ERP features. ERPs features were then used as input to an SVM classifier to access the temporal dynamics of the data and determine when in time CP was decodable from brain activity. State-of-the art variable selection (stability selection; see Section 3.6) (Meinshausen & Bühlmann, 2010) was then applied for identifying where in the brain (e.g., which ROIs) were involved in encoding and decision processes with regard to the categorization of speech. Before submitting to the SVM classifier, the data were z-score normalized to ensure all features were on a common scale range (Casale et al., 2008).

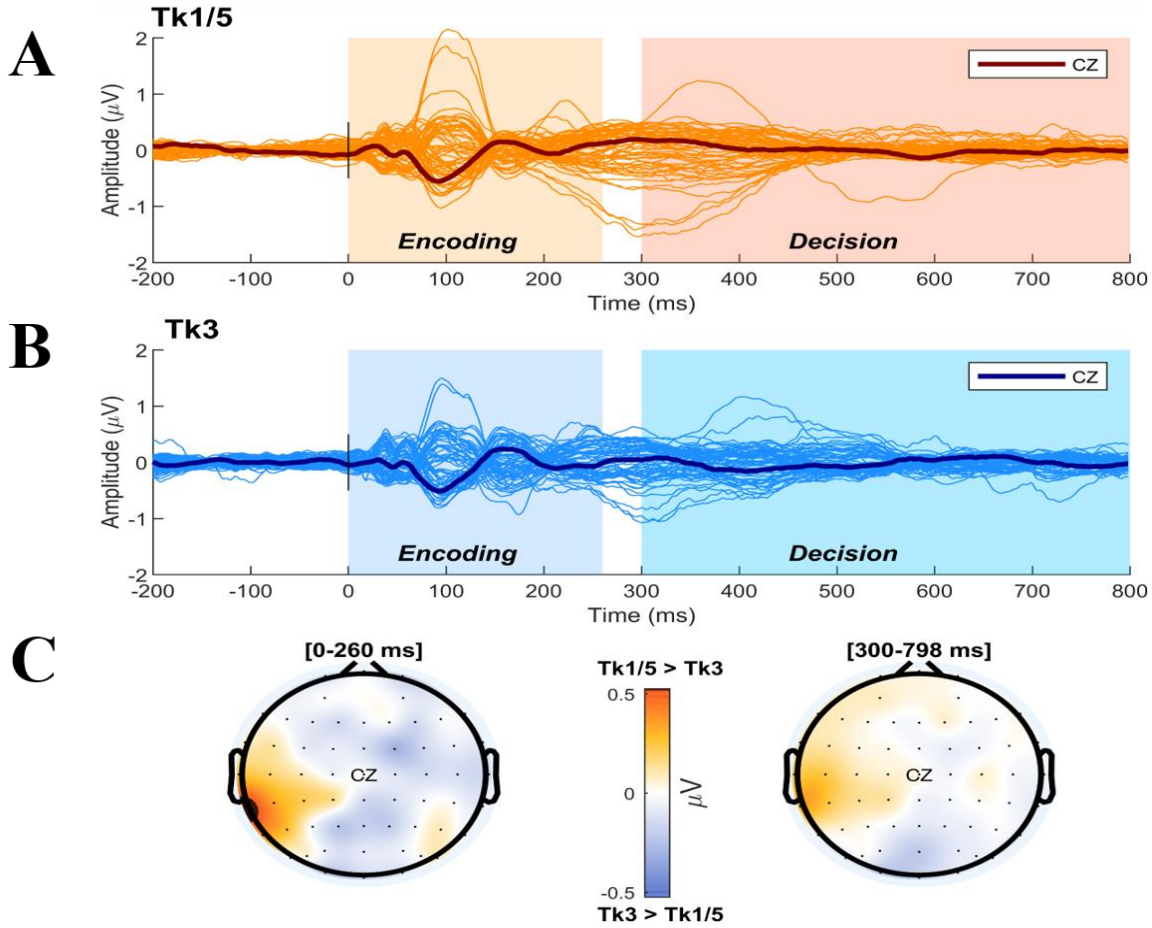

Figure 2: Grand averaged butterfly plots of scalp ERPs (64 channels) to prototypical (**A**; Tk1/5) vs. category-ambiguous (**B**; Tk3) vowels. Vertical lines demarcate segments for the stimulus encoding (0-260 ms) and decision period (300 ms-800 ms) analysis windows. *t*=0 marks stimulus onset. **C)** Topographic maps for encoding (left) and decision process (right).

4.2 Results

4.2.1 Behavioral results

Behavioral identification (%) functions and reaction time (ms) for speech categorization are depicted in Figure 1C and Figure 1D, respectively. Listeners' responses abruptly shifted in speech identity (/u/ vs. /a/) near the midpoint of the continuum, reflecting a change in perceived category. The behavioral speed of speech labeling (e.g., reaction time (RT)) were computed listeners' median response latency for a given condition across the all trials. RTs outside of 250-2500 ms were deemed outliers and excluded from further analysis (Bidelman et al., 2013; Bidelman & Walker, 2017). Listeners spent more time classifying the ambiguous (Tk3) than prototypical speech tokens (e.g., Tk1/5), further confirming categorical hearing (Pisoni & Tash, 1974). For each continuum, the identification scores were fit with a two parameters sigmoid function; $P = \frac{1}{[1+e^{-\beta 1(x-\beta 0)}]}$, where *P* is the proportion of the trial identification as a function of a given vowel, *x* is the step number along the stimulus continuum, and $\beta 0$ and $\beta 1$ the location and slope of the logistic fit estimated using the nonlinear least-squares regression (Bidelman et al., 2014; Bidelman & Walker, 2017). The slopes of listeners' sigmoidal psychometric function, reflecting the strength of their CP, is presented in Figure 1B.

4.2.2 Decoding the time-course of speech categorization from ERPs

We first examined how well categorical speech information could be decoded from whole-brain and individual hemisphere (e.g., LH and RH) ERPs data. During pilot modeling, we carried out a grid search approach (mentioned in Chapter 3 - 3.5) to develop parameters used for work shown here. The optimal values of C and γ parameters corresponding to the maximum speech decoding reported in Table 1 were: [C=10, γ=0.05 for whole-brain data; C=20, γ=0.01 for LH data; C=20, γ=0.01 for RH data]. We then selected the best model and

predicted the class labels (e.g., Tk1/5 vs. Tk3) by feeding the feature vectors only from the unseen test data. The performance metrics were calculated from predicted class labels and true class labels. Time-varying accuracy of the SVM classifier (i.e., distinguishing Tk1/5 vs. Tk3 responses) is shown in Figure 3.

Decoding was generally at chance level (54%) at stimulus onset (i.e., t = 0) but increased rapidly to a maximum accuracy of 95.16% by 120 ms (marked as circles in Figure 3). The individual hemispheres alone were less accurate and decoded speech categories later in time compared to whole-brain data (LH: 89.03% at 140 ms; RH: 86.45% at 200 ms) (marked as circles in Figure 3). Other important performance metrics of the SVMs at maximum decoding are reported in Table 1. Collectively, the earlier and improved ability of LH compared to RH in decoding phonetic categories is consistent with a LH bias in speech and language processing (Hickok & Poeppel, 2000). More critically, the early time course of decoding (120-150 ms) confirms that category level information, an abstract code, emerges very early in the neural chronometry of speech processing and well before listeners' execute their behavioral decision (cf. reaction times in Figure 1D) (Alho et al., 2016; Bidelman et al., 2013; de Taillez et al., 2020).

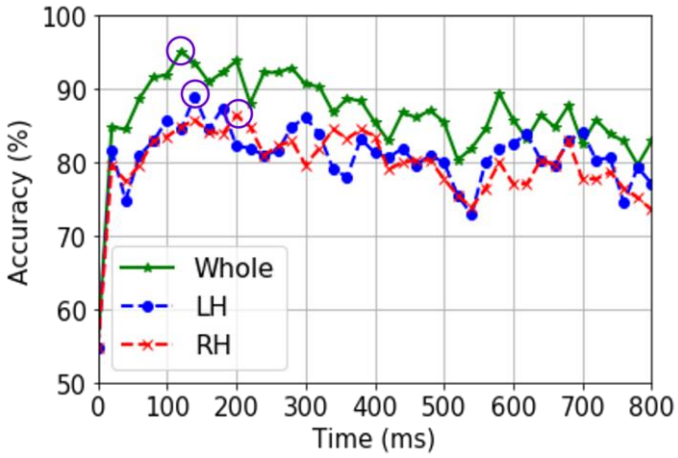

Figure 3: SVM classifier accuracy decoding speech categories from source ERPs. **A)** Decoding using whole-brain vs. hemispheres-specific data (LH and RH) across the epoch window. Maximum classification accuracies are marked by circles. Maximum classifier accuracy was observed at ~120 ms suggesting category representations emerge early, ~200 ms before listeners' behavioral categorization decisions (cf. Figure 1C).

Table 1: Performance metrics of the SVM classifier corresponding to maximal decoding of prototypical vs. ambiguous vowels from ERPs.

Metric (%)	Whole-brain features	LH features	RH features
Accuracy	95.16	89.03	86.45
AUC	95.14	89.18	86.45
F1-score	95.00	89.00	86.00
Precision	95.00	89.00	87.00
Recall	95.00	89.00	86.00

4.2.3 Decoding the spatial regions underlying categorization: stimulus encoding vs. decision

We used stability selection to find the most critical brain ROIs that were associated with categorical organization in the encoding (pre-perceptual) vs. decision (post-perceptual) periods of the task structure (see Figure 2). ERP features were considered stable (relevant) if they yielded a decoding accuracy performance >80%. The effect of stability threshold selection in the encoding and decision windows is illustrated in Figure 4. Each bin of histogram demonstrates the number of features in a range of stability threshold. The x-axis has four labels. The first line represents the stability score (0 to 1); the second and third line show the number of features and percentage of the selected features in the corresponding bin; line four represents the cumulative unique ROIs up to the lower boundary of the bin. The solid black and dotted red semi bell-shaped curves of Figure 4 represent classification

accuracy and AUC, respectively for the different stability thresholds. In this analysis, the number of unique brain ROIs represents distinct functional brain ROIs of the DK atlas and the number of features represents different time windows extracted from source ERPs. Features selected at each stability threshold were then submitted to an SVM classifier separately for the stimulus encoding and response decision periods.

During stimulus encoding (0-260 ms), 75% of features yielded stability scores 0 to 0.1. Thus, the majority of spatiotemporal ERP features were selected less than 10% out of 1000 model iterations and therefore carry weak importance in terms of describing categorical speech processing during stimulus encoding. In contrast, at a more conservative stability score of 0.3, 102 (11%) out of 884 ERP features selected from 52 ROIs were able to encode prototypical from ambiguous speech at near-ceiling accuracy (95.8%). Accuracy decreased precipitously with higher (more conservative) stability thresholds resulting in fewer (though more informative) brain ROIs describing category processing. For example, a stability score of 0.6—selecting only the most behaviorally-relevant features—still encoded speech categories well above chance (66.8%) with only 5 features from 5 ROIs. At stability score 0.5, speech encoding accuracy 82.6% only using 15 features from 13 unique ROIs. A BrainO visualization (Moinuddin et al., 2019) of relevant ROIs for the encoding period (threshold stability score $\geq$ 0.5) is shown in Figure 5 with additional details in Table 2.

During the decision period following stimulus encoding (> 300 ms), corresponding to the stability score 0.4, only 92 (5%) out of 1700 ERP features were selected, and the classifier showed decoding accuracy of 93.5% (AUC 93.6%). At a stability score 0.5 (corresponding to 83.2% accuracy), only 21 (1%) out 1700 ERP features from 15 unique ROIs were needed to describe categorical processing.

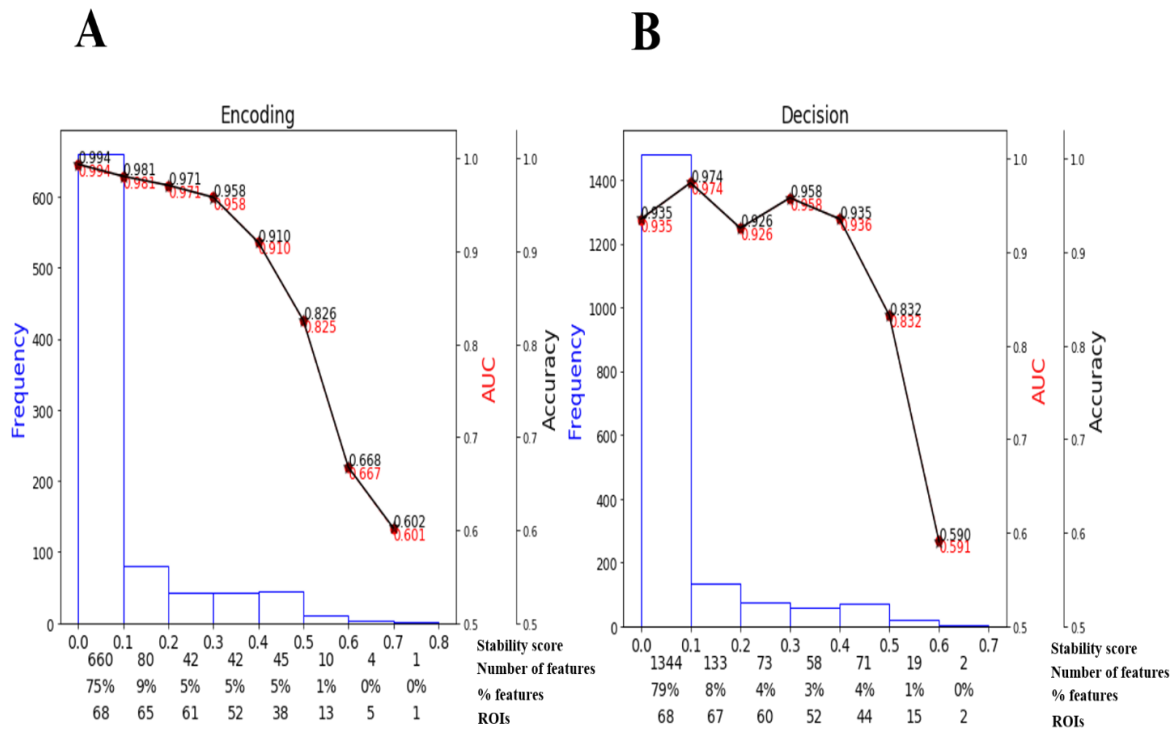

Figure 4: Effect of stability score threshold on model performance during (**A**) encoding and (**B**) decision period of the CP task. The bottom of the x-axis has four labels; *Stability score* represents the stability score range of each bin (scores: 0~1); *Number of features*, number of features under each bin; *% features*, the corresponding percentage of selected features; *ROIs*, number of cumulative unique brain regions up to the lower boundary of the bin.

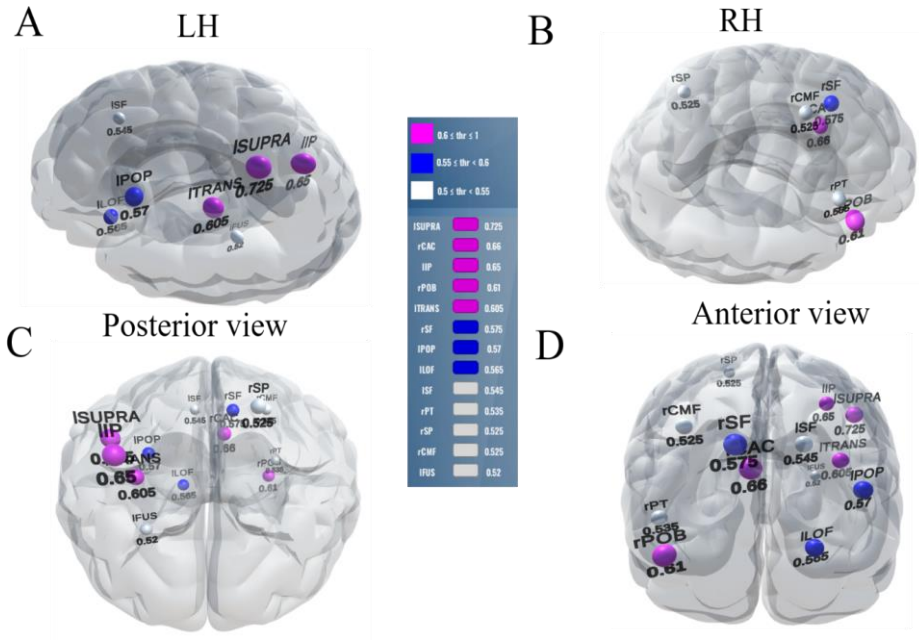

Figure 5: Stable (most consistent) neural network during the *encoding period* of CP. Visualization of brain ROIs corresponding to ≥ 0.50 stability threshold (13 top selected ROIs which show categorical organization (e.g., Tk1/5 ≠ Tk3) at 82.6%. **(A)** LH **(B)** RH **(C)** Posterior view (**D)** Anterior view. Color legend demarcations show high (pink), moderate (blue), and low (white) stability scores. l/r = left/right; SUPRA, supramarginal; CAC, caudal anterior cingulate; IP, inferior parietal; POB, pars orbitalis; TRANS, transverse temporal; SF, superior frontal; POP, pars opercularis; LOF, lateral orbitofrontal; PT, pars triangularis; SP, superior parietal; CMF, caudal middle frontal; FUS, fusiform.
.

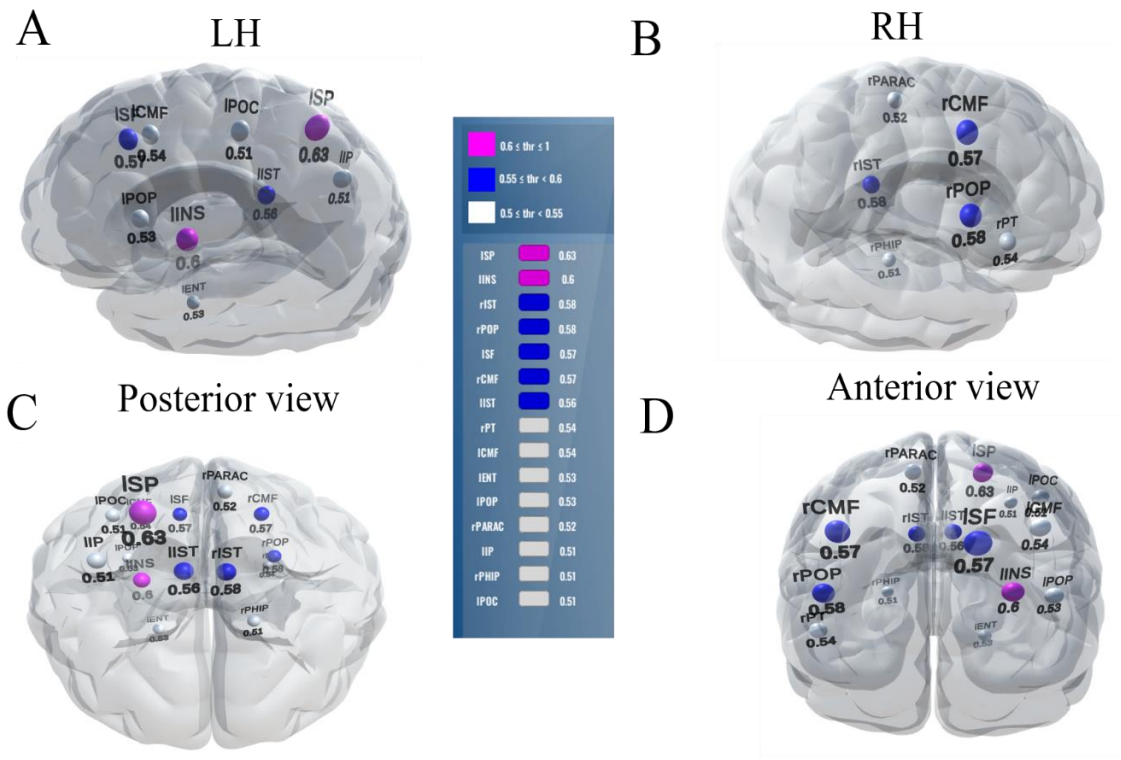

Figure 6: Stable (most consistent) neural network during the *decision period* of CP. Visualization of brain ROIs corresponding to ≥ 0.50 stability threshold (15 top selected ROIs which decode Tk1/5 from Tk3 at 83.2%. Otherwise as in Figure 5. SP, superior parietal; INS, Insula; POP, pars opercularis ; SF, superior frontal; CMF, caudal middle frontal; IST, isthmus cingulate; PT, pars triangularis; CMF, caudal middle frontal; ENT, entorhinal; PARAC, paracentral; IP, inferior parietal; PHIP, para hippocampal ;POC, postcentral.

Table 2: Most important brain regions describing speech categorization during stimulus encoding (13 ROIs) and response decision (15 ROIs) at a stability threshold ≥ 0.5.

Rank	**Encoding (82.6% total accuracy)**			**Decision (83.2% total accuracy)**		
	ROI name	**ROI abbrev.**	**Stability score**	**ROI name**	**ROI abbrev.**	**Stability score**
1	Supramarginal L	lSUPRA	0.73[a]	Superior parietal L	lSP	0.63
2	Caudal anterior cingulate R	rCAC	0.66	Insula L	lINS	0.60
3	Inferior parietal L	lIP	0.65	Isthmus cingulate R	rIST	0.58

4	Pars orbitalis R	rPOB	0.61	Pars opercularis R	rPOP	0.58
5	Transverse temporal L	lTRANS	0.61	Superior frontal L	lSF	0.57
6	Superior frontal R	rSF	0.58	Caudal middle frontal R	rCMF	0.57
7	Pars opercularis L	lPOP	0.57	Isthmus cingulate L	lIST	0.56
8	Lateral orbitofrontal L	lLOF	0.57	Pars triangularis R	rPT	0.54
9	Superior frontal L	lSF	0.55	Caudal middle frontal L	lCMF	0.54
10	Pars triangularis R	rPT	0.54	Entorhinal L	lENT	0.53
11	Superior parietal R	rSP	0.53	Pars opercularis L	lPOP	0.53
12	Caudal middle frontal R	rCMF	0.53	Paracentral R	rPARAC	0.52
13	fusiform L	lFUS	0.52	Inferior parietal L	lIP	0.51
14				Parahippocampal R	rPHIP	0.51
15				Postcentral L	lPOC	0.51

[a] A score of 0.73, for example, means that out of 1000 iterations, the ERP feature of this ROI was selected 730 times by stability selection.

4.2.4 Brain-behavior correspondences

Multivariate regression analysis is widely used to investigate when more than one predictor simultaneously influences an outcome variable (Hanley, 1983; Royston &

Sauerbrei, 2008). To evaluate the behavioral relevance of the brain regions identified via stability selection, we conducted multivariate regression using weighted least squares (WLS) regression (Ruppert & Wand, 1994). Regressions were computed between the 15 ROI ERPs identified in the decision interval and listeners' behavioral slopes (Figure 1B), which indexes their degree of categorical hearing. We computed the mean neural response (i.e., ERP) within each selected region across the stimuli [mean ERP of (Tk1/5 & Tk3)] and then regressed the 15 ROI responses simultaneously against listeners' behavioral slope. The inverse of the absolute error values of the ordinary least squares were used as weights in the WLS to reduce the effect of heteroscedasticity (Seabold & Perktold, 2010; *Weighted Regression in SAS, R, and Python*, n.d.). The multivariate model robustly predicted listeners' behavioral CP from neural data ($R^2 = 0.85$, $p<0.00001$; Table 3), demonstrating the selected 15 ROIs identified via ML (i.e., stability selection) carried behaviorally relevant information regarding CP.

Table 3: WLS regression results describing how individual brain ROIs predict behavioral CP.

Rank	**ROI name**	**ROI abbrev.**	**Coefficient**	**t-value**	**p-value**
1	Superior parietal L	lSP	-0.2163	-3.008	0.004920
2	insula L	lINS	0.1808	5.188	0.000010
3	Isthmus cingulate R	rIST	-0.2679	-3.764	0.000633
4	Pars opercularis R	rPOP	0.1231	4.429	0.000093
5	Superior frontal L	lSF	-0.1726	-3.190	0.003055
6	Caudal middle frontal R	rCMF	0.1544	2.367	0.023774
7	Isthmus cingulate L	lIST	0.2259	2.792	0.008545
8	Pars triangularis R	rPT	-0.0214	-0.679	0.501925
9	Caudal middle frontal L	lCMF	0.0153	0.345	0.732223
10	entorhinal L	lENT	0.1170	5.009	0.000013

11	Pars opercularis L	lPOP	0.1475	3.892	0.000441
12	paracentral R	rPARAC	0.2223	3.308	0.002226
13	Inferior parietal L	lIP	-0.1017	-1.364	0.181508
14	Parahippocampal R	rPHIP	-0.0422	-2.097	0.043540
15	Postcentral L	lPOC	0.1809	2.749	0.009512

4.3 Discussion

We conducted machine learning analyses on EEG to examine the spatiotemporal dynamics of speech processing during rapid speech sound categorization. We found that speech categories are best decoded via patterned neural activity occurring within 120 ms and no later than 200 ms. We also identified the most relevant brain regions that are involved in encoding and decision stages of the categorization process. Our findings show a small set of brain areas (15 ROIs) robustly predicts listeners' categorical decisions, accounting for 85.0% of the variance in behavior.

4.3.1 Speech categories are decoded early (<150 ms) in the time course of perception

We have replicated and extended previous work by using whole-brain EEG and SVM neural classifiers to examine the time-course and hemispheric asymmetry as the brain decodes the identity of speech sounds. We found optimal speech decoding in the time frame of the N1 wave (120 ms) of the auditory ERPs using full-brain data. Analysis by hemisphere further showed that LH yielded better and earlier decoding than the RH, where optimal decoding occurred 20-80 ms later (LH: 140 ms; RH: 200 ms). These latencies are compatible with the N1-P2 waves of the auditory ERPs and suggest a rapid speed to phonetic categorization (Alho et al., 2016; Bidelman et al., 2013; de Taillez et al., 2020). Our results are consistent with previous neuroimaging studies that have shown the N1 and P2 ERPs are sensitive to auditory perceptual object identification (Alain, 2007; Bidelman et al., 2013; Wood et al., 1971). The better decoding by LH as compared to RH activity is consistent with

the dominance of LH in phoneme discrimination and speech sound processing (Bidelman & Howell, 2016; Bidelman & Walker, 2019; Frost et al., 1999; Tervaniemi & Hugdahl, 2003; Zatorre et al., 1992). Our neural decoding results also corroborate previous hypothesis-driven work (Bidelman et al., 2013, 2014; Chang et al., 2010) by confirming speech sounds are converted to an abstract, categorical representation within the first few hundred milliseconds after stimulus onset.

4.3.2 Differential brain-networks involved in encoding and decision processing

Our results help identify the most stable, relevant, and invariant functional brain ROIs that support the brain-networks involved in encoding and decision processes of speech categorization using an entirely data-driven approach (stability selection coupled with SVM). During stimulus encoding, stability selection have identified 13 consistent ROIs that differentiate speech categories (82.6% accuracy; 0.5 stability threshold). Out of these 13 regions, eight of the ROIs are critically involved in the dorsal-ventral pathway for speech-language processing (Hickok & Poeppel, 2004). These included areas in frontal lobe including inferior frontal gyrus [BA 44, (i.e., pars opercularis L, pars triangularis R), i.e., "Broca's area"], three regions from parietal and two regions from temporal lobe including primary auditory cortex (i.e., transverse temporal L). For later decision stages of the task, the same criterion of decoding performance (83.2% @ 0.5 stability threshold) has identified 15 ROIs that showed categorical neural organization. Out of these 15 regions, eight areas are from inferior frontal areas including BA 44 (i.e., pars opercularis L, pars opercularis R) and BA 45 (i.e., pars triangularis R), four regions from parietal lobe, and three regions from temporal lobe. Our data reveal two, relatively sparse, and partially overlapping neural networks that support different stages of speech categorization process.

Among the encoding and decision networks identified from our EEG data, five regions were common between the two topologies. Notably were the inclusion of BA44/45 that are

heavily involved in speech-language processing (Hickok et al., 2011; Lee et al., 2012; Novick et al., 2010). Early activation of IFG (during encoding) could be due to higher order speech centers exerting an inhibitory influence on auditory representations in order to prevent interference from nonlinguistic cues (Dehaene-Lambertz et al., 2005; Liberman et al., 1981) and optimize categorization, particularly under states of uncertainty (Carter & Bidelman, 2020). The left inferior parietal lobe also appears as a common hub among the two networks. Superior parietal areas have been linked with auditory, phoneme, sound categorization, particularly when listeners are asked to resolve context or ambiguity (Dufor et al., 2007; Feng et al., 2018; Myers & Blumstein, 2008). Involvement of superior frontal lobe in both networks is perhaps consistent with its role in higher cognitive functions and working memory (Klingberg et al., 2002; Nyberg et al., 2003). The fact that these extra-sensory regions can decode category structure even during stimulus encoding (< 150 ms) suggests that the formation of speech categories might operate nearly in parallel within lower-order (sensory) and higher-order (cognitive-control) brain structures (Toscano et al., 2018). However, these category representations need not be isomorphic across the brain. For example, category formation might reflect a cascade of events where speech units are reinforced and further discretized by a recontact of acoustic-phonetic with lexical representation of the speech category (Myers & Blumstein, 2008).

Notable among the non-overlapping regions between stages were left primary auditory cortex (transverse temporal) and supramarginal gyrus, both of which were exclusive to the stimulus encoding period. Both regions have been implicated in the early acoustic analysis of the speech signal and related phonological processing (Deschamps et al., 2014; Geiser et al., 2008; Hickok et al., 2000; Oberhuber et al., 2016; Whitwell et al., 2013; Zatorre et al., 1992). Intuitively, their absence during the decision stage further suggests the categorical representation of speech, while present early in time (< 150 ms), might take different forms in auditory-sensory cortex before being broadcast to decision mechanisms downstream.

Left postcentral gyrus is also exclusive during decision. Activation of this area proximal to the behavioral response execution most probably reflects motor planning and/or speech reconstruction (Martin et al., 2014). Additional non-overlapping ROIs included pars opercularis in the RH. Right IFG has been implicated in attentional control and response imbibition (Hampshire et al., 2010), which is consistent with its exclusive involvement in the decision stage of our task. Presumably, the other non-overlapping regions identified via stability selection (superior parietal L, insula L, Isthmus cingulate (l/rIST), caudal middle frontal L, entorhinal L, paracentral R, parahippocampal R) are also involved in decision processes, though as of yet, in an unknown way. Minimally, the involvement parahippocampal regions implies putative memory and retrieval processes. Still, more detailed localization studies (e.g., using fMRI) are needed to validate our EEG data, which offers a much coarser spatial resolution.

It is noticeable that during encoding, 7 out of 13 ROIs are from LH; for decoding, 9 out of 15 ROIs. The left hemisphere bias in our decoding data is perhaps expected given the LH dominance in auditory language processing (Caplan, 1994; Hull & Vaid, 2006; Tzourio et al., 1998). Moreover, our results support previous studies by confirming a bilateral fronto-parietal network involved in auditory attentional, working memory (Belin et al., 2002; Schneiders et al., 2012), sound discrimination tasks (Hickok & Poeppel, 2000), and phoneme categorization (Bidelman & Walker, 2019; Lee et al., 2012; Loui, 2015). Interestingly, our study shows that only 15 brain regions (during decision) are needed to predict listeners' behavior CP with 85.0% accuracy.

Chapter 5 - Speech categorization from evoked versus induced responses

In this chapter, we discuss evoked vs. induced analysis to assess which mode of brain oscillations and frequency bands could categorize speech sound well. Our analysis shows that induced brain activity could categorize the speech sound better than the evoked activity. Particularly, we found that induced gamma frequency is the strongest predictable ability among all other frequency bands. Remarkably, induced high frequency (γ-band) oscillations dominate CP decoding in the left hemisphere, whereas lower frequency (θ-band) dominate decoding in the right hemisphere.

5.1 Evoked activity and induced features extraction

Here we have discussed time-frequency analysis for evoked and induced analysis. To separate the induced and evoked activity, we used the wavelet transform. We applied wavelet transform on each trial of each ROIs then took their absolute values and averaged them up. This averaged signal contains the evoked and induced activity. For evoked activity, we averaged over a number of trials that yielded evoked activity. Then we applied wavelet on each brain ROIs. To extract the induced activity, we subtracted the evoked activity from the total activity. We discuss detail in the following section.

5.2.1 Time-Frequency analysis

Time-frequency analysis was conducted via wavelet transform (Herrmann et al., 2014). First, we computed the ERP using bootstrapping by randomly averaged over 100 trials with the replacement 30 times (Al-Fahad et al., 2020) for each stimulus condition (e.g., Tk1/5 and Tk3) per subject and source ROI (e.g., 68 ROIs). We then applied the Morlet wavelet transform to each ROI average data (i.e., ERP) with time steps of 2 ms and an increment step frequency 1 Hz from low to high frequency (e.g., 1 to 100 Hz) across the epoch, which provided only evoked frequency-specific activity (i.e., time- and phase-locked to stimulus onset). For computing induced activity, we performed a similar Morlet wavelet transform on a single-trial basis for each ROI, and then computed the absolute value of each trial

spectrogram. We then averaged the resulting time-frequency decompositions (Herrmann et al., 2014), resulting in a spectral representation that contains total activity. To isolate induced responses, we subtracted the evoked activity from the total activity (Herrmann et al., 2014). We then extracted the different frequency band signals from evoked and induced activity time-frequency maps for each brain region (e.g., 68 ROIs). Examplary data showing evoked and induced time-frequency maps from the primary auditory cortex [i.e., transverse temporal (TRANs)] are shown in Figure 7. We did not separate early vs. late windows in this study as we have previously shown induced activity during speech categorization tasks is largely independent of motor responses (Bidelman, 2015).

Spectral features of different bands (θ, α, β, and γ) were quantified as the mean power over the full epoch. We concatenated four frequency bands that resulted in 4*68=272 features for each response type (e.g., evoked vs. induced) per speech condition (Tk1/5 vs. Tk3). We conducted a paired t-test between evoked and induced feature vectors [e.g., concatenating all frequency band features of each ROI and stimulus type (Tk1/5 vs. Tk3)] and found statistical significance [$t(783359) = 1212.53$, $p<0.001$] between the two brain modes. We also conducted one-way ANOVA tests to identify a possible band effects (θ, α, β, and γ band activity) within each brain-regime. Band modulations were evident in both induced [$F(4, 2876) = 247499.16$, $p<0.001$] and evoked [$F(4, 2876) = 108336.47$, $p<0.001$] activities. To assess which regime (evoked vs. induced) and oscillatory band (θ, α, β, and γ) is more important to speech categorization, we then used machine learning classifiers to decode the data. We separately (i.e., evoked and induced) submitted the individual frequency bands to the support vector machine (SVM) classifier and all concatenated features (e.g., θ, α, β, and γ bands) to stability selection to investigate which frequency bands and brain regions decode prototypical (e.g., Tk1/5) from ambiguous (Tk3) vowels. Features were z-scored prior to SVM to normalize them to a common range.

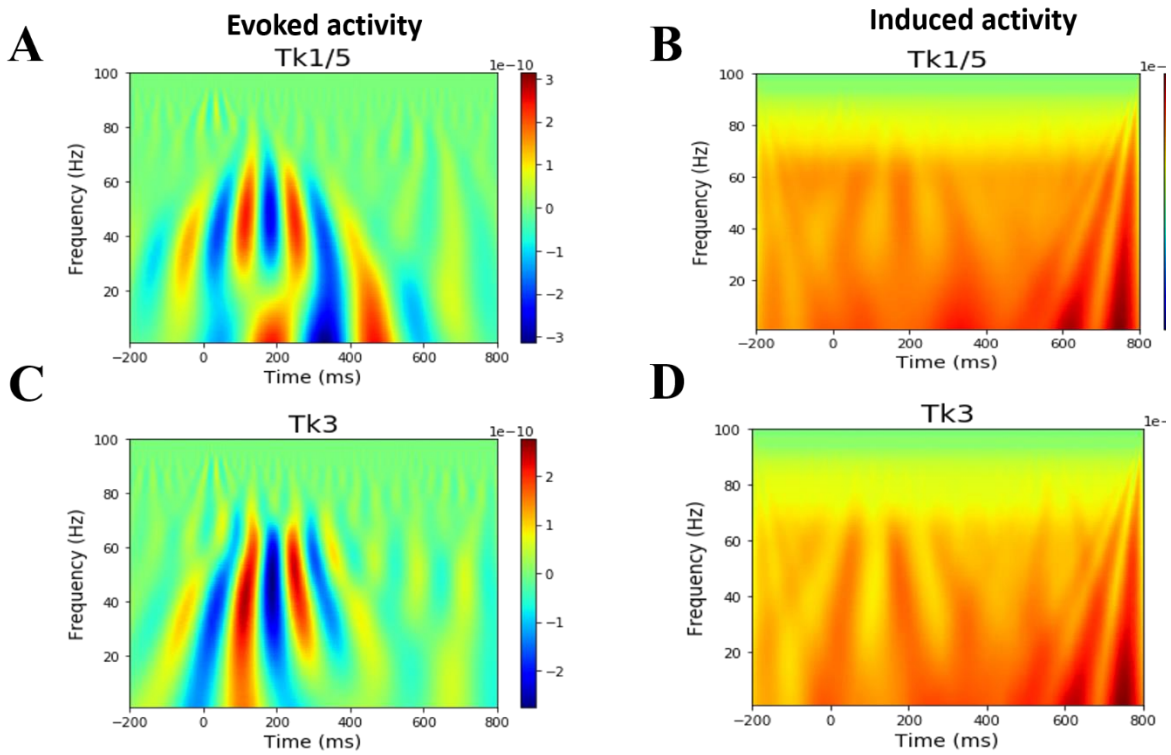

Figure 7: Grand average neural oscillatory responses to prototypical vowel (e.g., Tk1/5 and ambiguous speech token (Tk3) A,C) Evoked activity for prototypical vs. ambiguous tokens. B, D) Induced activity for prototypical vs. ambiguous tokens. Primary auditory cortex (PAC) [lTRANS, left transverse temporal gyrus].

5.2 Results

5.2.1 Decoding categorical neural responses using band frequency features and SVM

We investigated the decoding of prototypical from ambiguous vowels (i.e., category-level representations) using SVM neural classifier on whole-brain (all 68 ROIs) and individual hemisphere (LH and RH) data separately for induced vs. evoked activity. The best model performance on test-dataset is reported in Figure 8 and Table 4.

Using whole-brain *evoked* θ, α, β, and γ frequency responses, speech stimuli (e.g., Tk1/5 vs. Tk 3) were correctly distinguished at 66-69% accuracy. Among all evoked frequency bands, β-band was optimal to decode speech categories (69.61% accuracy). LH data revealed that θ, α, β, and γ bands decoded speech stimuli at accuracies between ~63-65% whereas decoding from RH was slightly poorer 57-62%.

Using whole-brain *induced* θ, α, β, and γ frequency responses speech stimuli were decodable at accuracies 89-95%. Among all induced frequency bands, γ band showed the best speech segregation (94.9% accuracy). Hemisphere specific data again showed lower accuracy. LH oscillations decoded speech categories at 76-87% accuracy whereas RH yielded 80-84%.

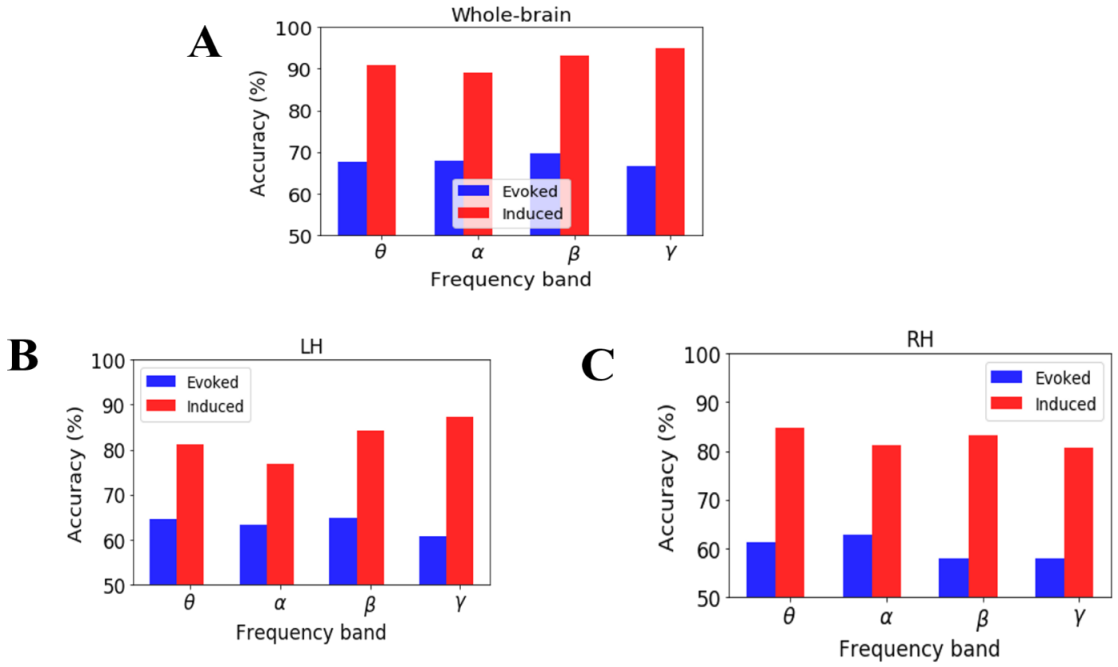

Figure 8: Decoding categorical neural encoding using different frequency band features of source-level EEG. SVM results classifying prototypical (Tk1/5) vs. ambiguous (Tk 3) speech sounds. A) Whole-brain data (e.g., 68 ROIs), B) LH (e.g., 34 ROIs) C) RH (e.g., 34 ROIs). Chance level =50%.

Table 4: Performance metrics of the SVM classifier corresponding to maximal decoding of prototypical vs. ambiguous vowels from ERPs.

Neural activity	**Frequency band**	**Accuracy (%)**	**AUC (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
Evoked	θ	67.53	67.59	68.00	68.00	67.00
	α	67.88	67.92	68.00	68.00	67.00
	β	69.61	69.64	68.00	68.00	68.00
	γ	66.66	66.65	67.00	67.00	67.00
	θ	90.97	90.98	91.00	91.00	91.00

Induced	α	89.06	89.05	89.00	89.00	89.00
	β	93.23	93.20	93.00	93.00	93.00
	γ	94.96	94.96	95.00	95.00	95.00

We also reported the mean accuracy and error variance of five-fold cross-validation accuracy in the Figure 9.

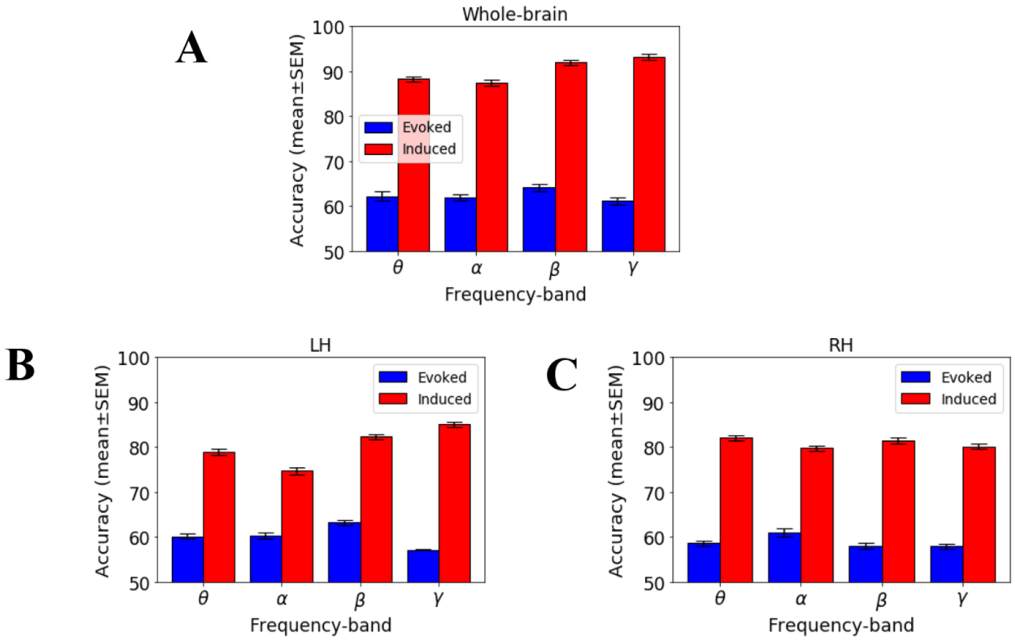

Figure 9: Decoding categorical neural encoding using different frequency band features of source-level EEG. Mean accuracy of SVM five-fold cross-validation results classifying prototypical (Tk1/5) vs. ambiguous (Tk 3) speech sounds. A) Whole-brain data (e.g. 68 ROIs), B) LH (e.g., 34 ROIs) C) RH (e.g., 34 ROIs). Chance level =50%. Errorbars = ±1 s.e.m.

Using whole-brain *evoked* θ, α, β, and γ frequency responses, speech stimuli (e.g., Tk1/5 vs. Tk 3) were correctly distinguished at a mean accuracy of 61-64% accuracy. LH data revealed that θ, α, β, and ɣ bands decoded speech stimuli at mean accuracies between ~57-63% whereas decoding from RH was slightly poorer 57-61%. The mean accuracy is ~5% less than the best model using the evoked activity.

Using whole-brain *induced* θ, α, β, and γ frequency responses speech stimuli were decodable at mean accuracies 87-93%. Still, among all induced frequency bands, ɣ band

showed the best speech segregation (93% mean accuracy). LH oscillations decoded speech categories at 75-85% mean accuracy whereas RH yielded 80-82%. The maximum deviation of accuracy ~2% from the best model using induced activity.

5.2.2 Decoding categorical neural responses using band frequency features and KNN

We used a KNN classifier to corroborate the main SVM findings with a different algorithm. We split the data into training and test sets of 80% and 20%, respectively. During the training phase, we tuned the value of *k* parameters from 1 to 10 to achieve maximum accuracy. Classification results of KNN on the test dataset is reported in the Figure 10. The KNN classifier exhibited similar though slightly inferior results than the SVM classifier, justifying our choice of the SVM model.

Using whole-brain *evoked* θ, α, β, and ɣ band frequency responses, speech stimuli (e.g., Tk1/5 vs. Tk 3) were correctly distinguished at 64-68% accuracy. Among all evoked frequency bands, β-band was optimal to decode speech categories (67.70% accuracy). LH data revealed that θ, α, β, and ɣ bands decoded speech stimuli at accuracies between ~58-63% whereas decoding from RH was slightly poorer 57-62%.

Using whole-brain *induced* θ, α, β, and ɣ frequency responses speech stimuli were decodable at accuracies 88-94%. Among all induced frequency bands, ɣ band showed the best speech segregation (94.42% accuracy). Hemisphere specific data again showed lower accuracy. LH oscillations decoded speech categories at 76-86% accuracy whereas RH yielded 79-84%.

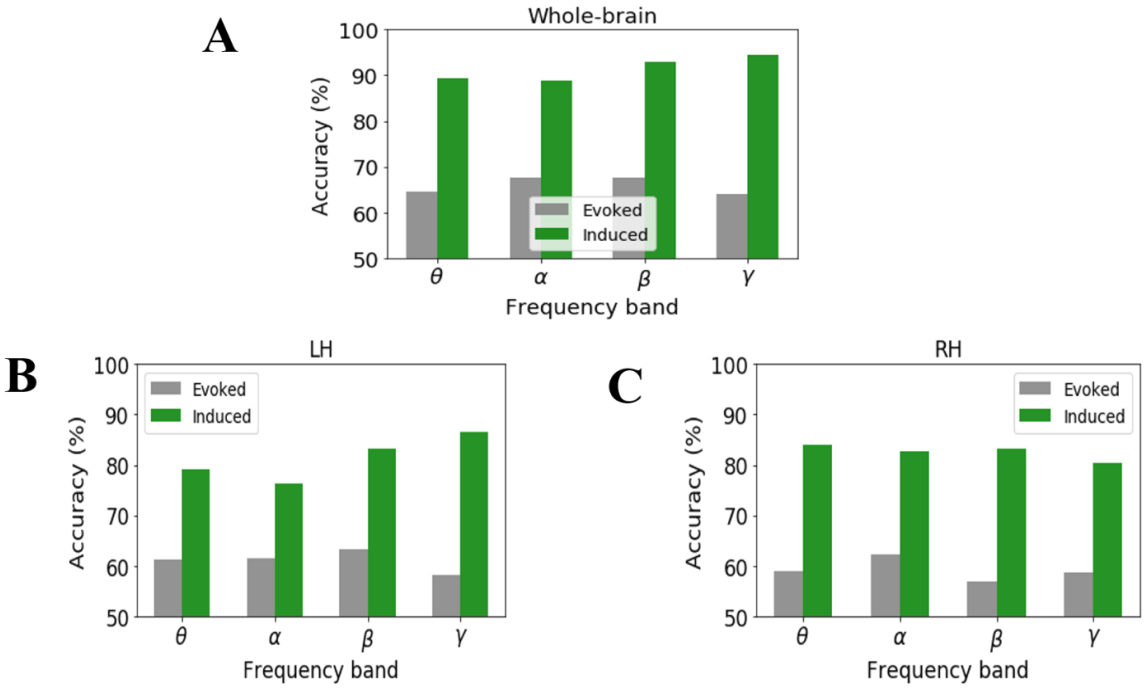

Figure 10: Grand Decoding categorical neural encoding using different frequency band features of source-level EEG. KNN results classifying prototypical (Tk1/5) vs. ambiguous (Tk 3) speech sounds. A) Whole-brain data (e.g., 68 ROIs), B) LH (e.g., 34 ROIs) C) RH (e.g., 34 ROIs). Chance level =50%.

5.2.3 Decoding brain regions associated with CP (evoked vs. induced)

We separately applied the stability selection (Meinshausen & Bühlmann, 2010) to induced and evoked activity features to identify the most critical brain areas (e.g., ROIs) that have been linked with speech categorization. Spectral features of brain ROIs were considered stable if the speech decoding accuracy was >70%. The effects of stability scores on speech sound classification is represented in Figure 11. Each bin of the histogram illustrates the number of features in a range of stability scores. In this work, the number of features (labeled in Figure 11) represents the neural activity of different frequency bands and the unique brain regions (labeled as ROIs in Figure 11) represent the distinct functional brain regions of the DK atlas. The semi bell-shaped solid black and dotted red lines demonstrate classifier accuracy and AUC, respectively. We submitted the neural features identified at different stability thresholds to SVMs. This allowed us to determine whether the collection of neural

measures identified via machine learning were relevant to classifying speech sound categorization.

For induced responses, most features (60%) yielded stability scores 0 to 0.1, meaning 163/272 (60%) were selected less than 10% out of 1000 iterations from 68 ROIs. A stability score of 0.2 selected 47/86 (32%) of the features from 47 ROIs that could decode speech categories at 96.9% accuracy. Decoding performance decreased with increasing the stability score (i.e., more conservative variable/brain ROIs selection) resulting in a reduced feature set that retained only the most meaningful features distinguishing speech categories from a few ROIs. For instance, corresponding to the stability threshold 0.5, 25 (10%) features were selected from 21 brain ROIs that yielded the speech categorization 92.7 % accurately. However, corresponding to the stability threshold 0.8 only 2 features were selected from 2 brain ROIs that decoded CP at 60.6%, still greater than chance level (i.e., 50%). Performance improved by ~10% (86.5%) when the stability score was changed from 0.7 (selected brain ROIs 9) to 0.6 (selected brain ROIs 14). A BrainO visualization (Moinuddin et al., 2019) of these brain ROIs is delineated in Figure 12.

Using evoked activity, maximum decoding accuracy was 78.0% at a 0.1 stability threshold. Here, 43 % of features produced a stability score between 0.0 to 0.1. These 118 (43%) features are not informative because they decreased the model's accuracy to properly categorize speech. Corresponding to the stability scores 0.9, only 8 features were selected from the 6 brain ROIs, which decoded speech at 65.8% accuracy. At stability score 0.6, 29 (1%) features were selected from 22 brain ROIs corresponding to 71.4 % accuracy performance.

Our goal is to build an interpretable model that can describe speech categorization with reasonable accuracy using the smallest number of brain ROIs/features. Usually, the knee-point is a location along the stability curve which balances model complexity (i.e., feature count) and decoding accuracy. Thus, 0.6 might be considered an optimal stability

score (i.e., knee point of a function in Figure 11) as it decoded speech well above change (>70%) with a minimal (and therefore more interpretable) feature set for both induced and evoked activity. Brain ROIs corresponding to the optimal stability score (0.6) are depicted in Figure 13 and Table 5 for both evoked and induced activities.

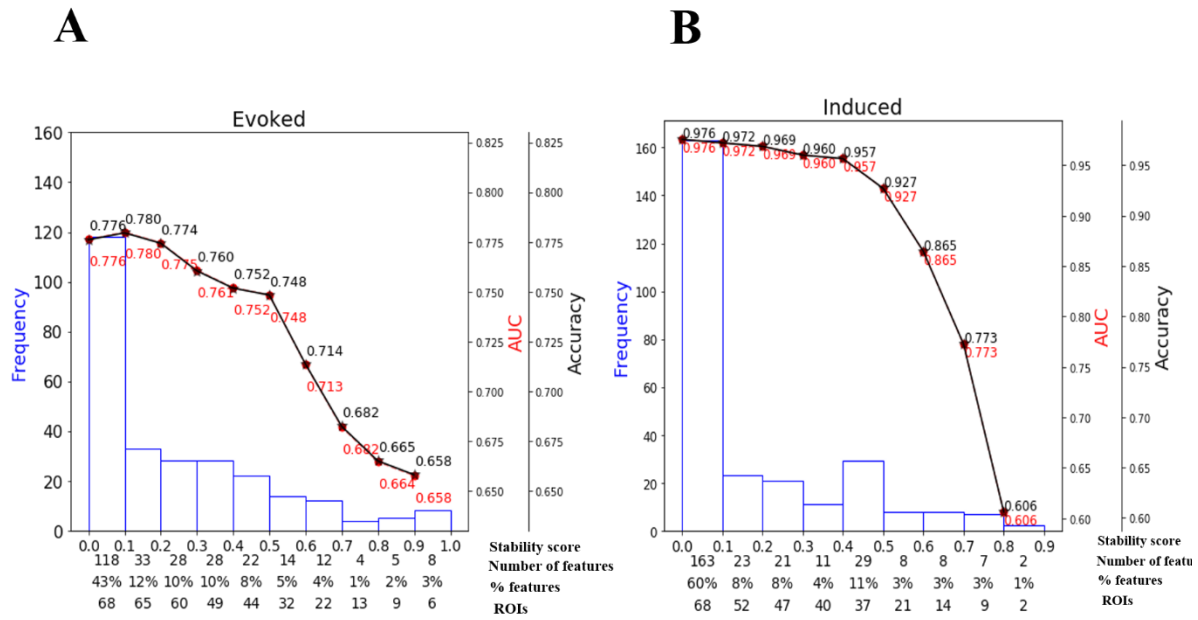

Figure 11: Effect of stability score threshold on model performance during (A) evoked activity and (B) induced activity during CP task. The bottom of the x-axis has four labels; *Stability score* represents the stability score range of each bin (scores range: 0~1); *Number of features*, number of selected features under each bin; *% features*, the corresponding percentage of selected features; *ROIs*, number of cumulative unique brain regions up to the lower boundary of the bin.

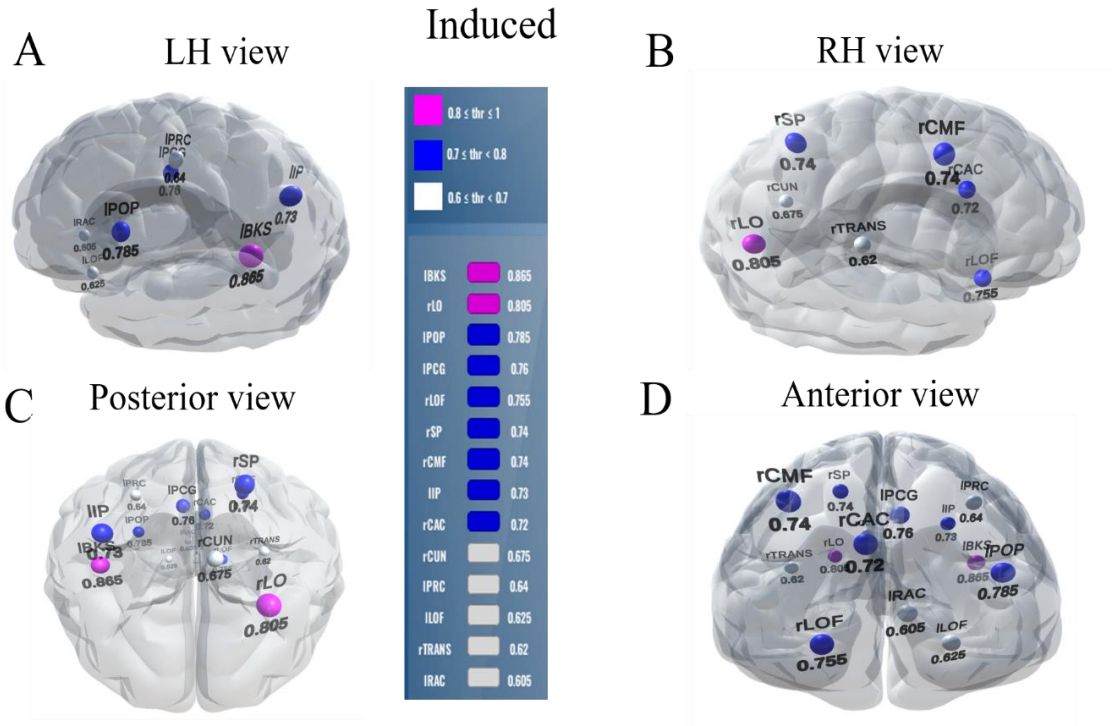

Figure 12: Stable (most consistent) neural network decoding using induced activity. Visualization of brain ROIs corresponding to ≥ 0.60 stability threshold (14 top selected ROIs which show categorical organization (e.g., Tk1/5 ≠ Tk3) at 86.5%. **(A)** LH view **(B)** RH view **(C)** Posterior view (**D)** Anterior view. Color legend demarcations show high (pink), moderate (blue), and low (white) stability scores. l/r = left/right; BKS, bankssts; LO, lateral occipital; POP, pars opercularis; PCG, posterior cingulate; LOF, lateral orbitofrontal; SP, superior parietal; CMF, caudal middle frontal; IP, inferior parietal; CAC, caudal anterior cingulate; CUN, cuneus; PRC, precentral; TRANS, transverse temporal; RAC, rostral anterior cingulate.

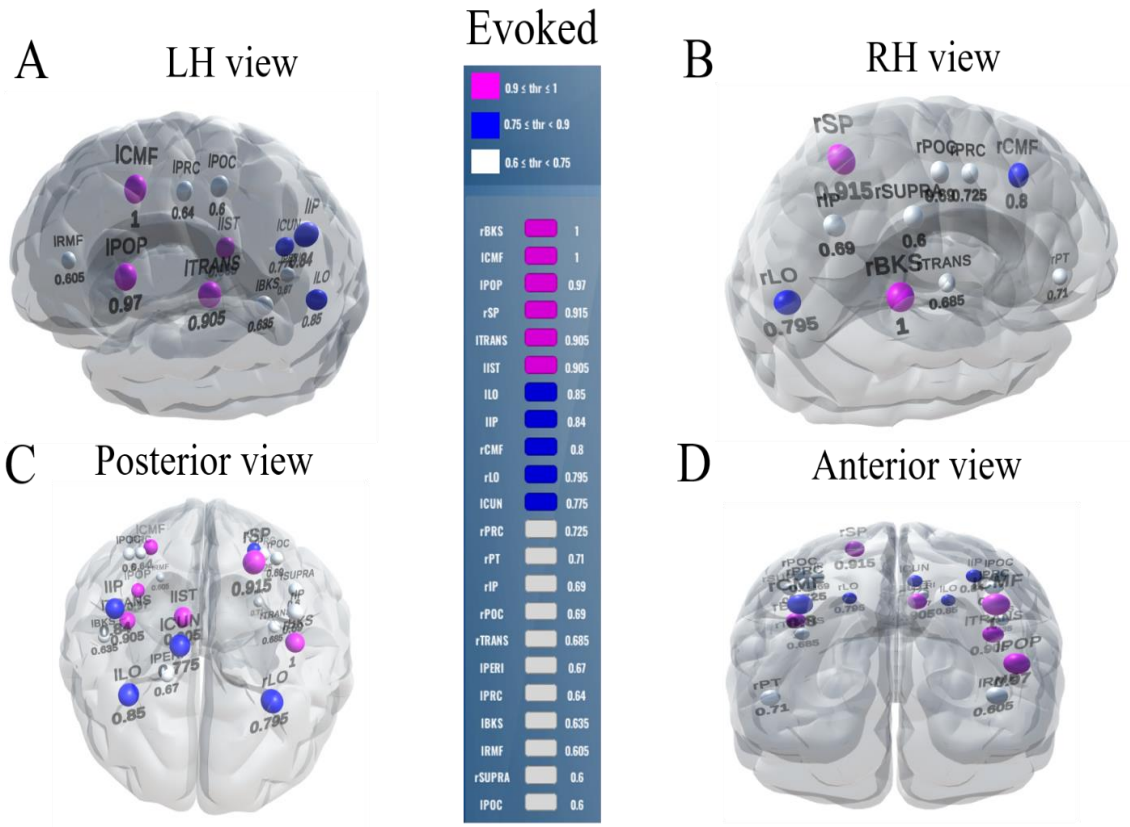

Figure 13: Stable (most consistent) neural network decoded using evoked activity. Visualization of brain ROIs corresponding to ≥ 0.60 stability threshold (22 top selected ROIs which decode Tk1/5 from Tk3 at 71.4%. Otherwise as in Figure 12. BKS, bankssts; CMF, caudal middle frontal; POP, pars opercularis; SP, superior parietal; TRANS, transverse temporal; IST, isthmus cingulate; LO, lateral occipital; IP, inferior parietal; CUN, cuneus; PRC, precentral; PT, pars triangularis; POC, postcentral; PERI, Pericalcarine; SUPRA, supra marginal.

Table 5: Brain-behavior relations of 14 brain ROIs in different frequency bands and behavioral prediction from the *induced* activity at a stability threshold ≥ 0.6 that yielded accuracy 86.5%.

Frequency Band and combined R^2	**ROI name**	**ROI abbrev.**	**Coefficient**	**p-value**	**Stability score**
θ, R^2 =0.807, p<0.00001	Pars opercularis L	lPOP	0.974	0.013	0.785
	Posterior cingulate L	lPCG	-1.759	0.001	0.760
	Caudal anterior cingulate R	rCAC	3.163	0.001	0.740

	Bankssts L	lBKS	0.645	0.249	0.865
α, R^2 =0.746,	Inferior parietal L	lIP	1.594	0.035	0.730
p<0.00001	Precentral L	lPRC	1.006	0.117	0.640
	Transverse temporal R	rTRANS	0.267	0.584	0.620
β, R^2 =0.876,	Rostral anterior		3.004	0.001	0.605
p<0.00001	cingulate L	lRAC			
	Lateral occipital R	rLO	0.768	0.804	0.805
	Lateral orbitofrontal R	rLOF	5.092	0.001	0.755
	Superior parietal R	rSP	16.472	0.004	0.740
	Caudal middle frontal R	rCMF	-3.243	0.188	0.740
γ, R^2 =0.915,	cuneus R	rCUN	-1.743	0.701	0.675
p<0.00001	Lateral orbitofrontal L	lLOF	0.709	0.553	0.625

Table 6: Brain-behavior relations of 22 brain ROIs in different frequency bands and behavioral slope prediction from the *evoked* activity at a stability threshold ≥ 0.6 that yielded accuracy 71.4%.

Frequency Band and combined R^2	**ROI name**	**ROI abbrev.**	**Coefficient**	**p-value**	**Stability score**
	Caudal middle frontal L	lCMF	-93.646	0.575	1
	Superior parietal R	rSP	-89.350	0.603	0.915
	Isthmus cingulate L	lIST	-46.527	0.671	0.905
θ, R^2 =0.349, p<0.0184	Lateral occipital L	lLO	-190.348	0.149	0.850

	Pars triangularis R	rPT	137.923	0.160	0.710
	Post central R	rPOC	180.073	0.015	0.690
	Rostral middle frontal L	lRMF	-69.220	0.326	0.605
	Post central L	lPOC	152.979	0.238	0.600
α, R^2 =0.863, p<0.00001	Bankssts R	rBKS	-64.139	0.775	1
	Transverse temporal L	lTRANS	62.583	0.729	0.905
	Inferior parietal L	lIP	986.399	0.027	0.840
	Caudal middle frontal R	rCMF	-254.140	0.338	0.800
	Inferior parietal R	rIP	-707.049	0.053	0.690
	Pericalcarine L	lPERI	-278.456	0.319	0.670
	Precentral L	lPRC	163.234	0.368	0.640
	Bankssts L	lBKS	947.797	0.0001	0.635
	Supra marginal R	rSUPRA	-466.985	0.107	0.600
β, R^2 =0.198, p<0.0184	Pars opercularis L	lPOP	475.923	0.240	0.970
	Lateral occipital R	rLO	-1119.991	0.157	0.795
	Precentral R	rPRC	1485.600	0.008	0.725

	Cuneus L	lCUN	-3730.107	0.160	0.775
γ, R^2 =0.604, p< 0.00001	Transverse temporal R	rTRANS	262.544	0.0001	0.685

5.2.4 Brain-behavior relationships

To examine the behavioral relevance of the brain ROIs identified in stability selection, we conducted the multivariate WLS regression (Ruppert & Wand, 1994). We conducted WLS between the individual frequency band features (i.e., evoked and induced) and the slopes of the behavioral identification functions (i.e., Figure 1B) which indexes the strength of listeners' CP. WLS regression for *induced* activity is shown in Table 5 and for *evoked* activity in Table 6. From the induced data, we found that γ frequency activity from 6 ROIs predicted behavior best among all other frequency R^2 = 0.915, p< 0.0001. Remarkably, only two brain regions (including PAC and Rostral anterior cingulate L) of β-band frequency could predict behavioral slopes (R^2 =0.876, p<0.00001). Except in the α frequency band, evoked activity was poorer at predicting behavioral CP.

5.3 Discussion

5.3.1 Speech categorization from evoked and induced activity

The present study aimed to examine which modes of brain activity and frequency bands of the EEG best decode speech categories and the process of categorization. Our results demonstrate that at the whole-brain level, evoked β-band oscillations robustly code (~70% accuracy) category structure of speech sounds. However, induced ɣ-band showed better performance, classifying speech categories at ~95% accuracy, better than all other induced frequency bands. Our data are consistent with notions that higher frequency bands

are associated with speech identification accuracy and carry information related to acoustic features and quality of speech representation (Yellamsetty & Bidelman, 2018). Our results also corroborate previous studies that suggest higher frequency channels of the EEG (β, γ) reflect auditory perceptual object construction (Tallon-Baudry & Bertrand, 1999) and how well listeners map vowel sounds to category labels (Bidelman, 2015, 2017).

Analysis by hemispheres showed that induced γ activity was dominant in LH whereas lower frequency band (e.g., θ) were more dominant in RH. These findings support the asymmetric engagement of frequency bands during syllable processing (Giraud et al., 2007; Morillon et al., 2012) and lower frequency band in RH dominance in inhibitory and attentional control (top-down processing during complex tasks) (Garavan et al., 1999; Price et al., 2019). Our results are consistent with the idea that cortical theta and gamma frequency bands play a key role in speech encoding (Hyafil et al., 2015). They also show that the machine learning model was able to decode acoustic-phonetic information (i.e., speech categories) in LH (using induced high frequency) and (using low frequencies) in RH.

5.3.2 Brain networks involved in speech categorization

Machine learning (stability selection coupled with SVM) further identified the most stable, relevant, and invariant brain regions that associate with speech categorization. Stability selection identified 22 and 14 critical brain ROIs using evoked and induced activity, respectively. Our results show that induced activity better characterizes speech categorization using less neural resources (i.e., fewer brain regions) as compared to evoked activity. Eight brain ROIs (e.g., bankssts L, lateral occipital R, pars opercularis L, superior parietal R, caudal middle frontal R, inferior parietal L, precentral L, transverse temporal R) are common in evoked and induced regimes. These eight areas included the primary auditory cortex (Transverse temporal R), Brocas's area (Pars opercularis L), and motor area (Precentral L) which are critical to speech-language processing. Superior parietal and inferior parietal areas

have been associated with auditory, phoneme, and sound categorization in particularly ambiguous contexts (Dufor et al., 2007; Feng et al., 2018). The non-overlapping areas in induced activity; orbitofrontal is associated with speech comprehension and rostral anterior cingulate with speech control (Sabri et al., 2008). Surprisingly, out of the identified 14 brain ROIs; three ROIs are in θ, three in α, two in β, and six in γ band.

Noticeably, we found that a greater number of brain regions were recruited in the γ-frequency band. This result is consistent with the notion that high-frequency oscillations play a role in network synchronization and widespread construction of perceptual objects related to abstract speech categories (Giraud & Poeppel, 2012; Haenschel et al., 2000; Si et al., 2017; Tallon-Baudry & Bertrand, 1999). Indeed, γ-band activity in only six ROIs were the best predictor of listeners' behavioral speech categorization. Interestingly, nine ROIs of evoked α-band activity were able to predict behavioral slopes better than induced α-band activity. This result supports notions that the α frequency band is associated with attention (Klimesch, 2012) and speech intelligibility (Dimitrijevic et al., 2017).

A main advantage of our data-driven approach is that it identifies the frequency bands and brain regions that are best linked to speech categorization behaviors from among the many thousands of features measurable from whole-brain EEG. It is a complement to conventional hypothesis-driven approaches (Bidelman, 2015; Bidelman & Alain, 2015a; Bidelman & Walker, 2019) but is perhaps more hands-off in that it requires fewer assumptions about the underlying brain mechanisms supporting speech perception. Additionally, our finding supports theoretical oscillatory models and empirical data (Doelling et al., 2019) that suggest induced activity can predict auditory-perceptual processing better than evoked activity. Nonetheless, our data suggest that induced neural activity plays a more prominent role in describing the perceptual-cognitive process of speech categorization than evoked modes of brain activity (Doelling et al., 2019). In particular, we demonstrate that among these two

prominent functional modes and frequency channels characterizing the EEG, induced γ-frequency oscillations best decode the category structure of speech and the strength of listeners' behavioral identification. In contrast, the evoked activity provides a reliable though weaker correspondence with behavior in all but the α frequency band.

Chapter 6 - Conclusion

In this chapter, we summarize the main findings and limitations of our works.

6.1 Summary of Contributions

We built data-driven frameworks for understanding the neural mechanisms underlying cognitive speech processing. Our spatiotemporal analysis demonstrates that early auditory ERPs could decode the speech sound over 95% accuracy. A smaller set of brain ROIs associate during encoding as compared to decision processing networks. Hemisphere-based analyses show that LH is dominating with earlier decoding ability. The selected 15 brain regions engage in the decision process that could robustly predict listeners' behavioral CP (i.e., the strength of listeners' categorical hearing) from neural data (e.g., ERPs). Moreover, the evoked vs. induced analysis demonstrates that higher induced frequency bands decode the speech sound best among all other frequency bands. Induced higher frequency band (gamma band) dominate in LH and lower frequency band (theta band) dominate in RH during speech categorization. Remarkably, only six brain ROIs' induced gamma-band activities were strongly associated with listeners' behavioral CP.

6.2 Limitation of this work

A disadvantage of this data-driven approach is that it is computationally expensive. In addition, our study only included vowel stimuli. Additional studies are required to examine if our findings generalize to other speech sounds (e.g., consonants) which elicit stronger/weaker categorical percepts or those which are more/less familiar to a listener (e.g., native vs. nonnative speech).

Relationship to published works

Chapter 4 -

1. Mahmud, M. S., Yeasin, M., & Bidelman, G. M. (2021). Speech categorization is better described by induced rather than evoked neural activity. *The Journal of the Acoustical Society of America*, *149*(3), 1644-1656.

Chapter 5 -

1. Mahmud, M. S., Yeasin, M., & Bidelman, G. M. (2021). Data-driven machine learning models for decoding speech categorization from evoked brain responses. *Journal of Neural Engineering*.

References

Alain, C. (2007). Breaking the wave: Effects of attention and learning on concurrent sound perception. *Hearing Research*, *229*(1–2), 225–236.

Al-Fahad, R., Yeasin, M., & Bidelman, G. M. (2020). Decoding of single-trial EEG reveals unique states of functional brain connectivity that drive rapid speech categorization decisions. *Journal of Neural Engineering*, *17*(1), 016045.

Alho, J., Green, B. M., May, P. J., Sams, M., Tiitinen, H., Rauschecker, J. P., & Jääskeläinen, I. P. (2016). Early-latency categorical speech sound representations in the left inferior frontal gyrus. *Neuroimage*, *129*, 214–223.

Bashivan, P., Bidelman, G. M., & Yeasin, M. (2014). Spectrotemporal dynamics of the EEG during working memory encoding and maintenance predicts individual behavioral capacity. *European Journal of Neuroscience*, *40*(12), 3774–3784.

Belin, P., McAdams, S., Thivard, L., Smith, B., Savel, S., Zilbovicius, M., Samson, S., & Samson, Y. (2002). The neuroanatomical substrate of sound duration discrimination. *Neuropsychologia*, *40*(12), 1956–1964.

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

Bidelman, G. M. (2017). Amplified induced neural oscillatory activity predicts musicians' benefits in categorical speech perception. *Neuroscience*, *348*, 107–113.

Bidelman, G. M., & Alain, C. (2015a). Hierarchical neurocomputations underlying concurrent sound segregation: Connecting periphery to percept. *Neuropsychologia*, *68*, 38–50.

Bidelman, G. M., & Alain, C. (2015b). Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *Journal of Neuroscience*, *35*(3), 1240–1249.

Bidelman, G. M., Bush, L., & Boudreaux, A. (2020). Effects of noise on the behavioral and neural categorization of speech. *Frontiers in Neuroscience*, *14*, 153.

Bidelman, G. M., & Howell, M. (2016). Functional changes in inter- and intra-hemispheric cortical processing underlying degraded speech perception. *NeuroImage*, *124*(Pt A), 581–590. https://doi.org/10.1016/j.neuroimage.2015.09.020

Bidelman, G. M., & Lee, C.-C. (2015). Effects of language experience and stimulus context on the neural organization and categorical perception of speech. *Neuroimage*, *120*, 191–200.

Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage*, *79*, 201–212.

Bidelman, G. M., & Walker, B. (2019). Plasticity in auditory categorization is supported by differential engagement of the auditory-linguistic network. *NeuroImage*, *201*, 116022.

Bidelman, G. M., & Walker, B. S. (2017). Attentional modulation and domain-specificity underlying the neural organization of auditory categorical perception. *European Journal of Neuroscience*, *45*(5), 690–699.

Bidelman, G. M., Weiss, M. W., Moreno, S., & Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *European Journal of Neuroscience*, *40*(4), 2662–2673.

Caplan, D. (1994). Language and the brain. *Academic Press*, 1023–1053.

Carter, J. A., & Bidelman, G. (2020). Auditory cortex is susceptible to lexical influence as revealed by informational vs. Energetic masking of speech categorization. *BioRxiv*.

Casale, S., Russo, A., Scebba, G., & Serrano, S. (2008). Speech Emotion Classification Using Machine Learning Algorithms. *2008 IEEE International Conference on Semantic Computing*, 158–165. https://doi.org/10.1109/ICSC.2008.43

Celsis, P., Doyon, B., Boulanouar, K., Pastor, J., Démonet, J.-F., & Nespoulous, J.-L. (1999). ERP correlates of phoneme perception in speech and sound contexts. *Neuroreport*, *10*(7), 1523–1527.

Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., & Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience*, *13*(11), 1428.

Cruz, J. A., & Wishart, D. S. (2006). Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics*, *2*, 117693510600200030.

David, O., Kilner, J. M., & Friston, K. J. (2006). Mechanisms of evoked and induced responses in MEG/EEG. *Neuroimage*, *31*(4), 1580–1591.

de Taillez, T., Kollmeier, B., & Meyer, B. T. (2020). Machine learning for decoding listeners' attention from electroencephalography evoked by continuous speech. *European Journal of Neuroscience*, *51*(5), 1234–1241.

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *Neuroimage*, *24*(1), 21–33.

Desai, R., Liebenthal, E., Waldron, E., & Binder, J. R. (2008). Left posterior temporal regions are sensitive to auditory categorization. *Journal of Cognitive Neuroscience*, *20*(7), 1174–1188.

Deschamps, I., Baum, S. R., & Gracco, V. L. (2014). On the role of the supramarginal gyrus in phonological processing and verbal working memory: Evidence from rTMS studies. *Neuropsychologia*, *53*, 39–46.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on

MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Dimitrijevic, A., Smith, M. L., Kadis, D. S., & Moore, D. R. (2017). Cortical alpha oscillations predict speech intelligibility. *Frontiers in Human Neuroscience*, *11*, 88.

Doelling, K. B., Assaneo, M. F., Bevilacqua, D., Pesaran, B., & Poeppel, D. (2019). An oscillator model better predicts cortical entrainment to music. *Proceedings of the National Academy of Sciences*, *116*(20), 10113–10121.

Domenech, P., & Dreher, J.-C. (2010). Decision threshold modulation in the human brain. *Journal of Neuroscience*, *30*(43), 14305–14317.

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2016). Increased activity in frontal motor cortex compensates impaired speech perception in older adults. *Nature Communications*, *7*, 12241. https://doi.org/10.1038/ncomms12241

Dufor, O., Serniclaes, W., Sprenger-Charolles, L., & Démonet, J.-F. (2007). Top-down processes during auditory phoneme categorization in dyslexia: A PET study. *Neuroimage*, *34*(4), 1692–1707.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, *32*(2), 407–499.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(3968), 303–306.

Eulitz, C., Maess, B., Pantev, C., Friederici, A. D., Feige, B., & Elbert, T. (1996). Oscillatory neuromagnetic activity induced by language and non-language stimuli. *Cognitive Brain Research*, *4*(2), 121–132.

Feng, G., Gan, Z., Wang, S., Wong, P. C., & Chandrasekaran, B. (2018). Task-general and acoustic-invariant neural representation of speech categories in the human brain. *Cerebral Cortex*, *28*(9), 3241–3254.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(4), 526.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1), 1–22.

Frost, J. A., Binder, J. R., Springer, J. A., Hammeke, T. A., Bellgowan, P. S., Rao, S. M., & Cox, R. W. (1999). Language processing is strongly left lateralized in both sexes: Evidence from functional MRI. *Brain*, *122*(2), 199–208.

Garavan, H., Ross, T. J., & Stein, E. A. (1999). Right hemispheric dominance of inhibitory control: An event-related functional MRI study. *Proceedings of the National Academy of Sciences*, *96*(14), 8301–8306.

Geiser, E., Zaehle, T., Jancke, L., & Meyer, M. (2008). The neural correlate of speech rhythm as evidenced by metrical speech processing. *Journal of Cognitive Neuroscience*, *20*(3), 541–552.

Giraud, A.-L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S., & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron*, *56*(6), 1127–1134.

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

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol*, *11*(12), e1001752.

Guenther, F. H., Nieto-Castanon, A., Ghosh, S. S., & Tourville, J. A. (2004). Representation of sound categories in auditory cortical maps. *Journal of Speech, Language, and Hearing Research*.

Haenschel, C., Baldeweg, T., Croft, R. J., Whittington, M., & Gruzelier, J. (2000). Gamma and beta frequency oscillations in response to novel auditory stimuli: A comparison of

human electroencephalogram (EEG) data with in vitro models. *Proceedings of the National Academy of Sciences*, *97*(13), 7645–7650.

Hampshire, A., Chamberlain, S. R., Monti, M. M., Duncan, J., & Owen, A. M. (2010). The role of the right inferior frontal gyrus: Inhibition and attentional control. *Neuroimage*, *50*(3), 1313–1319.

Hanley, J. A. (1983). Appropriate uses of multivariate analysis. *Annual Review of Public Health*, *4*(1), 155–180.

Herrmann, C. S., Rach, S., Vosskuhl, J., & Strüber, D. (2014). Time–frequency analysis of event-related potentials: A brief tutorial. *Brain Topography*, *27*(4), 438–450.

Hickok, G., Costanzo, M., Capasso, R., & Miceli, G. (2011). The role of Broca's area in speech perception: Evidence from aphasia revisited. *Brain and Language*, *119*(3), 214–220.

Hickok, G., Erhard, P., Kassubek, J., Helms-Tillery, A. K., Naeve-Velguth, S., Strupp, J. P., Strick, P. L., & Ugurbil, K. (2000). A functional magnetic resonance imaging study of the role of left posterior superior temporal gyrus in speech production: Implications for the explanation of conduction aphasia. *Neuroscience Letters*, *287*(2), 156–160.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, *4*(4), 131–138.

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

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception, & Psychophysics*, *72*(5), 1218–1227.

Hsu, C.-W., Chang, C.-C., & Lin, C. J. (2003). A practical guide to support vector classification technical report department of computer science and information engineering. *National Taiwan University, Taipei*.

Hull, R., & Vaid, J. (2006). Laterality and language experience. *Laterality*, *11*(5), 436–464.

Husain, F. T., Fromm, S. J., Pursley, R. H., Hosey, L. A., Braun, A. R., & Horwitz, B. (2006). Neural bases of categorization of simple speech and nonspeech sounds. *Human Brain Mapping*, *27*(8), 636–651.

Hyafil, A., Fontolan, L., Kabdebon, C., Gutkin, B., & Giraud, A.-L. (2015). Speech encoding by coupled cortical theta and gamma oscillations. *Elife*, *4*, e06213.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Increased brain activity in frontal and parietal cortex underlies the development of visuospatial working memory capacity during childhood. *Journal of Cognitive Neuroscience*, *14*(1), 1–10.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608.

Lee, Y.-S., Turkeltaub, P., Granger, R., & Raizada, R. D. (2012). Categorical speech processing in Broca's area: An fMRI study using multivariate pattern-based analysis. *Journal of Neuroscience*, *32*(11), 3942–3948.

Liberman, A. M., Isenberg, D., & Rakerd, B. (1981). Duplex perception of cues for stop consonants: Evidence for a phonetic mode. *Perception & Psychophysics*, *30*(2), 133–143.

Liebenthal, E., Desai, R., Ellingson, M. M., Ramachandran, B., Desai, A., & Binder, J. R. (2010). Specialization along the left superior temporal sulcus for auditory categorization. *Cerebral Cortex*, *20*(12), 2958–2970.

Loui, P. (2015). A dual-stream neuroanatomy of singing. *Music Perception: An Interdisciplinary Journal*, *32*(3), 232–241.

Luck, S. J. (2005). *An introduction to the event-related potential technique (pp. 45–64)*. Cambridge, Ma: MIT press.

Luo, H., & Poeppel, D. (2012). Cortical oscillations in auditory perception and speech: Evidence for two temporal windows in human auditory cortex. *Frontiers in Psychology*, *3*, 170.

Mahmud, M. S., Ahmed, F., Al-Fahad, R., Moinuddin, K. A., Yeasin, M., Alain, C., & Bidelman, G. (2020). Decoding hearing-related changes in older adults' spatiotemporal neural processing of speech using machine learning. *Frontiers in Neuroscience*, 1–14.

Mahmud, M. S., Yeasin, M., & Bidelman, G. M. (2021). Data-driven machine learning models for decoding speech categorization from evoked brain responses. *Journal of Neural Engineering*, *18*(4), 046012.

Mankel, K., Barber, J., & Bidelman, G. M. (2020). Auditory categorical processing for speech is modulated by inherent musical listening skills. *NeuroReport*, *31*(2), 162–166.

Martin, S., Brunner, P., Holdgraf, C., Heinze, H.-J., Crone, N. E., Rieger, J., Schalk, G., Knight, R. T., & Pasley, B. N. (2014). Decoding spectrotemporal features of overt and covert speech from the human cortex. *Frontiers in Neuroengineering*, *7*, 14.

Masmoudi, S., Dai, D. Y., & Naceur, A. (2012). *Attention, representation, and human performance: Integration of cognition, emotion, and motivation*. Psychology Press.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.

Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*(4), 417–473. https://doi.org/10.1111/j.1467-9868.2010.00740.x

Menon, V., & Desmond, J. E. (2001). Left superior parietal cortex involvement in writing: Integrating fMRI with lesion evidence. *Cognitive Brain Research*, *12*(2), 337–340.

Miller, C. T., & Cohen, Y. E. (2010). Vocalization processing. *Primate Neuroethology*, 237–255.

Miller, E. K., Freedman, D. J., & Wallis, J. D. (2002). The prefrontal cortex: Categories, concepts and cognition. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *357*(1424), 1123–1136.

Miller, E. K., Nieder, A., Freedman, D. J., & Wallis, J. D. (2003). Neural correlates of categories and concepts. *Current Opinion in Neurobiology*, *13*(2), 198–203.

Moinuddin, K. A., Yeasin, M., & Bidelman, G. M. (2019). *BrainO*. https://github.com/cvpia-uofm/BrainO

Molfese, D., Key, A. P. F., Maguire, M., Dove, G. O., & Molfese, V. J. (2005). Event-related evoked potentials (ERPs) in speech perception. *The Handbook of Speech Perception*, *99121*.

Morillon, B., Liégeois-Chauvel, C., Arnal, L. H., Bénar, C. G., & Giraud, A.-L. (2012). Asymmetric function of theta and gamma activity in syllable processing: An intra-cortical study. *Frontiers in Psychology*, *3*, 248.

Mostert, P., Kok, P., & De Lange, F. P. (2015). Dissociating sensory from decision processes in human perceptual decision making. *Scientific Reports*, *5*, 18253.

Myers, E. B., & Blumstein, S. E. (2008). The neural bases of the lexical effect: An fMRI investigation. *Cerebral Cortex*, *18*(2), 278–288.

Noe, C., & Fischer-Baum, S. (2020). Early lexical influences on sublexical processing in speech perception: Evidence from electrophysiology. *Cognition*, *197*, 104162.

Nogueira, S., Sechidis, K., & Brown, G. (2017). On the Stability of Feature Selection Algorithms. *Journal of Machine Learning Research*, *18*, 174–1.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*, 299–325.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2010). Broca's area and language processing: Evidence for the cognitive control connection. *Language and Linguistics Compass*, *4*(10), 906–924.

Nyberg, L., Marklund, P., Persson, J., Cabeza, R., Forkstam, C., Petersson, K. M., & Ingvar, M. (2003). Common prefrontal activations during working memory, episodic memory, and semantic memory. *Neuropsychologia*, *41*(3), 371–377.

Oberhuber, M., Hope, T. M. H., Seghier, M. L., Parker Jones, O., Prejawa, S., Green, D. W., & Price, C. J. (2016). Four functionally distinct regions in the left supramarginal gyrus support word processing. *Cerebral Cortex*, *26*(11), 4212–4226.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113.

Ou, J., & Law, S.-P. (2018). Induced gamma oscillations index individual differences in speech sound perception and production. *Neuropsychologia*, *121*, 28–36.

Park, Y., Luo, L., Parhi, K. K., & Netoff, T. (2011). Seizure prediction with spectral power of EEG using cost-sensitive support vector machines. *Epilepsia*, *52*(10), 1761–1770. https://doi.org/10.1111/j.1528-1167.2011.03138.x

Paus, T., Petrides, M., Evans, A. C., & Meyer, E. (1993). Role of the human anterior cingulate cortex in the control of oculomotor, manual, and speech responses: A positron emission tomography study. *Journal of Neurophysiology*, *70*(2), 453–469.

Perlovsky, L. (2011). Language and cognition interaction neural mechanisms. *Computational Intelligence and Neuroscience*, *2011*.

Picton, T. W., van Roon, P., Armilio, M. L., Berg, P., Ille, N., & Scherg, M. (2000). The correction of ocular artifacts: A topographic perspective. *Clinical Neurophysiology*, *111*(1), 53–65. https://doi.org/10.1016/S1388-2457(99)00227-8

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, *15*(2), 285–290.

Price, C. N., Alain, C., & Bidelman, G. M. (2019). Auditory-frontal Channeling in α and β Bands is Altered by Age-related Hearing Loss and Relates to Speech Perception in Noise. *Neuroscience*, *423*, 18–28.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718.

Royston, P., & Sauerbrei, W. (2008). *Multivariable model-building: A pragmatic approach to regression anaylsis based on fractional polynomials for modelling continuous variables* (Vol. 777). John Wiley & Sons.

Ruppert, D., & Wand, M. P. (1994). Multivariate locally weighted least squares regression. *The Annals of Statistics*, 1346–1370.

Russ, B. E., Lee, Y.-S., & Cohen, Y. E. (2007). Neural and behavioral correlates of auditory categorization. *Hearing Research*, *229*(1–2), 204–212.

Sabri, M., Binder, J. R., Desai, R., Medler, D. A., Leitl, M. D., & Liebenthal, E. (2008). Attentional and linguistic interactions in speech perception. *Neuroimage*, *39*(3), 1444–1456.

Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D., & Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science*, *326*(5951), 445–449.

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PloS One*, *10*(3), e0118432.

Schneiders, J., Opitz, B., Tang, H., Deng, Y., Xie, C., Li, H., & Mecklinger, A. (2012). The impact of auditory working memory training on the fronto-parietal working memory network. *Frontiers in Human Neuroscience*, *6*, 173.

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. *Proceedings of the 9th Python in Science Conference*, *57*, 61.

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

Shen, G., & Froud, K. (2019). Electrophysiological correlates of categorical perception of lexical tones by English learners of Mandarin Chinese: An ERP study. *Bilingualism*, *22*(2), 253–265.

Si, X., Zhou, W., & Hong, B. (2017). Cooperative cortical network for categorical processing of Chinese lexical tone. *Proceedings of the National Academy of Sciences*, *114*(46), 12303–12308.

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience*, *2011*, 8.

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, *3*(4), 151–162.

Tankus, A., Fried, I., & Shoham, S. (2012). Structured neuronal encoding and decoding of human speech features. *Nature Communications*, *3*(1), 1–5.

Tervaniemi, M., & Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Research Reviews*, *43*(3), 231–246.

Toscano, J. C., Anderson, N. D., Fabiani, M., Gratton, G., & Garnsey, S. M. (2018). The time-course of cortical responses to speech revealed by fast optical imaging. *Brain and Language*, *184*, 32–42.

Tsunada, J., & Cohen, Y. E. (2014). Neural mechanisms of auditory categorization: From across brain areas to within local microcircuits. *Frontiers in Neuroscience*, *8*, 161.

Tzourio, N., Crivello, F., Mellet, E., Nkanga-Ngila, B., & Mazoyer, B. (1998). Functional anatomy of dominance for speech comprehension in left handers vs right handers. *Neuroimage*, *8*(1), 1–16.

Von Stein, A., & Sarnthein, J. (2000). Different frequencies for different scales of cortical integration: From local gamma to long range alpha/theta synchronization. *International Journal of Psychophysiology*, *38*(3), 301–313.

Weighted Regression in SAS, R, and Python. (n.d.). Retrieved May 27, 2020, from https://jbhender.github.io/Stats506/F17/Projects/Abalone_WLS.html

Whitwell, J. L., Duffy, J. R., Strand, E. A., Xia, R., Mandrekar, J., Machulda, M. M., Senjem, M. L., Lowe, V. J., Jack Jr, C. R., & Josephs, K. A. (2013). Distinct regional anatomic and functional correlates of neurodegenerative apraxia of speech and aphasia: An MRI and FDG-PET study. *Brain and Language*, *125*(3), 245–252.

Wood, C. C., Goff, W. R., & Day, R. S. (1971). Auditory evoked potentials during speech perception. *Science*, *173*(4003), 1248–1251.

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America*, *120*(2), 1063–1074.

Yellamsetty, A., & Bidelman, G. M. (2018). Low-and high-frequency cortical brain oscillations reflect dissociable mechanisms of concurrent speech segregation in noise. *Hearing Research*, *361*, 92–102.

Yin, Q.-Y., Li, J.-L., & Zhang, C.-X. (2017). Ensembling Variable Selectors by Stability Selection for the Cox Model. *Computational Intelligence and Neuroscience*, *2017*. https://doi.org/10.1155/2017/2747431

Youssofzadeh, V., Stout, J., Ustine, C., Gross, W. L., Conant, L. L., Humphries, C. J., Binder, J. R., & Raghavan, M. (2020). Mapping language from MEG beta power modulations during auditory and visual naming. *NeuroImage*, *220*, 117090.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, *256*(5058), 846–849.