The Thesis committee for Garret Kurteff
Certifies that this is the approved version of the following thesis:

Modulation of Neural Responses to Naturalistic Speech Production and Perception

APPROVED BY

SUPERVISING COMMITTEE:

J. Liberty Hamilton, Supervisor

Rosemary Lester-Smith

Modulation of Neural Responses to Naturalistic Speech Production and Perception

by

Garret Kurteff

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE IN SPEECH, LANGUAGE, AND HEARING SCIENCES

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2020

Dedicated to Willow and Cherry

Acknowledgments

I acknowledge the participants who graciously volunteered their time to participate in my study. I acknowledge the many undergraduate research volunteers that made linguistic analysis of this dataset possible through their transcription work: Nicole Currens, Jade Holder, Claire Huber, Amanda Martinez, Valerie Mercado, Paranjaya Pokharel, Christopher Truong, and Cassandra Villarreal. I acknowledge graduate students Maansi Desai and Mary Lowery and research assistant Ian Griffith for their assistance in data collection and analysis. I acknowledge Dr. Rosemary Lester-Smith for her assistance in EMG electrode placement and her contributions to my literature review. Lastly, I acknowledge Dr. Liberty Hamilton for her unparalleled support as a mentor throughout my first research project in my PhD program.

Modulation of Neural Responses to Naturalistic Speech Production and Perception

Garret Kurteff, M.S.S.L.H.S.
The University of Texas at Austin, 2020

Supervisor: J. Liberty Hamilton

Speech production is under-studied compared to speech perception largely due to complications in data collection caused by articulation. In electroencephalography (EEG), these complications manifest as electromyographic activity (EMG) originating from the muscles that control articulation (Chen et al. 2019). This is unfortunate because EEG is well-suited for studying the rapid temporal changes in speech production. In addition, the few EEG studies of speech production are limited to the single-word level, which limits the generalizability of studies to how speech is used in everyday contexts.

In this thesis I present an EEG study of the differences between speech production and perception using sentence-level naturalistic stimuli. Participants overtly produced sentences from the MOCHA-TIMIT (Wrench 1999) corpus then listened to playback of themselves producing the sentences. Perception trials were then split into *predictable* and *unpredictable* trials. Predictable trials consisted of playback of the previously produced sentence, while

unpredictable trials consisted of playback of a randomly selected previously produced sentence. In this thesis, two contrasts are compared: (1) overt production of sentences versus passive listening to sentences, and (2) passive listening to predictable sentences versus passive listening to unpredictable sentences. Canonical correlation analysis (CCA) was used to remove EMG artifact from the recorded EEG.

To demonstrate removal of EMG and preservation of neural responses in CCA-corrected EEG data, event-related potential (ERP) analysis was used on neural responses to perception stimuli, inter-trial click tones, and activity recorded from auxiliary facial EMG electrodes. These ERP analyses revealed a reduction in amplitude for production trials and facial EMG activity after CCA artifact correction and a preservation of early auditory responses in inter-trial click tones, suggesting that EMG was successfully removed while preserving neural responses. After validation of EMG removal, perception and production trials were compared using ERP analysis. Responses to produced sentences were found to have reduced amplitude when compared with perceived sentences, which is consistent with previous research on speaker-induced suppression. Differences in stimulus predictability during speech perception had an effect on response amplitude as well; however, this difference was weaker than the difference in amplitudes observed while comparing the differences between perception and production trials. Multivariate temporal receptive field modeling was used to examine phonological tuning in perception and production. Models demonstrated that speaker-induced suppression does not reflect

a change in neural encoding of phonological features but instead a generalized reduction in response amplitude during speech production. Understanding the differences between speech perception and production in a naturalistic context has implications for developing brain-computer interfaces and understanding the neural basis of communication disorders such as apraxia of speech and stuttering. This thesis also serves as a proof-of-concept for studying sentence-level speech production using EEG by demonstrating an effective way of removing EMG artifact while preserving integrity of neural responses.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The neuroscience of language is frequently studied in heavily constrained experiments that bear little immediate connection to the language we use in our daily lives. While such studies have contributed to our understanding of how language works in the brain, the desire to specifically study more "naturalistic" language as it is represented in the brain is strong (Hamilton & Huth 2020). The objective of this thesis is to expand the study of speech production using naturalistic stimuli.

There is an additional degree of separation between the study of language and how it used in daily life: language research is often restricted to speech perception. While there are definitely overlaps in neural representation between speech *perception* and speech *production* (Wilson et al. 2004; D'Ausilio et al. 2009; Meister et al. 2007; Watkins et al. 2003), the processes are often studied in isolation. This separation of perception and production is likely reinforced by the propagation of neurobiological models of language that emphasize this dichotomy (Broca 1861; Wernicke 1874; Hickok & Poeppel 2007; Tremblay & Dick 2016). Additionally, it can be difficult to compare production and perception responses in a single experiment, as different stimuli

are usually used to study the two independently.

This dichotomy has led to speech production being comparatively understudied, as researchers must consider additional methodological constraints when attempting to study speech production. Every widely used noninvasive neuroimaging technique is impacted by head movement during image acquisition (Jiang et al. 2019; Burgess 2020; Friston et al. 1996). Because speech production involves movement of speech articulators (tongue, lips, jaw, etc.), head movement is a fundamental component of speech production. To ensure that reliable data are acquired, many researchers have used methods of studying speech production that are additional degrees of abstraction away from natural speech, such as covert speech, where the movement of articulators is imagined instead of executed (Shuster 2003; Okada et al. 2018). Another technique for avoiding movement is to acquire imaging data before articulation begins or after articulation has completed, as examining time windows where articulation is not actively taking place can prevent the influence of movement caused by articulation (Singh et al. 2018). A common characteristic of these methods is that they do not directly examine the speech production that occurs in everyday scenarios. In order to study speech production in a naturalistic context, methods for dealing with the recording errors caused by articulatory movement must be developed.

1.1 Speech Production: What We Know

Although the neuroscience of speech production is relatively understudied, especially when using naturalistic stimuli, there are well-developed models of the neurobiology of speech production supported by neuroscientific results (Perkell et al. 1997; Tourville & Guenther 2011; Parrell et al. 2019; Houde & Chang 2015). A universal inclusion in these models is the concept of sensorimotor control of speech, which describes the mechanisms by which a speaker can detect and correct errors while speaking. This is accomplished in part through *feedback* (corrective) control, a term used in speech production models such as Directions Into Velocities of Articulators (Tourville & Guenther 2011) (DIVA) to refer to mechanisms for detecting changes in real time between predicted and actual somatosensory and auditory consequences of speech production. This is in contrast with *feedforward* (predictive) control, which contains motor programs of speech production that are updated in real-time by error correction signals from the feedback system.

The expected sensory consequence of the control system is often referred to as the *efference copy*, and neuropsychological evidence for this mechanism is found in studies that use altered auditory feedback to create mismatches between the expected and perceived sensory consequences of speech (Hawco et al. 2009; Hashimoto & Sakai 2003; Zheng et al. 2010; Behroozmand & Larson 2011; Greenlee et al. 2013), usually through altering the fundamental frequency (f_o) or the first formant (F_1) of produced speech. The way speakers automatically respond to altered auditory feedback differs based on the de-

gree of perturbation. For small changes in feedback, participants *oppose* the perturbation by adjusting their f_o/F_1 in the opposite direction of the perturbation (Burnett et al. 1998; Greenlee et al. 2011; Keough et al. 2013). For larger changes in feedback, participants *follow* the perturbation by adjusting their f_o/F_1 in the same direction of the perturbation (Burnett et al. 1998; Houde 1998). For example, in a study that changed feedback by altering f_o, Burnett et al. (1998) found the proportion of "opposing" responses decreased and the proportion of "following" responses increased as f_o perturbation increased from 25 cents to 300 cents. However, Hawco et al. (2009) found that the magnitude of correction to f_o perturbation decreased for changes above 200 cents, which they explain by theorizing that larger changes in f_o are no longer perceived as internally generated, making them less relevant to the correction mechanisms described above.

Other behavioral parameters can also influence subjects' responses to feedback perturbation, which gives us insight into what characteristics of speech modulate the control system. Lester-Smith et al. (2020) altered f_o and F_1 individually during speaking and listening, and also in predictable and unpredictable blocks. In unpredictable blocks of perturbation, f_o and F_1 were shifted suddenly in random trials, while in predictable blocks f_o and F_1 were gradually shifted over the course of consecutive trials. The authors found that participants were able to adapt to and correct shifted f_o and F_1 in predictable blocks and had difficulty correcting for unpredictable feedback in f_o and F_1. Furthermore, in F_1 but not f_o, participants' degrees of response to

unpredictable alteration was correlated with both degree of correction in the predictable perturbation block and also acuity to changes in F_1 during passive listening. The differences in feedback correction in predictable and unpredictable contexts presented in this study demonstrate that the predictability of a perturbation plays a role in an individual's ability to correct that perturbation, and the differences in how f_o and F_1 respond to the parameters of the experiment suggests that there may be disparate feedback mechanisms for the voice (which controls f_o) and the articulatory tract (which controls F_1).

The feedforward and feedback systems are complicated and still not fully understood, but the incorporation of auditory feedback during speech production provides a fundamental link between speech perception and production. Additionally, feedforward and/or feedback responses to altered auditory feedback can be disrupted in various disorders including schizophrenia (Heinks-Maldonado et al. 2007; McGuire et al. 1995; Woodruff et al. 1997), apraxia of speech (Ballard et al. 2018), stuttering (Daliri et al. 2018), dyslexia (van den Bunt et al. 2017), and neurodegenerative disorders such as Parkinson's disease (Hoffman 2014; Parrell et al. 2017).

1.1.1 Perceptual Responses to Speech Production are Suppressed

While some aspects of speech perception occur during speech production, there are some key differences in how these responses differ to speech perception in the absence of speaking. While imaging studies have demonstrated that regions of the temporal lobe associated with speech perception

are active during production, the responses in these regions during production are relatively suppressed compared to those during pure speech perception (Martikainen et al. 2005; Brumberg & Pitt 2019), a phenomenon known as *speaker-induced suppression* (SIS). The exact neural mechanisms behind SIS are not well understood. Changes in neural responses can be traced to specific neural components in EEG and MEG studies such as the N100(m) (Brumberg & Pitt 2019; Martikainen et al. 2005). Previous MEG research has suggested that SIS does not reflect a general suppression of the auditory cortex during speaking (Houde & Nagarajan 2011), which is supported by models such as DIVA that posit sensory feedback as an important component of motor speech control (Tourville & Guenther 2011). Because SIS does not represent a general suppression of auditory cortex activity, it most likely involves the suppression of specific components of speech perception (for example phonological feature encoding (Mesgarani et al. 2014)) that are not necessarily involved in feedback control of speech. That being said, which components of speech perception are suppressed during speaking are unknown. Studies examining the effect of altered auditory feedback on SIS have found SIS is absent when auditory feedback does not match the speaker's expectation (i.e., efference copy) for the feedback (Heinks-Maldonado et al. 2006; Niziolek et al. 2013).

Speaker-induced suppression provides an example of how perception and production systems interact while speaking; however, many aspects of this interaction are not well-documented. The primary region involved in perceiving speech is the posterior superior temporal gyrus (pSTG), which has

been shown to encode linguistic information about speech, such as phonological features (Mesgarani et al. 2014). The pSTG has also been implicated in processing auditory feedback during speaking alongside the superior parietal temporal (Spt) area (Houde & Nagarajan 2011; Chang et al. 2013).

Different regions of the cortex involved in speech production also behave differently in response to changes in predictability. Unpredictable feedback situations such as altered auditory feedback can result in an *increase* of activity in the pSTG, Spt and right inferior frontal/premotor regions (Heinks-Maldonado et al. 2006; Tourville et al. 2008). The middle temporal gyrus (MTG), on the other hand, shows a *decrease* in activity during altered auditory feedback (Zheng et al. 2010; Gauvin et al. 2016). That is, MTG activity is suppressed during altered auditory feedback while pSTG, Spt and right inferior frontal/premotor region activity is suppressed during normal speaking. While the MTG and pSTG/Spt are activated in seemingly competing patterns of suppression during speaking, it is possible that these processes work in tandem with each other: regions exhibiting *suppressive* behavior in unpredictable contexts (MTG) could be filtering external stimuli so that regions that exhibit *amplifying* behavior in unpredictable contexts (pSTG, STG) can properly attenuate and correct errors in speech production (Houde et al. 2002; Chang et al. 2013).

1.2 Predictability

As discussed above, a proposed bridge between the phenomenon of speaker-induced suppression and altered auditory feedback studies is predictability. Speaker-induced suppression appears to only function in predictable contexts, such as errorless speech production. Using fMRI, Okada et al. (2018) found a repetition suppression effect in the left inferior frontal gyrus was modulated by phonetic similarity in dyads of monosyllabic words: phonetically similar dyads demonstrated a greater degree of suppression. This result suggests predictability can affect the degree of speaker-induced suppression, possibly due to the predictive nature of feedforward control of speech.

In auditory perturbation studies, altering auditory feedback causes the perceptual component of speech production to become unpredictable, and thus responses are not suppressed so that the errorful production may be corrected. Stimulus predictability has been observed as a modulator of neural activity in speech perception experiments, where unpredictable stimuli result in an increased neural response known as *mismatch negativity* (Fitzgerald & Todd 2020; Bishop & Hardiman 2010; Hawco et al. 2009; Näätänen et al. 2007) (MMN). MMN is believed to be a distinct neural component that occurs 150-200ms after stimulus onset (Hawco et al. 2009; Näätänen et al. 2007). Given its polarity and temporal proximity to the N100 (see §1.3.1), the components are often compared but are believed to be different. For example, Hawco et al. observed MMN but no N100 response in an f_{o} perturbation task (Hawco et al. 2009). Additional studies have shown that perceiving a predictable au-

ditory stimulus will result in relative suppression of responses compared to an unpredictable auditory stimulus in lower-level auditory processing (Fitzgerald & Todd 2020; Bishop & Hardiman 2010) as well as in speech perception (Astheimer & Sanders 2011; Bendixen et al. 2014).

Error correction mechanisms provide a theoretical link between stimulus predictability and speech production. For speech perception, a theoretical motivation for the importance of stimulus predictability is speech segmentation. Speech is a continuous signal that needs to be segmented into smaller units of representation, such as sentences/words/phonemes, for proper comprehension by the listener (Giraud & Poeppel 2012; Bahl et al. 1983). The predictability of syllable sequences is used in the transitional probability theory of language acquisition, where unpredictable syllable sequences that are statistically less likely to occur in a language are used to identify word boundaries by infants acquiring a language (Saffran et al. 1996). In adults, the superior temporal gyrus (STG) has been functionally parcellated into "onset" and "sustained" response profiles (Hamilton et al. 2018), with onset responses occurring at sound edges. While onset responses occur regardless of linguistic content, their involvement in acoustic edge detection could implicate them in the speech segmentation process and therefore predictability. This suggests onset responses to speech could differ between perception and production; that is, onset responses present during speech perception could be suppressed during production.

1.3 Using Electroencephalography (EEG) to Study Speech Production

Changes in articulation and linguistic content (i.e., phonemes) occur in a very rapid timescale (around 50ms (Chang 2015)). While the anatomical specificity of fMRI makes the method appealing, fMRI measures hemodynamic response and can often take multiple seconds to acquire an image. The rapid temporal resolution of electrophysiologal methods such as noninvasive EEG and invasive electrocorticography (ECoG) makes them well-suited for studying speech production.

1.3.1 Event-Related Potentials and Components

The event-related potential (ERP) technique is a popular method for analyzing data with EEG (Luck 2014). In an ERP study, changes in EEG amplitude are analyzed relative to an event of interest to the researcher. For example, a researcher interested in studying the phoneme /m/ could average all trials relative to the onset of /m/ to create an event-related potential for the EEG response to /m/. The process of obtaining ERP data by timelocking to events is referred to as *epoching*.

ERP *components* are observed positive and negative deflections in the EEG response (Luck 2014; Luck, Steven J., & Kappenman, Emily S. 2011). Components have a highly predictable timecourse and are well-documented to generalize across studies. Two components of interest to this thesis are the N100 and P200 components, which have been associated with early automatic

responses to auditory stimuli (Lightfoot 2016; Lijffijt et al. 2009). The N100 is a negative deflection that occurs around 100ms (80-150ms (Lijffijt et al. 2009)) after stimulus onset, and the P200 is a positive deflection that occurs around 200ms (150-250ms (Lijffijt et al. 2009)) after stimulus onset. These two components are often grouped together as the "N1-P2 complex." The N1-P2 complex has been shown previously to be modulated by the predictability of the auditory stimulus (Hawco et al. 2009; Martikainen et al. 2005; Lijffijt et al. 2009) and is believed to be suppressed during speaker-induced suppression (Brumberg & Pitt 2019; Martikainen et al. 2005). The N100 component is also theorized as a neural marker of the efference copy (Brumberg & Pitt 2019; Heinks-Maldonado et al. 2007).

1.3.2 EEG Artifacts

While EEG's high temporal resolution makes it appealing for the study of speech production, there are a number of potential confounds that manifest in the form of artifacts. An *artifact* is any extraneous signal recorded by an EEG electrode that does not originate from the neurons the electrode intends to record from (Islam et al. 2016). The techniques used to remove artifact from EEG signal can be classified as either artifact *rejection* or artifact *correction* (Luck 2014). Artifact rejection is the complete exclusion from analysis of trials contaminated with artifact and, while effective, has the potential drawback of substantially reducing the number of trials in an analysis, possibly affecting the statistical power. Artifact correction allows for trials with artifact to be

included in the analysis by subtracting voltage from the EEG signal to "zero out" artifacts. The drawback of artifact correction is that it can lead to Type I (false positives) or Type II (false negatives) error in the "cleaned" signal. A visualization of all the common types of EEG artifact is present in Figure 1.1.

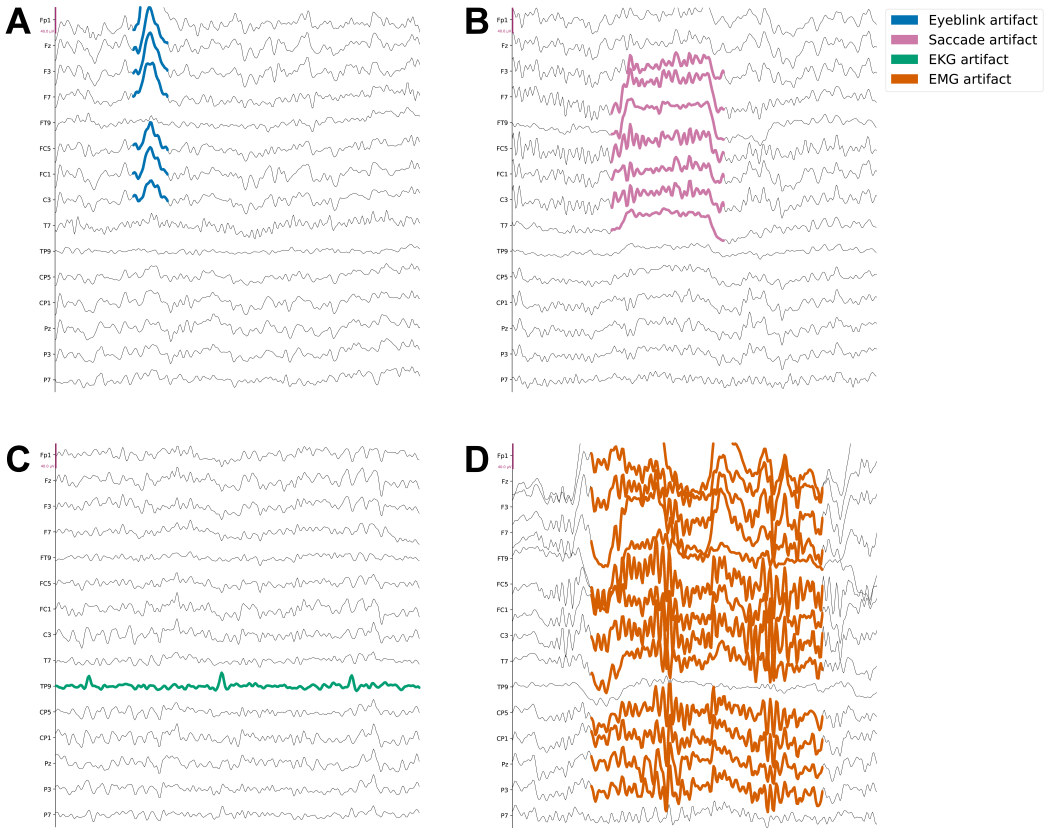

Figure 1.1: Common artifacts displayed in single-subject plots of raw EEG data.

1.3.2.1 Electrooculographic (EOG) Artifact

Because EEG records scalp activity, eye movement is a potential cause of artifact. EOG takes two primary forms depending on the direction of eye

movement: (1) blink artifacts (vertical EOG) and (2) saccade artifacts (horizontal EOG) (Berg & Scherg 1991; Makeig & Others 1996; Keren et al. 2010). Both blinks and saccades occur spontaneously and automatically. Volitional blinks can be minimized by instructing the participant to blink as a little as possible. The removal/correction of EOG artifacts is a well-documented procedure (Gomez-Herrero et al. 2006; Dimigen 2020; Gao et al. 2009; Keren et al. 2010; Jiang et al. 2019; Islam et al. 2016) and most software suites for analysis of EEG data (e.g., BrainVision Analyzer, EEGLab, MNE) come bundled with blind source separation techniques for isolation and removal of EOG artifact. For most EEG researchers who do not use complex visual stimuli that would elicit an increased amount of EOG artifact, correction of EOG is trivial.

1.3.2.2 Electrocardiographic (EKG) Artifact

EKG artifact (sometimes referred to as ECG) is another spontaneous and automatic source of artifact, generated by the heartbeat (Barlow & Dubinsky 1980; Tamburro et al. 2019). It is visible in the EEG as rhythmic, transient spikes at a frequency that roughly corresponds with heart rate. The removal of EKG is also well-documented, and the majority of blind source separation techniques used to remove EOG (such as Independent Component Analysis (ICA)) can be used to remove EKG as well (Tamburro et al. 2019; Jiang et al. 2019; Barlow & Dubinsky 1980; Islam et al. 2016).

1.3.2.3 Electromyographic (EMG) Artifact

EMG artifact is a broad category of artifact and is used to describe any activity recorded by the EEG electrodes originating from a muscle. In the context of speech production research, EMG refers to activity originating in muscles specifically associated with speech production. This includes the muscles of the tongue, lips and jaw, but also muscles from the pharynx, larynx, and velum. Most muscles of speech production are volitionally controlled, which adds variability to their timecourse. The large number of source muscles and variability associated with volitional control makes EMG artifact much more difficult to isolate than EOG and EKG artifact using methods like ICA (McMenamin et al. 2010; Shackman et al. 2009; Chen et al. 2019). In addition, the frequency range of EMG responses is very wide: some researchers categorize EMG within the alpha (8-13Hz) and beta (13-20Hz) bands (Friedman & Thayer 1991), while others report broader frequency distributions in the 1-200Hz range (Goncharova et al. 2003). ICA is very successful at isolating neural components with similar topographic distribution, voltage change and timecourse from trial to trial, similarities which EMG artifact often does not have trial-to-trial.

At the single-word level, there has been success at isolating and correcting EMG activity through a blind source separation technique similar to ICA called Canonical Correlation Analysis (CCA) (Vos et al. 2010; De Clercq et al. 2006). The primary difference between ICA and CCA is that CCA excels in identifying data that have low autocorrelation within a specified sliding

window. EMG activity is weakly autocorrelated compared to EOG, EKG and EEG signal due to how variable its timecourse can be. CCA has been successfully used to remove EMG in a synthetic data set (De Clercq et al. 2006) and at the single-word level in a speech production task (Vos et al. 2010).

1.3.3 Multivariate Temporal Receptive Field Modeling

Another advantage to the high temporal resolution of EEG is that a large number of samples of neural activity can be acquired, which allows for computationally intensive modeling techniques. The multivariate temporal receptive field (mTRF) is a method for modeling activity of neural populations (in this case, electrodes) in response to certain stimulus features and is often utilized with EEG and ECoG datasets (Martin et al. 2018; Hullett et al. 2016; Crosse et al. 2016; Di Liberto et al. 2015; Hamilton et al. 2018). These models use linear regression to predict the EEG time series from a combination of acoustic, linguistic, or behavioral features, which allows researchers to test hypotheses about which specific acoustic, linguistic, or behavioral features of speech drive neural responses and how these responses can be modulated by context. Phonological features such as place and manner of articulation are common stimulus features used in mTRF modeling, as they are theorized to be neurally represented in the STG (Mesgarani et al. 2014; Hamilton et al. 2018). mTRF models, while commonly used with speech perception data, are less commonly used with speech production data. Using mTRF modeling on speech production data could provide insights into whether speaker-induced

suppression (see §1.1.1) reflects a general reduction in response intensity or if specific features of speech perception are not encoded when perceiving self-generated speech.

1.4 Objective and Hypotheses

To summarize, the objective of this thesis is to study speech production in a naturalistic context using EEG. To feasibly accomplish this, novel artifact correction techniques must be implemented. I am interested in how neural responses to speech production and perception differ and which behavioral stimulus characteristics can modulate the degree of response suppression during speech perception, one possible modulator being stimulus predictability.

The hypotheses of the study are presented below:

- **Hypothesis 1:** To develop methods for removing EMG from EEG data recorded during naturalistic speech production, canonical correlation analysis (CCA) will be utilized. I hypothesize that CCA will be able to successfully remove EMG artifact from EEG data while preserving the integrity of the recorded neural responses.

- **Hypothesis 2:** Using naturalistic speech production and perception stimuli, I hypothesize that a relative suppression in amplitude of the production responses compared to the perception responses will be observed in the N100/P200 components of the EEG.

- **Hypothesis 3:** When comparing responses to predictable and unpredictable naturalistic speech perception stimuli, I hypothesize that the responses to predictable stimuli will be reduced compared to the unpredictable stimuli in the N100/P200 components of the EEG.

Chapter 2

Methods

2.1 Participants

21 EEG participants were recruited from flyers placed around the University of Texas at Austin campus (11 female, 10 male; age range 18-35; age mean 24.4±3.9). One participant (OP0020) was excluded due to a recording error, and another participant (OP0004) was excluded due to delays in transcription (see §2.5). All participants were native speakers of English. Pure tone audiometry and a speech-in-noise hearing test were used to ensure all participants had typical hearing. Pure tone audiometry followed standard clinical guidelines (Working Group on Manual Pure-Tone Threshold Audiometry 2005). Hearing responses to the pure tone audiogram consisted of bilateral hearing thresholds of <25dB in the range of 125 to 8000Hz. The QuickSIN test (Killion et al. 2004) was used to assess hearing in noise after confirmation of typical hearing via audiogram. A range of 0-3 dB SNR loss was observed during QuickSIN testing, which was within normal limits. Participants provided written consent for participation in the study and were compensated at a rate of $15/hour for their participation. Sessions lasted an average of one hour. All experimental procedures were approved by the Institutional Review Board at the University of Texas at Austin. Table 2.1 summarizes

demographic information of the participants in this study.

ID	Gender	Age	Languages Spoken	EMG Placement
OP0001	M	24	English	N/A
OP0002	F	25	English, Gujarati	N/A
OP0003	F	21	English, Spanish	orbicularis oris & mandible
OP0004	F	23	English	orbicularis oris & mandible
OP0005	F	18	English, Mandarin	orbicularis oris & mandible
OP0006	F	22	English	orbicularis oris & mandible
OP0007	F	27	English, Spanish	orbicularis oris & mandible
OP0008	F	23	English	orbicularis oris & mandible
OP0009	F	24	English, Spanish, Polish	orbicularis oris & mandible
OP0010	M	30	English	orbicularis oris & mandible
OP0011	F	21	English	submentalis
OP0012	M	26	English	masseter
OP0013	M	35	English	orbicularis oris & mandible
OP0014	M	23	English	masseter
OP0015	F	23	English	mylohyoid
OP0016	M	25	English	orbicularis oris & mandible
OP0017	M	30	English	masseter
OP0018	M	28	English	masseter
OP0019	M	23	English	masseter
OP0020	M	21	English	orbicularis oris & mandible
OP0021	F	20	English	masseter

Table 2.1: Participant information.

2.2 EEG Data Acquisition

Neural responses were recorded continuously using a 64-channel scalp EEG cap connected to a BrainVision actiChamp amplifier (Brain Products, Gilching, Germany). Data were acquired at a sampling rate of 25kHz, and impedance level was kept below 15kΩ throughout recording. Conductive gel

was applied between the scalp and electrodes using a flat-tipped syringe to reduce impedance. Pycorder, software developed by Brain Products, was used to control and record responses from the amplifier. For 2 participants (OP0015, OP0016), an amplifier battery change necessitated pausing of the task. Consequently, for these subjects, recording was split into two different data files. Trials at the end of the session where the battery needed to be changed were excluded from analysis to prevent edge artifact.

Audio levels were tested prior to the start of task and were presented to the participant via 3M E-A-Rtone Gold 10Ω insert earbuds (3M, Minnesota, USA) at a comfortable volume, and the participant's responses were recorded via a wall-mounted Audio Technica U853rw cardioid condenser microphone (Audio Technica, Tokyo, Japan). Insert earbuds were not noise-cancelling, so participants could hear their live auditory feedback during sentence production[1]. Soundproof paneling in the recording booth and throughout the recording suite minimized background noise during data collection. Auditory stimuli were recorded and synchronized using a StimTrak stimulus processor (Brain Products). Visual stimuli were presented on an Apple iPad Air 2 (Apple, California, USA). The stimuli were controlled by the participant during data collection through custom interactive software developed in Swift (Apple). Stimulus changes were locked to the refresh rate of the screen at 60Hz to

[1]Earbuds were foam-tipped which likely reduced the amplitude of auditory feedback; however, auditory feedback is still available to the participant through a mixture of attenuated air conduction and bone conduction, and lack of access to auditory feedback is not a concern for this study.

minimize jitter and obtain high temporal precision data, as is commonly done in psychophysics software such as the MATLAB PsychToolbox.

To isolate vEOG and EMG activity, four auxiliary electrodes were used. vEOG electrodes were placed above and below the participant's left eye. An abrasive conductive gel was used to reduce impedance at facial electrodes (Abralyt gel, Brain Products). Because EMG activity can originate from many muscles, different placements were trialed across participants. Placement efficacy was assessed by the number of EMG activity-related events that were detected using the MNE function `mne.preprocessing.create_eog_epochs()`, which uses a peak detection process to find artifacts on the specified channel (Figure 3.1). One participant (OP0008) consented to additional time during pre-recording setup to test multiple EMG electrode placements; for this subject, placement on the origin and insertion of the masseter muscle resulted in the largest number of identified EMG events. Across all subjects, EMG activity recorded from the masseter muscle, with one electrode on the origin (zygomatic arch) and one on the insertion (angle of mandible) resulted in the largest number of EMG events. However, a placement on the mental protuberance of the mandible and the superior orbicularis oris also resulted in a large number of identified EMG events and was used on participants for whom electrode adhesion to the angle of the mandible was unreliable. A summary of facial electrode placement and responses can be found in Figure 2.1[2].

[2]Images in Figure 2.1, panels A and B, used under the Creative Commons Attribution-Share Alike 3.0 Unported license: `https://creativecommons.org/licenses/by-sa/3.0/`.

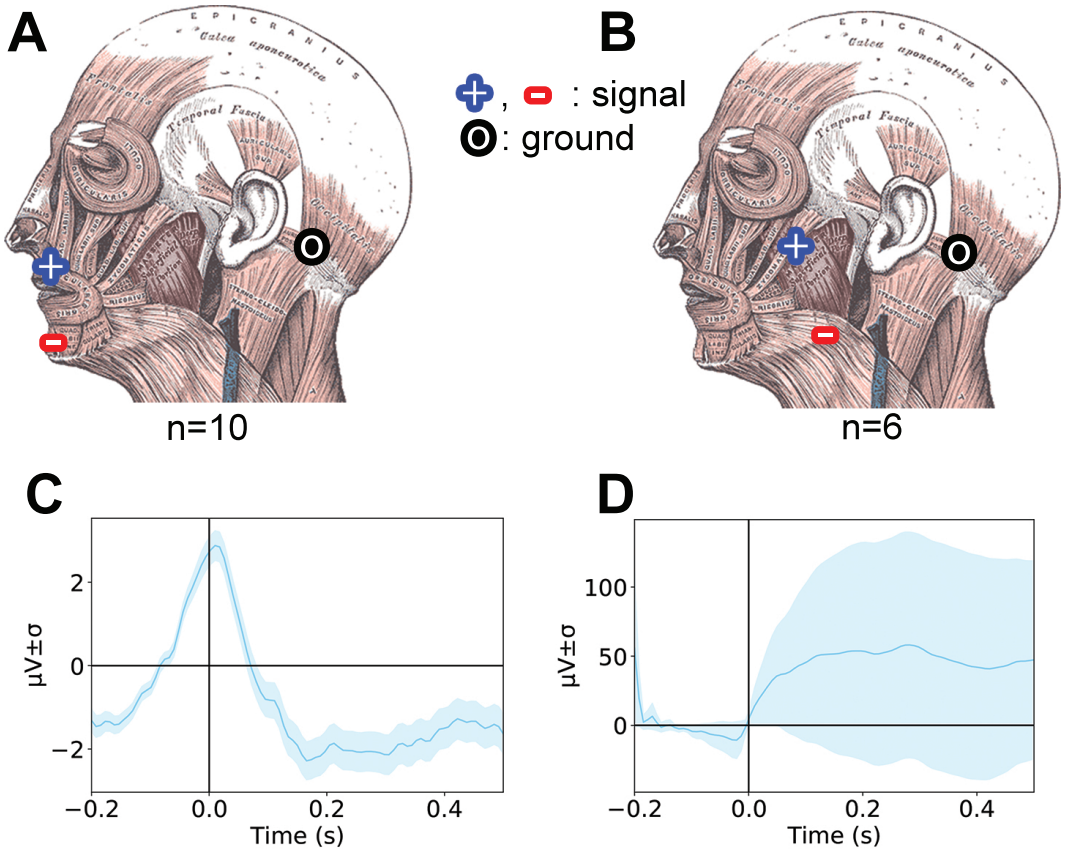

Figure 2.1: EMG placement. A: Electrode schematic for orbicularis oris & mentalis placement. B: Electrode schematic for masseter placement. Both placements used a reference electrode on the zygomatic process. C,D: EEG response in the raw data to detected peak activity from the electrode placements in A and B, respectively. More details about how these epochs were generated can be found in §3.1.2.

Additional electrode placements, such as on the submental triangle and mylohyoid, were trialed in individual subjects. Two subjects (OP0001, OP0002) did not have auxiliary electrodes for identification of EMG activity. The purpose of these auxiliary EMG electrodes was to detect EMG activity

associated with the onset of articulation, as speech-onset articulation would cause the largest artifact in the temporal window of interest for the analyses described in §2.8. The masseter, orbicularis oris, mylohyoid and submentalis have all been implicated in jaw movement by previous facial EMG studies (Stepp 2012; Van Eijden et al. 1993; Rastatter & De Jarnette 1984).

2.3 Experimental Design

The task was designed using a dual perception-production block paradigm, where trials consisted of a dyad of production component followed by perception component. Trials began with participants reading a sentence displayed on the stimulus presentation iPad. The following perception component was split into two experimental conditions: predictable and unpredictable. During predictable perception trials, audio from the production component of the dyad was played back, making auditory stimuli between perception and production identical[3] (Figure 2.2[4], panel B). During unpredictable perception trials, audio from a previous predictable dyad was played back, causing a stimulus mismatch between the perception and production components of the current dyad. Dyads switched between predictable and unpredictable perception every

[3]In Figure 2.2 panel B, the amplitude between production and perception trials is visibly different. The spectral characteristics of the stimuli are identical (except amplitude). Production stimuli were recorded into the amplifier through the wall-mounted microphone while perception stimuli were recorded through the built-in iPad microphone. Earbud volume was calibrated for each participant to ensure a comfortable listening level, so amplitude between perception and production as perceived by the participant are assumed to be similar.

[4]Icons in Figure 2.2 panel A used with permission from Flaticon: `https://freepikcompany.com/legal`.

50 trials. Sessions lasted between 300 and 400 trials to ensure a large number of trials for each condition (Luck 2014). Sentences used in the task were taken from the MultiCHannel Articulatory (MOCHA) database, a corpus of sentences designed to include a wide distribution of phonemes and phonological processes typically found in spoken English (Wrench 1999). A subset of 50 sentences of the total 460 sentences in MOCHA was randomly chosen for use in the task, with the exception of one subject (OP0001) for whom 100 sentences were used. Before random selection, 61 sentences were manually removed by the author for containing offensive semantic content (e.g., "Women may never become completely equal to men.") or being difficult for an average reader to produce (e.g., "Many wealthy tycoons splurged and bought both a yacht and a schooner.") to reduce extraneous cognitive effects of the task and error production, respectively.

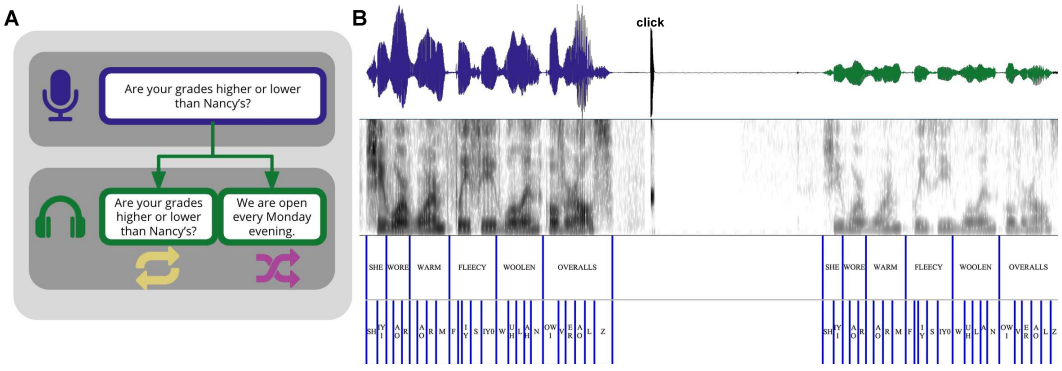

Figure 2.2: Task overview. A: schematic of experimental design. Production trials (blue) are used to create stimuli for perception trials (green), which are then divided into predictable (yellow) and unpredictable (magenta) blocks. B: Waveform, spectrogram and Praat TextGrid comparing perception and production trials with identical phonetic information.

Stimulus presentation was controlled by the participant on an iPad running custom software written in Swift (version 4, Apple). During production trials, stimuli were presented in a white font on a black background after a 1000ms fixation cross to minimize visual artifact in the EEG signal. The stimulus iPad was placed on an overbed table to minimize arm travel needed to interact with the screen and ensure that stimuli were presented at a comfortable reading distance. Participants were given as much time as they needed to read the sentence, after which they pressed a "Next" button to begin the perception component of the trial. Following a 1000ms fixation cross, perception stimuli were played to the participant accompanied by a black screen on the iPad to again minimize visual artifact. During the task, the Swift software collected timing information on button presses and stimulus presentation that was used to generate a log file to assist in data preprocessing. Log files contained: (1) within-block trial number, (2) cumulative trial number, (3) whether the trial was perception or production, (4) whether the trial was predictable or unpredictable, (5) a timestamp of when the stimulus was presented, (6) transcriptions of the perceived/produced sentence[5], and (7) how many times the presented sentence had been repeated during the data collection session.

[5]Transcriptions of produced sentences were taken directly from the MOCHA corpus and assume no errors were made during production. Additional steps were taken to make sure any errors produced during the task were included in the transcription (see §2.5).

2.4 EEG Signal Preprocessing

All preprocessing was performed offline. EEG, EOG and EMG data were downsampled from 25kHz to 128Hz using BrainVision Analyzer (Brain Products). Data were recorded at 25kHz because the audio signals from the microphone and iPad were recorded on the same amplifier as EEG for easy synchronization of stimulus audio with the EEG, and a high sampling rate for audio stimuli was necessary in order to use a match filter for transcription purposes (see §2.5). Subsequent preprocessing steps were performed using custom Python scripts and functions from the MNE-python software package (Gramfort et al. 2014). Data were then concatenated across multiple blocks for the 2 participants who had multiple blocks of data collected. A linked mastoid (electrodes TP9, TP10) reference was applied followed by a 60Hz notch filter to remove electrical line noise artifact. For one subject (OP0017), one of the reference electrodes (TP9) was a bad channel and was interpolated prior to referencing. A 1-30Hz bandpass filter was applied to facilitate artifact rejection and independent component selection; however, artifact correction was performed on non-bandpassed data, with a 1-15Hz bandpass filter being applied to artifact-corrected data before analysis. All filters applied to the EEG were designed using the finite impulse response method (Saramaeki et al. 1993).

Next, power spectral density by channel was visually inspected to isolate bad channels. Bad channels and segments were then manually annotated for rejection. EKG, EOG, and EMG artifacts were corrected using procedures

detailed below.

2.4.1 Artifact Correction through Independent Component Analysis and Canonical Correlation Analysis

Independent component analysis (ICA) was performed using MNE to remove EKG and EOG artifact from the data. ICA is a blind source separation technique widely utilized by EEG researchers to identify and correct for artifact in the data (Jiang et al. 2019; Barlow & Dubinsky 1980; Gao et al. 2009; Dimigen 2020). The number of ICA components selected was set to the number of non-bad channels. ICA was performed using an information-maximization approach (Bell & Sejnowski 1995), and manually rejected channels and segments were ignored during the fitting of ICA. EOG epochs were generated using activity from the vEOG auxiliary electrodes and used to detect ICA components related to EOG activity. Components related to EOG and EKG activity were visually inspected and rejected.

While ICA was fit on 1-30Hz bandpass filtered data, it was applied to unfiltered data because canonical correlation analysis (CCA) works best on minimally filtered data due to the wide range of frequencies in which EMG artifact can occur (Goncharova et al. 2003). The post-ICA data were saved as a `.fif` file that was converted to a `.vhdr` file using a custom version of the Philistine Python package (Alday 2018) to make the ICA-corrected data compatible with the MATLAB script used to run CCA. After conversion, the data were highpass filtered at 0.16Hz to remove low frequency EMG before

CCA (Vos et al. 2010). CCA was completed in the MATLAB (MATLAB 2017) script EEGLab (Delorme & Makeig 2004) using the AAR plugin (Gomez-Herrero et al. 2006). EEGLab was used to delete any channels that were annotated as bad in MNE before the AAR plugin was used. Bad channels were interpolated prior to analysis. The AAR plugin estimated power spectral density of EEG and EMG components using a Welch spectrum estimator with a Hamming window. The approximate frequency separating the EEG and EMG was set at the default 15Hz. The options passed to the BSS algorithm included `'eigratio'=1e6` and the default criterion options for optimization: `('ratio'=10, 'range'=[0, 32])`. CCA was run in two passes: first, a 30-second window to remove tonic muscle activity; second, a 2-second window to remove rapid bursts of EMG associated with speech production. For these analyses, both the window length and window shift were set to 30 or 2 seconds, respectively.

2.5 Transcription

Accurate timing information for words, phonemes and sentences was generated to allow epoching of EEG data to multiple levels of linguistic representation (see §2.2, panel B). To expedite this process, a modified version of the Penn Phonetics Forced Aligner (Yuan & Liberman 2008) (P2FA) was used to automatically generate Praat (Boersma 2001) TextGrids (Figure 2.2, panel B). As an input, P2FA used a transcription of the task generated by the iPad log file that was then manually edited to check for errors by dyads of undergradu-

ate research volunteers. P2FA-generated TextGrids were checked for errors by undergraduate research volunteers in a similar fashion. The author supervised the transcription process and checked the final TextGrids for accuracy before generating the event files used in the analyses. Because auditory stimuli for the perception and production components of the task were identical, the production TextGrids with confirmed accurate timing information were used in conjunction with a match filter, which aligned recorded production audio to recorded perception audio using convolution. The match filter was used to find the onset and offset of each sentence and automatically generate TextGrids for the perception component of the task. Perception component TextGrids were manually inspected by the author for accuracy before generation of event files.

2.6 Event-Related Potential Analysis

Event files were automatically generated using the iPad log files and the semi-automatically generated TextGrids. Event files contained start and stop times for each phoneme, word and sentence in each recording session, as well as information about each: perception versus production, predictable versus unpredictable, phoneme/word/sentence ID, phoneme/word/sentence transcription, and current repetition within block. Additional event files with start and stop times for each inter-trial click sound were generated using a match filter.

All epochs used for event-related potential (ERP) analysis in this study were generated from these event files. Neural data was bandpass filtered 1-

15Hz before ERP analysis.

2.7 Multivariate Temporal Receptive Field Analysis

The multivariate temporal receptive field (mTRF) approach was used to describe the selectivity of a neural response to given stimuli (Crosse et al. 2016; Di Liberto et al. 2015; Hamilton et al. 2018). Forward modeling mTRFs attempt to describe the statistical relationship between the input (given stimuli) and the output (the predicted EEG response). Crosse et al. (2016) conceptualized TRFs as "a filter that describes the linear transformation of an ongoing stimulus to an ongoing neural response."

$$\mathrm{EEG}(t,n) = \sum_{f}\sum_{t} w(f,\tau,n)s(f,t-\tau) + \epsilon(t,n) \tag{2.1}$$

Figure 2.3: Equation for the forward model temporal receptive field. This model demonstrates the neural response EEG at time t from electrode n as a convolution between two matrices: the input stimulus property $s(f, t-\tau)$ with the EEG TRF $w(f,\tau,n)$. $\epsilon(t,n)$ represents the residual response not explained by the model. Adapted from Crosse et al. (2016).

Forward modeling TRF linear regression was used with different sets of linguistic and behavioral features. Phonological features based on place and manner of articulation were adapted from Hayes (2011) and used as features in the TRF model (Table 2.2). Behavioral information about the current trial (perception, production, predictable, unpredictable), as well as normalized EMG activity recorded from facial electrodes, was also included in the model.

EMG was normalized by dividing the amplitude at each timepoint by the maximum amplitude of the auxiliary electrode throughout the task.

	Dorsal	Coronal	Labial	Plosive	Fricative	Nasal	Voiced	Obstruent	Sonorant		High	Front	Low	Back	Syllabic	Voiced	Sonorant
p	-	-	+	+	-	-	-	+	-	ɑ	-	-	+	+	+	+	+
b	-	-	+	+	-	-	+	+	-	æ	-	+	+	-	+	+	+
t	-	+	-	+	-	-	-	+	-	ɔ	-	-	+	+	+	+	+
d	-	+	-	+	-	-	+	+	-	ʌ	-	-	+	+	+	+	+
k	+	-	-	+	-	-	-	+	-	ə	-	-	+	+	+	+	+
g	+	-	-	+	-	-	+	+	-	ɚ	-	-	+	+	+	+	+
m	-	-	+	-	-	+	+	-	+	i	+	+	-	-	+	+	+
n	-	+	-	-	-	+	+	-	+	ɪ	+	+	-	-	+	+	+
ŋ	+	-	-	-	-	+	+	-	+	ɨ	+	+	-	-	+	+	+
ʔ	-	-	-	+	-	-	+	+	-	ɛ	+	+	-	-	+	+	+
f	-	-	+	-	+	-	-	+	-	u	+	-	-	+	+	+	+
v	-	-	+	-	+	-	+	+	-	ʊ	+	-	-	+	+	+	+
θ	-	+	-	-	+	-	-	+	-	ʉ	+	-	-	+	+	+	+
ð	-	+	-	-	+	-	+	+	-	aʊ	-	-	+	+	+	+	+
s	-	+	-	-	+	-	-	+	-	oʊ	-	-	+	+	+	+	+
z	-	+	-	-	+	-	+	+	-	eɪ	+	+	-	-	+	+	+
ʃ	-	+	-	-	+	-	-	+	-	aɪ	+	+	-	-	+	+	+
ʒ	-	+	-	-	+	-	+	+	-	oɪ	+	-	-	+	+	+	+
h	-	-	-	-	+	-	-	+	-								
r	-	+	-	-	-	-	+	-	+								
j	+	-	-	-	-	-	+	-	+								
l	-	+	-	-	-	-	+	-	+								
w	+	-	+	-	-	-	+	-	+								
tʃ	-	+	-	-	+	-	-	+	-								
dʒ	-	+	-	-	+	-	+	+	-								

Table 2.2: Consonant and vowel feature matrices for stimuli used in TRF modeling.

All 64 channels were used during mTRF modeling, with bad channels

interpolated prior to modeling. Each channel was fit with a separate model to predict the EEG signal in that channel. For all linguistic and behavioral feature models, mTRFs were fit using a time delay of −300ms to 500ms. This delay range encompasses the temporal integration times to similar responses found in previous research (Hamilton et al. 2018), with an added negative delay to encompass potential prearticulatory neural activity and activity reflecting neural control of the motor response. The data were split into a training set and a validation set, with 80% of sentences in the task being used for training and 20% being used for validation. Training and validation sets were split by unique sentence to avoid potential overfitting of the TRF by including the same sentence in both sets. The weights (w) were fit using ridge regression on the training set and a regularization parameter chosen by a bootstrap procedure (n=10 bootstraps). The performance of the model was then calculated on the held out test set, and the ridge parameter was selected at the value that provided the highest average correlation performance across all bootstraps. The ridge parameter was the same across all electrodes. Ridge parameters between $10e$-4 and $10e4$ were tested in 30 logarithmically-scaled intervals.

2.8 Statistical Analysis

An advantage of naturalistic datasets is their size: it becomes possible to perform many computational analyses. The analyses used in this thesis are presented in this section.

2.8.1 Event-Related Potential Evaluation

To determine the statistical significance of the event-related potential analysis, a linear mixed-effects (LME) model was created and assessed using the `lmertest` package (Kuznetsova et al. 2017) in R (Computing & Others 2013). LME models are well-suited to analysis of large EEG datasets due to minimal assumptions made about the structure of the data and the ability to examine behavioral effects across subjects that accounts the high degree of within-subject variation that exists in EEG datasets. Assessment of differences in perception and production and differences in predictable and unpredictable perception required two separate models to be fit, as the predictable/unpredictable split occurring in only half of the total trials (i.e., the perception trials) would cause a single model to be unbalanced. For both models, the behavioral distinction (perception & production; predictable & unpredictable) of interest was used as a fixed effect while the individual subject was used as a random effect (Figure 2.4, Equations 2.2 and 2.3).

$$x \sim \text{Condition} + (1|\text{Subject}) \tag{2.2}$$

$$x \sim \text{Predictability} + (1|\text{Subject}) \tag{2.3}$$

$$x \sim \text{Epoch type} + (1|\text{Subject}) \tag{2.4}$$

Figure 2.4: Equations for linear mixed-effects models, where x is the response variable. For Equations 2.2 and 2.3, response variables were peak to peak amplitude of N1-P2 complex, N100 response amplitude/latency, and P200 response amplitude/latency. For Equation 2.4, the response variable was the mean difference wave between the raw and CCA-corrected EEG response.

Response variables focused on the N100 and P200 components (see §1.3.1). The peak amplitude of the N100 component was used as a response variable and obtained as the peak amplitude in the 80-150ms time window after sentence onset (Lijffijt et al. 2009). The peak amplitude for the P200 was used as a response variable in a similar fashion but was obtained from the 150-250ms range. The response latency of both these components was also included as response variables, obtained as the time in milliseconds at which the peak amplitudes of these components occurred. The last response variable included in analysis was the peak-to-peak amplitude of the N100/P200 components, which was operationalized as the difference in peak amplitude between the two. Responses from nine EEG channels were included in the models due to their topographic relevance for the N100 component (see Figure 3.3 panel F and Figure 3.4 panel F): F1, Fz, F2, FC1, FCz, FC2, C1, Cz and C2. Full model parameterization and results for the perception/production LME and predictable/unpredictable LME are summarized in Tables 3.2 and 3.3, respectively.

An LME model was also used to evaluate the effectiveness of CCA correction at removing EMG from the data and preserving response integrity. The root mean square (RMS) difference wave of the epoched responses averaged across the same nine channels included in the above LME models was used as the response variable. Difference waves were calcuated by subtracting the channel-averaged response of the CCA-corrected data from the channel-averaged response of the raw data at each epoch. Three sets of difference

waves were created in this fashion corresponding to three sets of epochs: (1) peak EMG activity from facial electrodes obtained via the MNE (Gramfort et al. 2014) function `mne.preprocessing.create_eog_epochs()` (§2.2, Figure 2.1), (2) sentence onsets (§2.6) and (3) inter-trial click tone responses (§2.3). All three sets of epochs were obtained from -200ms to +500ms relative to the event. The LME model had a fixed effect of epoch type and a random effect of subject (Figure 2.4, Equation 2.4). RMS values were used so response polarity would not influence model evaluation.

As an additional confirmation of CCA efficacy, RMS values were obtained from manually-annotated jaw EMG epochs in a single subject (OP0008). A Wilcoxon signed-rank test implemented in the SciPy (Jones et al. 2001) function `scipy.stats.wilcoxon()` was used to compare the mean of these epochs across channels between the raw and CCA-corrected datasets. The Wilcoxon signed-rank test is a nonparametric statistical test used to assess differences in matched pairs of data (Woolson 2007). Because the behavioral distinctions of interest come from the same dataset, it can be assumed that their means do not follow a Gaussian distribution, which makes these data well-suited for Wilcoxon signed-rank tests. p-values from this test were false discovery rate-corrected for multiple comparisons using the negative Benjamini/Yekutieli method (Benjamini & Yekutieli 2001).

2.8.2 mTRF Model Evaluation

Evaluation of mTRF model performance was obtained by observing the linear correlation coefficients (r) between the observed EEG response and the EEG response predicted by the model. Signifance of these correlations was obtained through within-subject[6] channel-by-channel reshuffling, model refitting and subsequent bootstrap tasks with 100 iterations each. The significance threshold for each correlation was set as $p < 0.01$ ($1/nboots$). Between-model significance of linear correlation coefficients was compared using the negative-Benjimani-Yekuteli-corrected Wilcoxon signed-rank test used to assess CCA artifact correction efficacy described above. This nonparametric test was also used to calculate the significance of differences of perception/production (Figure 3.7) and predictable/unpredictable (Figure 3.8) feature weights at individual timepoints.

[6]It is important that significance correlations are obtained on a within-subject basis due to the large individual variations in baseline that can occur in EEG datasets. This also explains why some channels in Figure 3.6 appear to have insigificant correlation values that are greater than significant correlation values.

Chapter 3

Results

3.1 Validation of Artifact Correction Techniques

Because studies of speech production using naturalistic stimuli above the word level are rare, it is important to confirm that EMG artifact correction techniques were successful. Because there is no guaranteed method of confirming an artifact correction technique is both successful and accurate, a custom method of confirming that both Type I (false positive) and Type II (false negative) error were absent from the dataset was developed.

3.1.1 Preservation of Neural Responses after CCA Artifact Correction

In the context of this dataset, a lack of Type I error means that neural signal is not falsely identified as EMG and removed from the dataset. To confirm signal integrity, well-studied neural components related to auditory processing were observed before and after CCA correction. The N1-P2 response is a low-level neural response to auditory stimuli that is automatic and not affected by cognition (Lightfoot 2016). If EEG was falsely removed during CCA artifact correction, the N1-P2 response would likely be degraded when CCA-corrected data are compared to pre-CCA-corrected data. Figure

3.1 (panels A,D) demonstrates that the N1-P2 complex is preserved after CCA correction in ERP data epoched to the inter-trial broadband click tone. The integrity of task-related ERP (Figures 3.3, 3.4) responses provides further corroboration that neural signal is not falsely removed from the dataset.

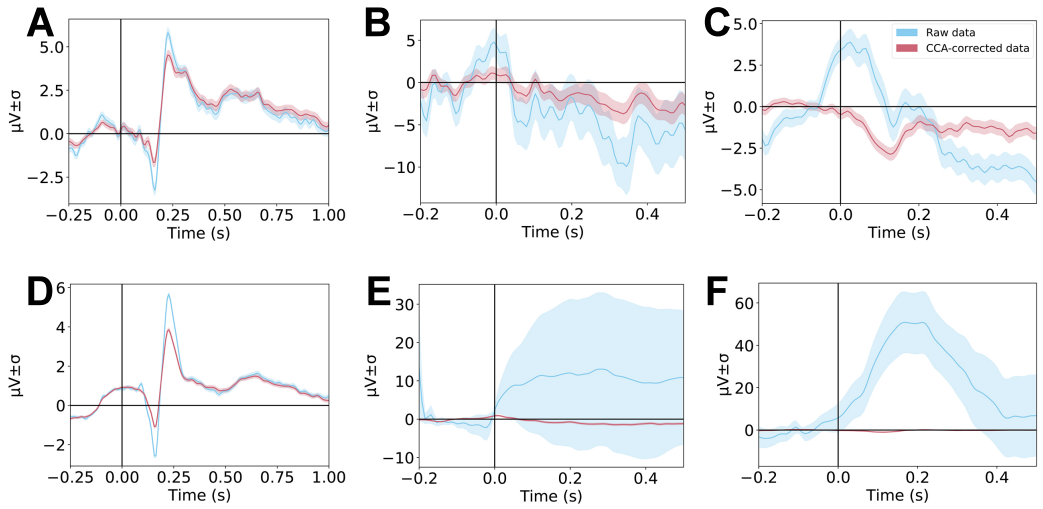

Figure 3.1: CCA correction removes EMG artifact without significantly affecting auditory responses, as shown by comparison of event-related potential activity between raw data (blue) and CCA-corrected data (red). The top row of panels (A,B,C) shows responses in a single subject (OP0008) while the bottom row of panels (D,E,F) shows grand average responses in 17 subjects. Left column (A,D): ERP responses epoched to the inter-trial click tone. Middle column (B,E): ERP responses epoched to EMG activity recorded from facial electrodes. Right column (C,F): ERP responses epoched to the onset of sentence articulation. All panels include data averaged across nine electrodes: F1, Fz, F2, FC1, FCz, FC2, C1, Cz, and C2.

3.1.2 Removal of EMG activity after CCA Artifact Correction

A lack of Type II error in this dataset means that EMG activity is accurately removed from the dataset. EMG activity associated with articula-

tion was identified automatically using activity from the facial electrodes (see Figure 2.1, §2.8). A comparison of EMG-epoched responses between the raw dataset and the CCA-corrected dataset showed a large reduction in amplitude corresponding with the onset of EMG activity, which suggests that CCA was successful in removing EMG activity from the dataset (Figure 3.1, panels B,E). A similar comparison using sentence-onset-epoched instead of EMG-epoched responses showed that there is a large reduction in amplitude corresponding with sentence articulation (Figure 3.1, panels C,F). Manual inspection of 109 jaw clench artifacts in an individual subject (OP0008) demonstrated that EMG activity not associated with peak activity from facial electrodes or sentence production is also removed from CCA-corrected data ($p < 0.0001$, Wilcoxon signed-rank test). Linear mixed effects models comparing sets of root mean square difference waves are summarized in Table 3.1 and Figure 3.2. LME modeling of difference waves revealed a significant contrast between the click responses and EMG/sentence responses but no significant contrast between EMG and sentence responses, suggesting that EMG artifact correction is being applied to peak EMG activity and sentence-level epochs but not to click responses. This result is consistent with the simultaneous removal of EMG artifact and preservation of neural responses. Overall there was an average of close to zero difference in amplitude of the click response before and after artifact correction. On the other hand, EMG epochs and sentence epochs showed reductions in amplitude following CCA, likely related to removal of artifact (Figure 3.2).

Epoch Type	Estimated marginal mean	95% Confidence interval
Click	-1.43µV	-53.57 to 50.71µV
EMG	51.50µV	-1.63 to 104.62µV
Sentence	40.27µV	-11.21 to 91.74µV
Contrast	Estimated marginal mean	p value
Click−EMG	-52.93±11.47µV	< 0.0001
Click−Sentence	-41.70±8.73µV	< 0.0001
EMG−Sentence	11.23±10.31µV	0.52

Table 3.1: Linear mixed-effects model results comparing three sets of mean difference waves between raw and CCA-corrected data.

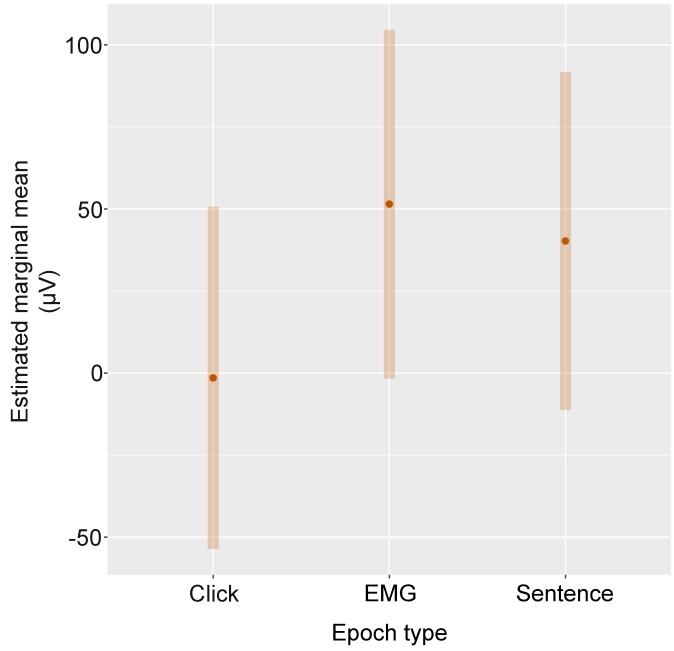

Figure 3.2: Plot of estimated marginal means of difference wave amplitude split by epoch type. The shaded area represents the confidence interval for each epoch type's estimated marginal mean calcuated via Kenward-Roger approximation.

3.2 Event-Related Potential Results

To examine differences between speech production and perception and differences between unpredictable and predictable speech perception, two methods were used. The first, event-related potential (ERP, §2.6) analysis, was used to compare sentence-level responses between behavioral conditions.

3.2.1 Differences Between Speech Production and Perception

Previous research has demonstrated a suppression effect in speech production relative to perception at the word and syllable level (Okada et al. 2018; Houde & Nagarajan 2011; Toyomura et al. 2020; Behroozmand & Larson 2011), as well as a general suppression of self-generated compared to externally-generated sounds (Martikainen et al. 2005; Brumberg & Pitt 2019). To examine if responses to speech production are suppressed in naturalistic contexts, differences between the speech production and speech perception components of the task were compared using ERP analysis.

Comparison of perception and production trials is summarized in Figure 3.3. Data were epoched to sentence onset for analysis: for production trials, this corresponds to the onset of articulation; for perception trials, this corresponds to the onset of the first presented phoneme in the trial sentence. Topographic EEG activity in response to sentence onset revealed a frontal/central ROI of activity in both the perception condition at 100ms and the production condition at -100ms.

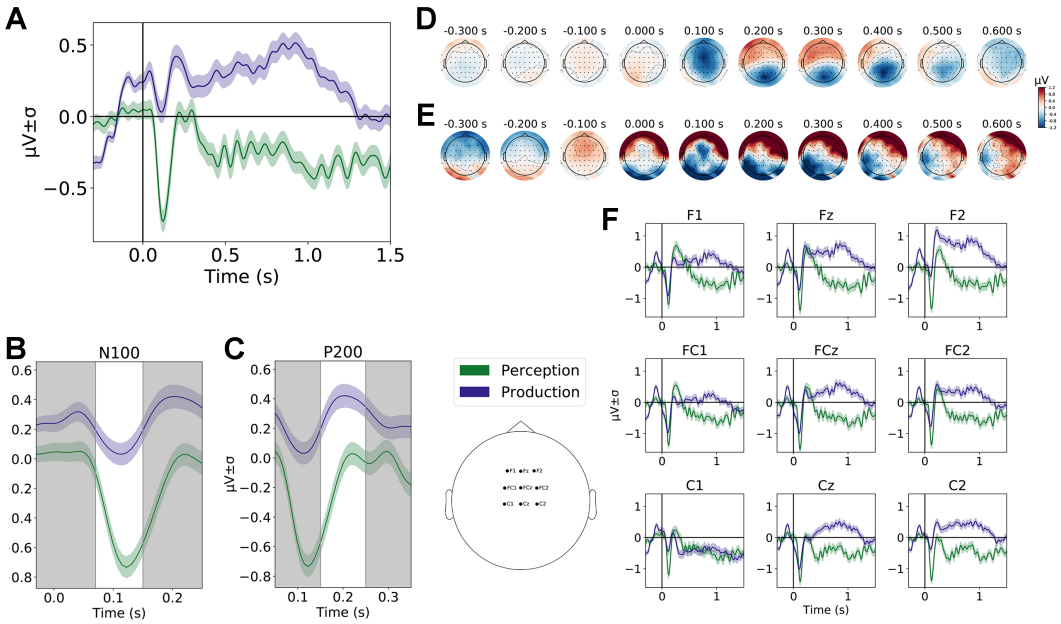

Figure 3.3: Sentence-level ERP activity demonstrates relative suppression of production (blue) compared to perception (green) trials. Panel A: Grand average ERP plot of activity epoched to sentence onset averaged across 64 channels, n=19 subjects. B,C: Time ranges for N100(B) and P200(C) components in-between shaded gray areas. D,E: Topographic plots of perception (D) and production (E) activity. F: Locations of nine frontal/central region of interest (ROI) electrodes on an EEG montage. G: Grand average plots comparing perception and production ERP activity in 19 subjects split by frontal/central ROI channels.

The N1-P2 complex (see §3.1.1) is present at the sentence level in both production and perception conditions, but relatively reduced in amplitude for the production trials. Linear mixed-effects model results, summarized in Table 3.2, suggest a significant difference in the amplitudes and latencies of both the N100 and the P200 components between perception and production conditions, as well as the peak-to-peak amplitude between the N100 and P200 components.

Amplitudes were reduced during production and response latencies decreased during production.

Response Variable	Fixed Effects	Random Effects	Perception estimated marginal mean	Production estimated marginal mean	95% Perception confidence interval	95% Production confidence interval	t ratio	p value
Peak to peak amplitude (µV)	Condition	Subject	12.1	8.1	9.8 to 14.3	5.9 to 10.3	25.6	< 0.001
N100 amplitude (µV)	Condition	Subject	-6.0	-3.7	-7.0 to -5.0	-4.7 to -2.7	-15.6	< 0.001
N100 latency (s)	Condition	Subject	0.116	0.114	0.115 to 0.117	0.113 to 0.115	3.5	< 0.001
P200 amplitude (µV)	Condition	Subject	6.0	4.3	4.6 to 7.3	2.9 to 5.7	11.2	< 0.001
P200 latency (s)	Condition	Subject	0.205	0.202	0.204 to 0.207	0.201 to 0.204	4.2	< 0.001

Table 3.2: Linear mixed-effects model results comparing speech perception and production.

Additionally, there was an increase in activity prior to stimulus onset ($\sim$ -100ms) for the production condition relative to the perception condition. This increase in pre-articulatory activity could be a component of feedforward speech motor control or motor speech programming (see §4.2).

3.2.2 Differences Between Predictable and Unpredictable Speech Perception

Differences between predictable and unpredictable speech perception were compared in a similar fashion to the differences between perception and production (see §3.2.1). Overall, the differences between these two trial types were less pronounced than the differences between perception and production trials. See §4.3 for an interpretation of this difference.

Comparison of predictable and unpredictable perception trials is summarized in Figure 3.4. Data were epoched to the onset of the first perceived phoneme in the trial sentence. The same frontal/central ROI shown in Figure 3.3 was present in the plots, although the magnitude of response was larger for unpredictable perception.

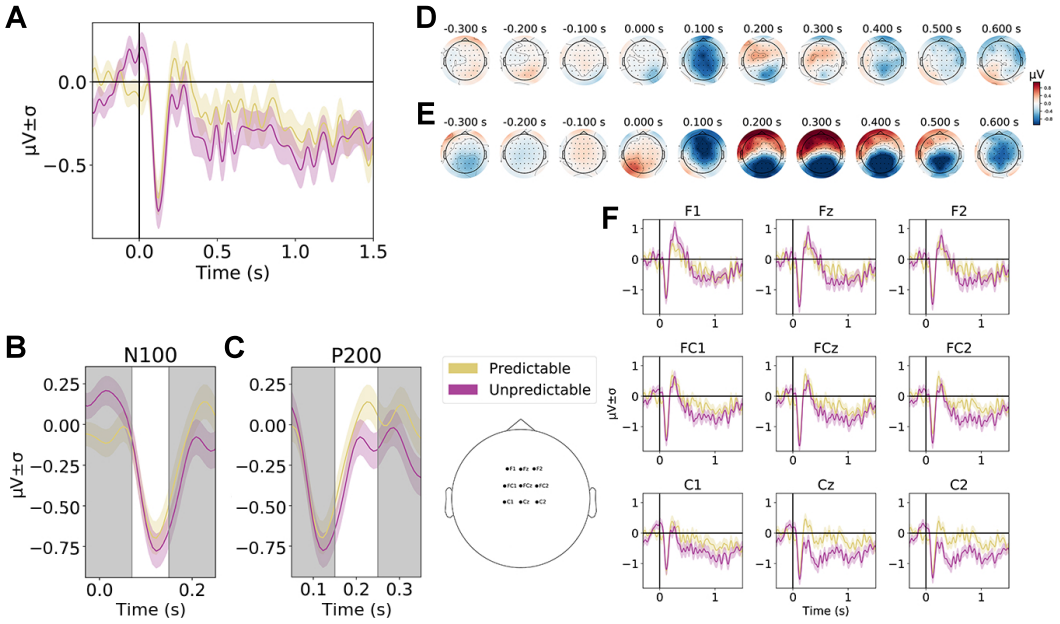

Figure 3.4: ERP comparison between predictable (yellow) and unpredictable (magenta) speech perception trials. Panel A: Grand average ERP plot of activity epoched to sentence onset in 19 subjects. B,C: Time ranges for N100(B) and P200(C) components in-between shaded gray areas. D,E: Topographic plots of predictable (D) and unpredictable (E) perception activity. F: Locations of nine frontal/central ROI electrodes on an EEG montage. G: Grand average plots comparing predictable and unpredictable ERP activity in 19 subjects split by frontal/central ROI channels.

Linear mixed-effect model results, summarized in Table 3.3, suggest

there was no significant difference in N100 and P200 amplitudes between predictable and unpredictable perception trials; however, peak-to-peak amplitude and N100 latency differed significantly between predictable and unpredictable trials. To further investigate this result, a series of Wilcoxon signed-rank tests with negative Benjimani-Yekuteli false discovery rate-correction comparing N100-P200 peak-to-peak amplitude was performed on a within-subject basis. Three individual subjects demonstrated a suppression of peak-to-peak amplitude at a $p = 0.05$ significance level; however, there was a high degree of inter-subject variation in these responses, plots of which have been included in Appendix 1.

Response Variable	Fixed Effects	Random Effects	Predictable estimated marginal mean	Unpredictable estimated marginal mean	95% Predictable confidence interval	95% Unpredictable confidence interval	t ratio	p value
Peak to peak amplitude (μV)	Predictability	Subject	11.7	12.2	8.5 to 15.0	9.0 to 15.5	-2.1	0.03
N100 amplitude (μV)	Predictability	Subject	-5.8	-6.1	-7.5 to -4.1	-7.8 to -4.5	1.5	0.12
N100 latency (s)	Predictability	Subject	0.117	0.115	0.116 to 0.118	0.114 to 0.117	2.3	0.02
P200 amplitude (μV)	Predictability	Subject	5.9	6.1	4.2 to 7.6	4.3 to 7.8	-0.9	0.37
P200 latency (s)	Predictability	Subject	0.205	0.205	0.203 to 0.207	0.203 to 0.207	0.21	0.84

Table 3.3: Linear mixed-effects model results comparing predictable and unpredictable speech perception trials.

3.3 Multivariate Temporal Receptive Field Model Performance

To understand how phonological information is modulated by the behavioral context of the task (i.e., in production versus perception, predictable versus unpredictable contexts) while controlling for the possible presence of residual EMG, I fit a series of multivariate temporal receptive field models. Multivariate temporal receptive field (mTRF) modeling serves to provide a model of how specific features were encoded in the EEG signal. Model performance was evaluated using the linear correlation coefficient (r) between the EEG activity predicted by the model and the actual EEG response. To evaluate how the inclusion/exclusion of specific features affected model performance, five models were run and evaluated using identical methods (see §2.7). Figure 3.5 provides a summary of comparative model performance. The "full" model (Model 1) contained 14 phonological features, 4 binary features encoding trial information (perception, production, predictable, unpredictable) and normalized EMG activity from facial electrodes for a total of 19 features. If Model 1 did not perform significantly different from the other models, the interpretation is that speech features were encoded similarly between perception and production. Model 2 is identical to Model 1 but has two additional sets of phonological features split by perception and production trials for a total of 47 (19 + 14 + 14) features. If Model 2 outperformed other models, the interpretation is that phonological features were encoded differently between perception and production. Model 3 and Model 4 were identical to Model 1

but with two binary features removed: predictable/unpredictable and perception/production, respectively, for a total of 17 $(19-2)$ features. If Model 3 and 4 performed as well as other models, the interpretation is that the modality (perception/production) and predictability of speech were not encoded by the EEG response. Model 5 is identical to the full model but with two additional features corresponding to the first phoneme of perception and production trials in an effort to see how onset responses contribute to the model's predictive power, for a total of 21 $(19 + 2)$ features. The differences in Model 2-5 performance from the full model (Model 1) were evaluated (see §2.8). Models 2, 3 and 4 performed significantly different from the full model at a $p < 0.01$ significance threshold, while Model 5 did not ($p = 0.21$). These results suggest that phonological features were encoded differently during perception and production (Model 2), predictability of speech perception was encoded by the EEG response (Model 3), modality (perception/production) of speech was encoded by the EEG response (Model 4), and sentence-onset phonemes were not encoded differently by the EEG (Model 5).

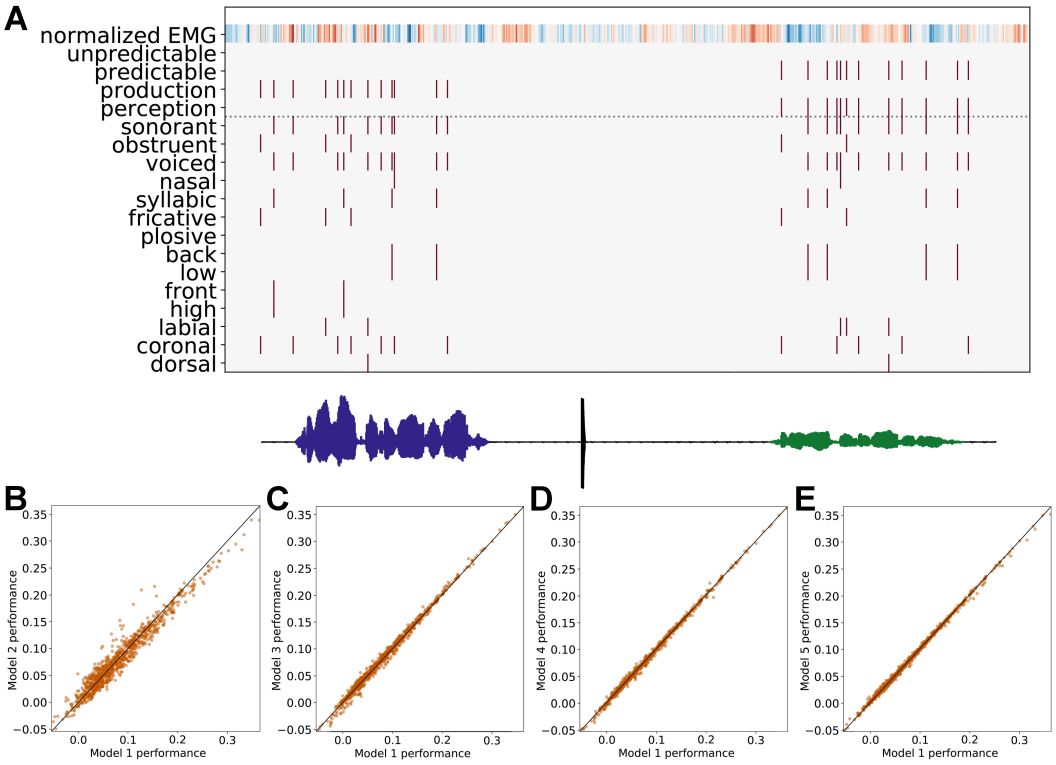

Figure 3.5: Regression schematic and model comparison. Panel A: Regression schematic displaying all features in Model 1 for an example trial color-coded by production (blue) and perception (green). A horizontal dotted line divides phonological features (bottom) from task-related features and normalized EMG (top). Panels B,C,D,E: Scatterplots comparing correlation values between Model 1 and the other four models assessed.

Model Number	Number of Features	Feature Contents	Performance Interpretation
1 (full)	19	14 phonological, 2 binary perception / production, 2 binary predictable / unpredictable, 1 normalized EMG	Speech features are encoded similarly in perception and production
2	47	14 phonological, 14 phonological (perception only), 14 phonological (production only), 2 binary perception / production, 2 binary predictable / unpredictable, 1 normalized EMG	If this outperforms the full model, phonological features are encoded differently during perception and production
3	17	14 phonological, 2 binary perception / production, 1 normalized EMG	If this performs as well as the full model, stimulus predictability during speech perception does not contribute to the EEG response
4	17	14 phonological, 2 binary predictable / unpredictable, 1 normalized EMG	If this performs as well as the full model, differences between speech production and perception do not contribute to the EEG response
5	21	14 phonological, 1 binary perception trial first phoneme, 1 binary production trial first phoneme, 2 binary perception/production, 2 binary predictable / unpredictable, 1 normalized EMG	If this outperforms the full model, sentence-initial phonemes are encoded differently than other phonemes

Table 3.4: Description of assessed models.

The full model performance was also evaluated per channel to observe any topographic differences in encoding accuracy. Linear correlation coefficients (r) and significance values were obtained at each channel (see §2.8).

At each channel, subjects who had significant and insignificant p-values at a $p < 0.01$ significance threshold were tallied (Figure 3.6). Overall, frontal and central electrodes demonstrated higher linear correlation coefficients and a greater proportion of significant correlations compared to parietal and occipital electrodes. There did not appear to be an effect of hemispheric laterality.

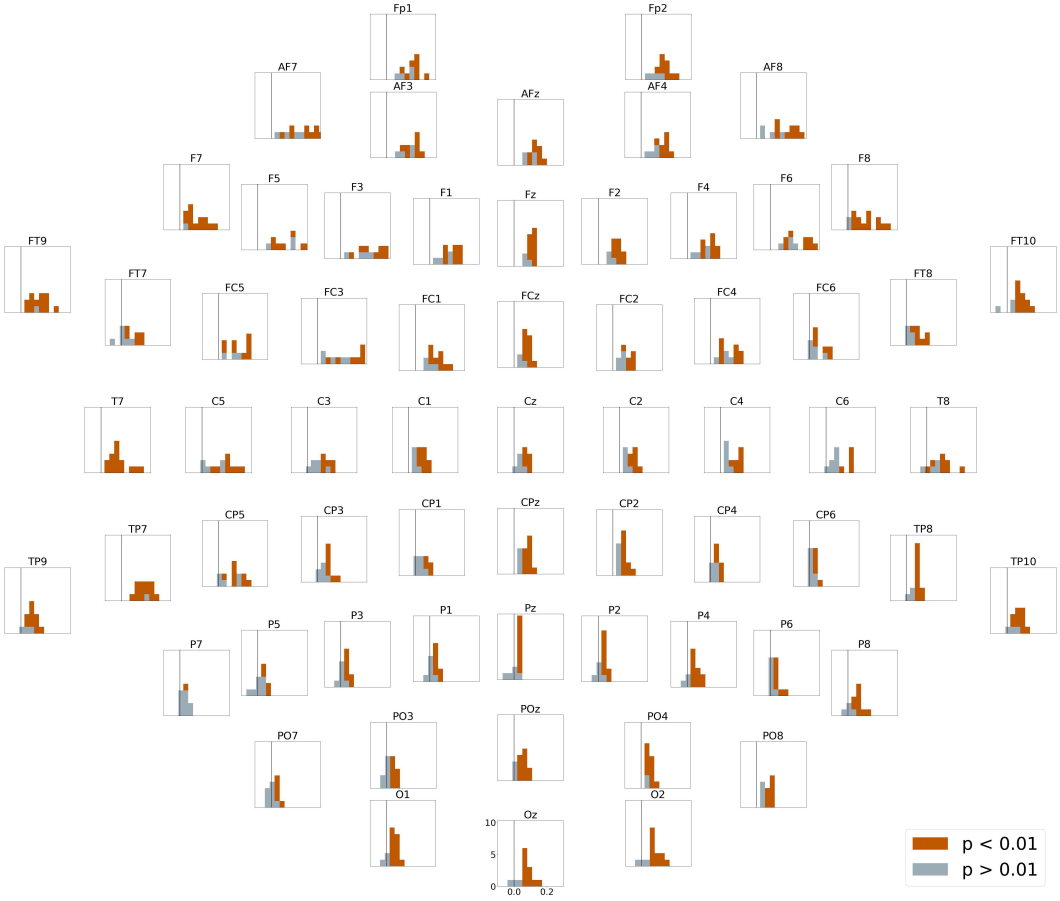

Figure 3.6: Histogram tallying individual subjects' correlation values for Model 1 split by channel. Bins are color-coded according to significance threshold.

3.3.1 Differences Between Speech Production and Perception

As described in Section 3.3, comparing mTRF models that include or exclude stimulus features that contrast between the perception and production components of the task (i.e., Model 1 versus Model 2) suggested that phonological features during speech production and perception were differentially encoded. A significant difference in model performance with (Model 1) and without (Model 4) the inclusion of binary perception/production features (Figure 3.5) also suggested that a distinction between these trial types played a substantial role in the encoding of EEG activity during the task.

To further examine differences in how perception and production are encoded by the EEG, mTRF model weights were examined. In an mTRF model, the weight w of a feature f at a given electrode n provides a measure of how much the feature f contributes to the predicted EEG response (see Figure 2.3 and §2.7). By examining weight differences at different delays in the model, one can construct a visualization of how much individual features contribute to the predicted EEG activity over time. Feature weights for production and perception show a divergent timecourse, with production weights contributing more to model performance before articulation. Delays at which there is a significant difference between perception and production weights are indicated in Figure 3.7 with a black line at the bottom of each channel's plot.

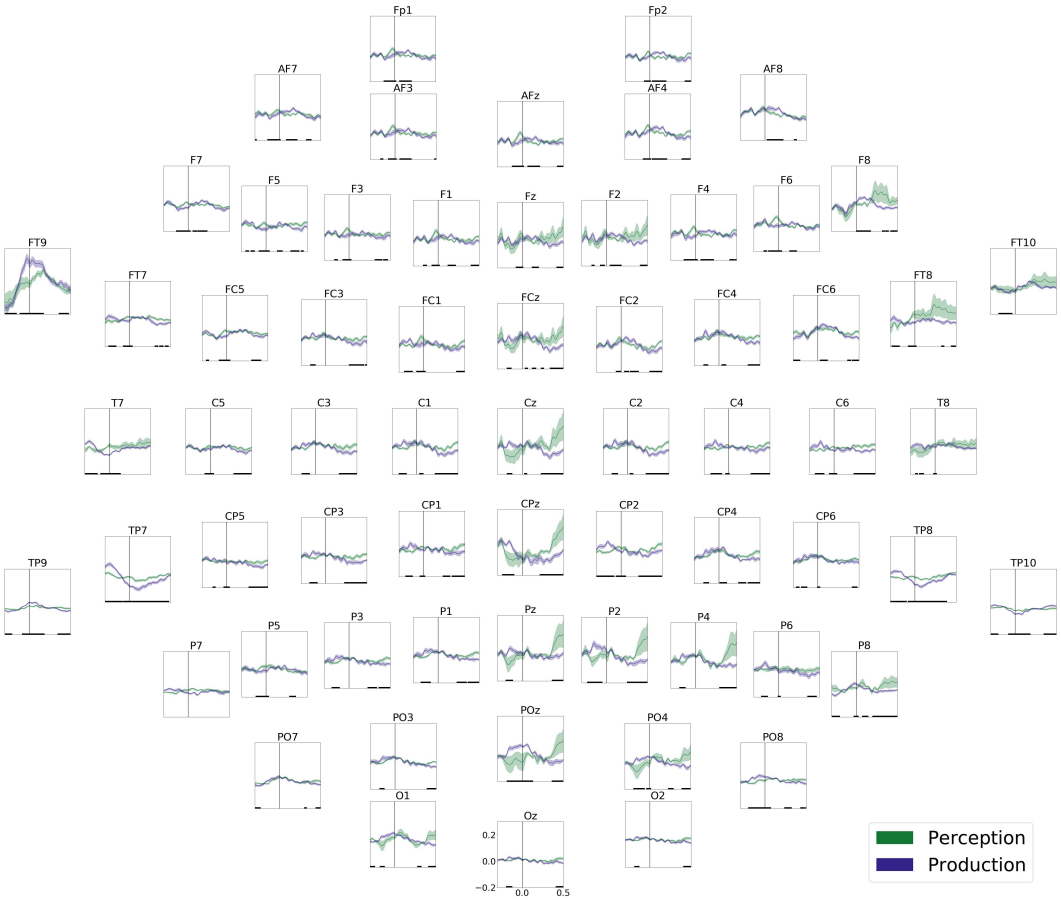

Figure 3.7: Production (blue) versus perception (green) mTRF weights relative to onset of neural activity at the phoneme level by channel. Black horizontal lines indicate delays at which there is a significant difference between the weights as determined via Wilcoxon signed-rank test.

3.3.2 Differences Between Predictable and Unpredictable Speech Perception

In a similar fashion to perception and production trial weights (see §3.3.1), the timecourse of predictable and unpredictable perception trial weights was examined (Figure 3.8). When compared to the contrast between percep-

tion and production weights, the predictable and unpredictable weights had less of a difference in timecourse, although unpredictable weights did show an increase before onset of neural activity, potentially reflecting the block design of the experiment allowing participants to anticipate predictable trials (Figure 3.8, channels FC1, C1, CP1, P1). Interestingly, the channels with larger standard error margins followed a similar pattern to the standard error margin of weights compared in Figure 3.7: midline channels (Fz, FCz, Cz, CPz, Pz, POz) and right/left-lateralized channels (FT9, F8, FT8) had the largest standard error margins.

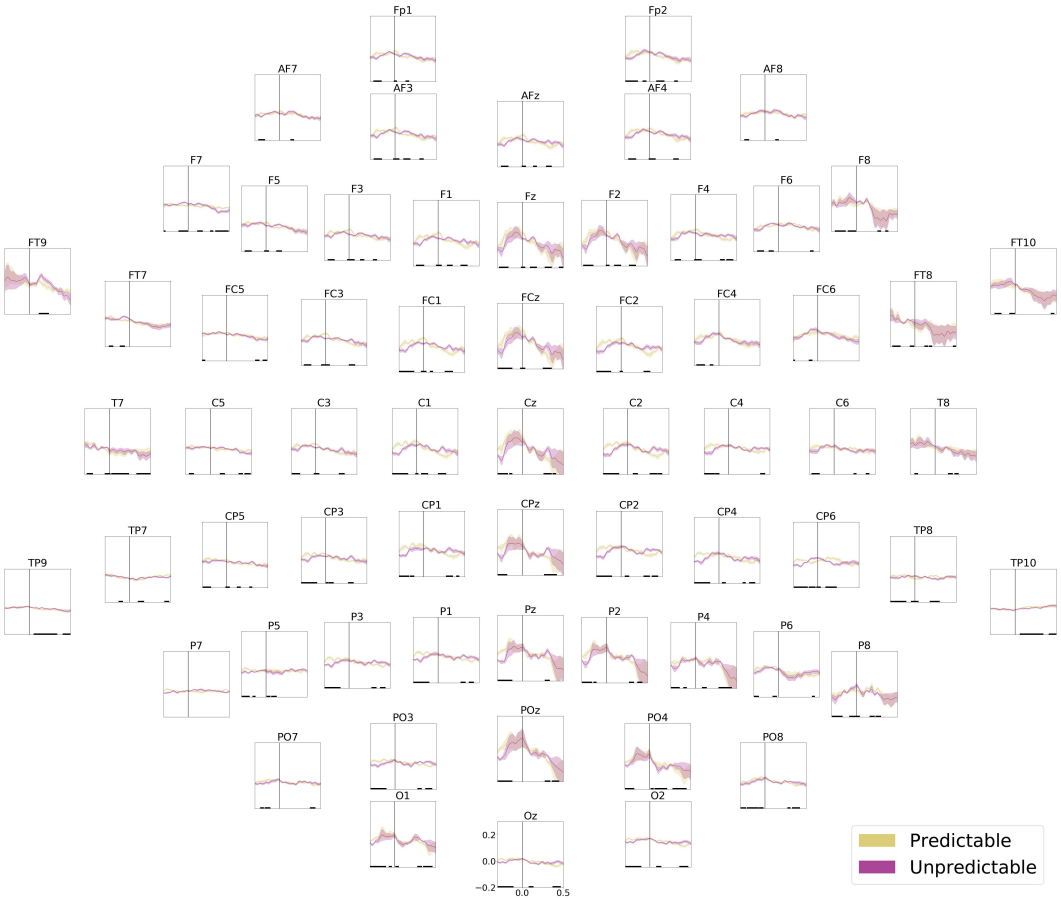

Figure 3.8: Predictable (yellow) versus unpredictable (magenta) speech perception mTRF weights relative to onset of neural activity at the phoneme level by channel. Black horizontal lines indicate delays at which there is a significant difference between the weights as determined via Wilcoxon signed-rank test.

Although N100/P200 components suggest that there is no significant difference between predictable and unpredictable speech perception in this dataset (see §3.2.2), a significant peak-to-peak amplitude difference and significant performance difference between models that include (Model 1) and

exclude (Model 3) binary predictable/unpredictable perception features suggest there is still a trend towards suppression of predictable speech perception compared to unpredictable speech perception.

Chapter 4

Discussion

4.1 EMG Artifact Correction in Naturalistic Speech Production

EMG artifact was perceived by speech production EEG researchers as an insurmountable hurdle for a long period of time (see §1.3.2.3). While studies occasionally successfully analyze speech production data after correcting for EMG artifact, the procedure is far from standardized (Chen et al. 2019; Shackman et al. 2009; Jiang et al. 2019; Islam et al. 2016; Fargier et al. 2018) and analysis of these results is only done at the single-word level (Fargier et al. 2018; Vos et al. 2010; Behroozmand & Larson 2011; Ganushchak et al. 2011). The problem is further complicated by the lack of ground truth in artifact correction, meaning there is not a guaranteed method of confirming both the success and accuracy of an artifact correction technique. In this thesis, I present the application of an EMG artifact correction technique known as Canonical Correlation Analysis (CCA) to naturalistic, sentence-level speech production and demonstrate ways to confirm the effectiveness and accuracy of the technique (see §2.4.1). Analysis of event-related potential (ERP) data in raw EEG data and CCA-corrected EEG data reveals a significant reduction in artifact for CCA-corrected data relative to EMG-related activity recorded from

facial electrodes and task-related activity. These results confirm that EMG activity is being reduced from the data (see §3.1). A nonsignificant reduction in amplitude of responses to inter-trial click tones between CCA-corrected EEG and non-CCA-corrected EEG suggests that the integrity of the neural signal is preserved after CCA.

These techniques are sufficient for validation of CCA artifact correction techniques in this naturalistic data set, but potential additional analyses of CCA efficacy are discussed in Section 4.4.1.

4.2 Differences Between Speech Production and Speech Perception

Differences between speech production and perception were assessed using a linear mixed-effects (LME) model and a multivariate temporal receptive field (mTRF) model. The LME model included the modality of the trial (perception, production) as a fixed effect and the subject as a random effect. Peak to peak amplitude of the N100 and P200 components, as well as the amplitude and latency of the N100 and P200 components, were used as response variables. All response variables examined differed significantly between perception and production trials at a $p < 0.05$ significance rate. This difference is characterized by a reduction in amplitude of the N100 and P200 components in speech production relative to speech perception. This amplitude reduction is likely a reflection of speaking-induced suppression, a commonly observed phenomenon in speech production where self-generated audi-

tory stimuli are reduced in amplitude during speech production (Martikainen et al. 2005; Behroozmand & Larson 2011; Brumberg & Pitt 2019; Cao et al. 2017). Furthermore, I attempted to elaborate on the nature of suppression by using mTRF modeling: is the suppression seen during speech production a general reduction of response amplitude or is certain information encoded during speech perception not encoded during speech production? The significant difference in performance between a model that did not include a perception/production distinction (Model 4) and one that did (Model 1) suggests that differentiating between these modalities is relevant to the recorded EEG response. Furthermore, a model that explicitly differentiated phonological features between perception and production (Model 2) performing differently from one that does not (Model 1) suggests that phonological features are encoded differently between perception and production.

4.3 Differences Between Predictable and Unpredictable Speech Perception

Differences between predictable and unpredictable speech perception were examined using a linear mixed-effects (LME) model and a multivariate temporal receptive field (mTRF) model. The LME model included the stimulus predictability as a fixed effect and the subject as a random effect. Multiple response variables were evaluated using this model (see §3.2.2), and only peak to peak amplitude of the N100/P200 components ($p = 0.03$) and N100 latency ($p = 0.02$) demonstrated a significant difference between predictable and un-

predictable sentences. This effect is reduced when compared to the production versus perception LME (see §3.2.1), as the production versus perception LME had many more significant response variables. The predictable and unpredictable stimulus features in the mTRF model also appeared to have a less divergent timecourse than the comparison between production and perception weights (see §3.3.1 & §3.3.2). However, the fact that a model that excluded predictable and unpredictable stimulus features (Model 3) performed significantly worse than a model that included them (Model 1) suggests that predictability contributes in some way to the recorded EEG activity (see §3.3). Interestingly, a model that did not separately encode sentence-initial phonemes (Model 1) did not perform significantly different from a model that did (Model 5), which suggests that onset responses are not differentially encoded in the EEG, at least in this task. This is in contrast to the onset and sustained response profiles (Hamilton et al. 2018) observed in an ECoG study of sentence perception. As onset and sustained response profiles were both localized to the superior temporal gyrus, it is possible EEG lacks the spatial resolution to properly differentiate between these response profiles. Additionally, production data were included in Model 5, while (Hamilton et al. 2018) solely used perception data, which could explain the difference in results.

The presence of a relatively weaker significant result comparing predictable and unpredictable speech perception has implications for Hypothesis 3. The reduced significant difference between conditions could be due to adjacent competing functional regions averaging out the neural response to these

trials. Areas of the middle temporal gyrus (MTG) have been previously implicated in speech monitoring during production (Zheng et al. 2010; Gauvin et al. 2016), with studies showing a reduction in MTG activity in unpredictable contexts. In altered auditory feedback studies, feedback becomes unpredictable when it is altered to a degree that makes it no longer recognizable to the participant as being internally generated (Behroozmand & Larson 2011; Hashimoto & Sakai 2003). A crucial distinction between altered auditory feedback studies and the study presented in this thesis is that feedback in altered auditory feedback studies is immediate, while feedback in this study was delayed due to the "produce then listen to playback" design of the experiment. The MTG has also been implicated in resolving competing stimuli during speech perception tasks, which suggests it more generally plays a role in feedback situations where the perceived stimulus mismatches the expected stimulus provided by the efference copy (Ashtari et al. 2004; Luthra et al. 2019). Another piece of evidence for the MTG being involved in this auditory process comes from research conducted in people with schizophrenia. Multiple studies have shown abnormal MTG activity in people with schizophrenia who have auditory hallucinations, the mechanism for which is theorized to be an inability to successfully determine if speech is internally or externally produced (McGuire et al. 1995; Woodruff et al. 1997). This suggests that a critical part of normal MTG functioning is processing self-produced speech.

Conversely, regions of the superior temporal gyrus (STG) have been implicated in general speech processing as well as selectively responding to

unpredictable stimuli (Fitzgerald & Todd 2020; Astheimer & Sanders 2011; Ding et al. 2016; Cao et al. 2017). The anatomical proximity of these regions is not discernible using the low spatial resolution of EEG. If these areas are truly active during opposite experimental conditions (MTG for predictable speech perception and STG for unpredictable speech perception), then it is possible these regional differences are cancelling out any amplitude changes that could be observed in a grand average ERP analysis. Further explanations for the lack of a result across these conditions are discussed in Section 4.4.2.

4.4 Limitations

4.4.1 EMG Artifacts

Limitations arise when exploring a new methodological space, such as the study of naturalistic speech production using EEG. A large issue for the interpretation of these results is the lack of ground truth data in EEG: because it is impossible to observe the neural components that make up the electroencephalogram individually, it is always a possibility that a substantial amount of artifact remains in the data. While it is true that this is a fundamental limitation of EEG as a method, it is especially salient as a limitation in this study, as the overt production nature of the task inherently causes large amounts of EMG artifact to be present in the EEG.

Although techniques that confirm the accuracy and effectiveness of CCA artifact correction in removing EMG artifact and preserving signal integrity are presented in this study, additional reliability checks could be em-

ployed by a skeptic researcher. A correlation between the degree of N100/P200 suppression in speech production and the amplitude of the pre-articulatory responses discussed in Section 3.2.1 would strengthen the argument that neural responses are preserved in CCA-corrected data, as the observed suppression is theorized to be related to efference copy generation during pre-articulatory motor speech planning. Pre-articulatory EEG activity would not be affected by EMG artifact while the N100/P200 amplitudes would, so a lack of a correlation between the two would demonstrate N100/P200 amplitude reduction in the production condition was not due to efference copy-related suppression but instead due to CCA subtracting neural activity from the EEG signal. Similarly, examining within-subject N100 amplitude could serve as a check that EMG artifact is removed in the CCA-corrected data. The N100 component is more likely to be affected by EMG activity due to its temporal proximity to the onset of articulation than later neural components, so its integrity could serve as a marker of EMG artifact correction; however, this analysis is not included in this thesis. To quantify N100 integrity, one could compare the difference in mean amplitude of the N100 before and after artifact correction to later components such as P200, and a lack of significant difference in amplitude reduction would suggest the N100 is preserved equally. Another way to quantify N100 integrity would be comparing post-correction perception and production N100 peak amplitudes using nonparametric statistics: a lack of significant difference between the two would suggest N100 preservation after artifact correction.

A possible surface-level interpretation of the reduction in amplitude of N100/P200 components in production trials relative to perception trials (see §3.2.1) is that artifact correction techniques asymmetrically affected production relative to perception due to speech production-related EMG artifact. Such an interpretation is reliant on the presence of Type II error in the corrected dataset. The preservation of the N1-P2 response to inter-trial broadband click tones (see §3.1.1) suggests that the reduction in amplitude seen in sentence-level ERP analyses is not due to overcorrection of the data.

4.4.2 Stimulus Predictability

The effect size between predictable and unpredictable speech perception is less than what I expected when designing this study. One potential explanation is that activity from two adjacent but functionally different regions of the temporal cortex are cancelling each other out during the averaging process. Some neural populations demonstrate an increase in activity to errorful/unpredictable stimuli (Fitzgerald & Todd 2020; Bishop & Hardiman 2010; Hawco et al. 2009) while other neural populations instead have shown suppression to errorful/unpredictable stimuli (Niziolek et al. 2013; Zheng et al. 2010; Gauvin et al. 2016). If this is the case, then the low spatial resolution of EEG is an inherent limitation to this study. It is possible that EEG lacks the spatial resolution to have isolated these two competing regions/patterns of activation in the context of this task. Future research that incorporates invasive electrocorticography would preserve the necessary high temporal resolution of

EEG for speech production research but allow for examination of individual cortical structures' contributions to the recorded response.

The lack of a significant difference between predictable and unpredictable speech perception stimuli may be also due to the scale of the task. While using naturalistic stimuli in neurolinguistic experiments has many benefits over using more constrained stimuli, one potential limitation of naturalistic stimuli is an inability to operationalize and isolate specific neural responses using classic "independent variable, dependent variable" experimental design. Many studies using naturalistic stimuli instead opt for computational methods that do not make assumptions about how the recorded data is organized. That is to say, a hypothesis about how neural responses differ within two very constrained conditions (i.e., predictable and unpredictable) may not be appropriate for a study using naturalistic stimuli.

Another consideration is that a robust suppression of predictable trials was not observed because the task, although using naturalistic stimuli for EEG research standards, was still relatively constrained compared to self-generated conversational speech. Reading scripted sentences off an iPad is a less naturalistic task than effortlessly conversing with other people as a part of daily life. However, studies that have used much less naturalistic stimuli, such as vocalization (Behroozmand & Larson 2011; Hawco et al. 2009), single words (Zheng et al. 2010; Astheimer & Sanders 2011) and pressing a button to play a sound (Martikainen et al. 2005) have demonstrated suppression to predictable stimuli, which suggests that abstraction from natural speech is not an expla-

nation for the weaker difference between predictable and unpredictable trials. These studies also differ in whether they distribute unpredictable trials in a random manner (Behroozmand & Larson 2011; Hawco et al. 2009; Astheimer & Sanders 2011) or in a predictable-unpredictable block design, similar to what this study used (Zheng et al. 2010; Martikainen et al. 2005). Previous studies that find significiant differences between predictable and unpredictable stimuli using a block design suggest that the block design of this study is not the reason for fewer significant differences between predictable and unpredictable trials when compared to perception versus production trials.

4.5 Future Directions

One benefit to a study using naturalistic stimuli is that large quantities of data compared to more constrained tasks can be gathered (Hamilton & Huth 2020). This opens the door for many future analyses to be performed on this dataset.

4.5.1 Levels of Linguistic Representation and Other Parameters

The ERP analyses present in this thesis are restricted exclusively to the sentence level; however, word and phoneme-level timing information was also collected (see §2.5). This presents many different possible analyses, for example, does production-related suppression exist at smaller units of linguistic representation? Additionally, Hamilton et al's work on onset and sustained responses (Hamilton et al. 2018) which motivated the predictable/unpredictable

condition split focused on the sentence level, studies examining voice onset time have found onset-specific responses at the word level (Fargier et al. 2018; Luthra et al. 2019). Although stimulus predictability was not manipulated at the word level in this experiment, it is possible that examination of ERP responses to words and phonemes could differ by other behavioral modifiers present in the task, such as phonetic or morphosyntactic content.

Returning to sentence-level ERP analysis, there were multiple neural components that appeared to differ between perception and production (see §3.2.1 and Figure 3.3). Prearticulatory activity was increased for production relative to perception and could be reflective of some aspect of motor speech planning or programming (such as the efference copy), but this effect was not a part of my hypotheses and therefore was not quantitatively analyzed in this thesis. Neural activity after the P200 component also appears to be divergent between production and perception, with production trials having an increase in amplitude compared to perception. Incorporating a wider range of time series into the linear mixed effects models used in this thesis could provide insight about which later components are potentially involved in differences between speaking and listening.

4.5.2 Error Analysis

Because the findings for differences between predictable and unpredictable stimuli were not robust, I am interested in studying other possible causes of suppression within the context of this dataset. One potential cause

is errorful speech production: neural responses to errorful speech have been shown to be different from errorless speech in previous studies (Niziolek et al. 2013; Hawco et al. 2009; Gauvin et al. 2016; Masaki et al. 2001). Due to the complexity of sentences used as production stimuli in this task, (see §2.3) errors are present to some degree in most participants. Additionally, because production stimuli are used to create perception stimuli, any produced errors are also present in the perception condition. Analyzing this subset of errorful trials using similar ERP and mTRF techniques presented in this thesis could yield interesting observations about how the brain monitors and responds to inaccurate speech.

4.5.3 Decoding Speech Features from EEG

As stated above, the size of this dataset allows for implementation of computational analyses, such as mTRF modeling (see §1.3.3). mTRF models attempt to explain how neural populations *encode* stimulus features; however, *decoding* models are increasing in popularity in the field of computational auditory neuroscience (Cooney et al. 2018; Pasley et al. 2012; Moses et al. 2016), motivated in large part by brain-computer interface (BCI) research. Because studying sentence-level speech production with EEG is a novel method, attempting to decode various linguistic features using this dataset could serve as a proof-of-concept for those wishing to use EEG-based speech BCI systems in naturalistic contexts.

4.5.4 Naturalistic Speech Production in Communication Disorders

Lastly, as mentioned in the introduction, feedback speech motor control is theorized to break down in multiple psychological and communication disorders (Heinks-Maldonado et al. 2007; McGuire et al. 1995; Woodruff et al. 1997; Ballard et al. 2018; Daliri et al. 2018; Toyomura et al. 2020; Hoffman 2014; Parrell et al. 2017), including schizophrenia, apraxia of speech, fluency disorders, and neurodegenerative diseases such as Parkinson's disease. Treatment of neurogenic communication disorders in particular often focuses on functional aspects of communication, that is, how a disorder prevents an individual from participating in the activities of daily living (Ingham et al. 2012; Chapey et al. 2000; Stokes 2011). As treatment goals shift from more constrained to more naturalistic, advances in the treatment of such disorders could be benefitted by their study in more naturalistic contexts, such as the ones presented in this thesis. Understanding how the dynamics of speech production function at higher levels of linguistic representation (such as sentences) could provide more direct insights about how disorders such as stuttering and apraxia of speech impact the functional communication these individuals use in daily life.

Chapter 5

Conclusion

A suppression of production trials relative to perception trials is indicative of speaking-induced suppression. This result, alongside mTRF modeling demonstrating the preservation of phonological tuning in speech production, points to the efference copy present in models of motor speech control as a potential reason for an amplitude reduction in production trials. However, speaking-induced suppression does not appear to be caused solely by stimulus predictability, as predictable and unpredictable speech perception were significantly different for fewer response variables than speech production and perception.

Application of CCA to sentence-level EEG demonstrates it is possible to preserve integrity of neural responses while simultaneously removing EMG artifact. Neuroscientists interested in the study of language have avoided studying naturalistic speech production using EEG because of the fear, albeit a valid fear, of EMG artifact rendering the data too noisy to analyze. The results presented in this thesis have implications for those looking to study the neural basis of communication disorders that affect speech production in a more naturalistic context, those looking to use EEG to create brain-computer

interfaces for speech production, or any researchers interested in utilizing naturalistic stimuli in speech production studies.

Appendix 1

Individual Subject Predictability Plots

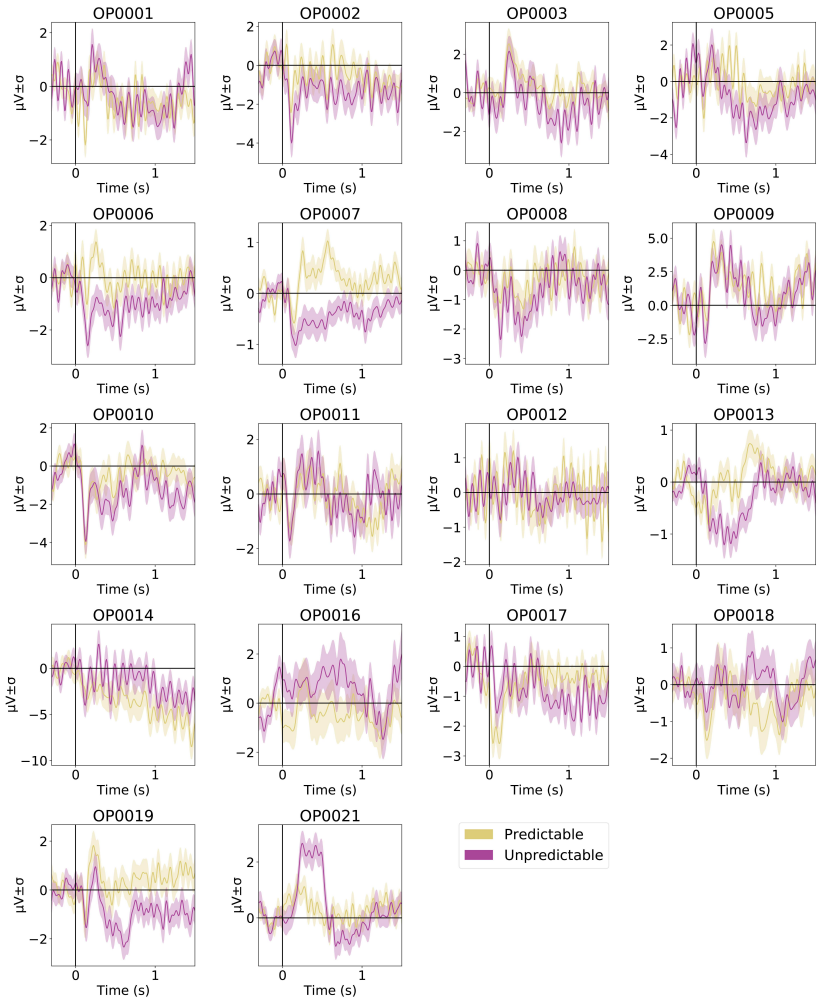

Figure 1.1: ERP comparison between predictable (yellow) and unpredictable (magenta) speech perception trials separated by individual subject. Activity is epoched to sentence onset.

Bibliography

Alday, P. (2018). Philistine: Utility functions for EEG and statistical analysis in Python, version 0.1.

Ashtari, M., Lencz, T., Zuffante, P., Bilder, R., Clarke, T., Diamond, A., Kane, J., & Szeszko, P. (2004). Left middle temporal gyrus activation during a phonemic discrimination task. Neuroreport, 15, 389–393.

Astheimer, L. B. & Sanders, L. D. (2011). Predictability affects early perceptual processing of word onsets in continuous speech. Neuropsychologia, 49, 3512–3516.

Bahl, L. R., Jelinek, F., & Mercer, R. L. (1983). A maximum likelihood approach to continuous speech recognition. IEEE Trans. Pattern Anal. Mach. Intell., 5, 179–190.

Ballard, K. J., Halaki, M., Sowman, P., Kha, A., Daliri, A., Robin, D. A., Tourville, J. A., & Guenther, F. H. (2018). An investigation of compensation and adaptation to auditory perturbations in individuals with acquired apraxia of speech. Front. Hum. Neurosci., 12, 510.

Barlow, J. S. & Dubinsky, J. (1980). EKG-artifact minimization in referential EEG recordings by computer subtraction. Electroencephalogr. Clin. Neurophysiol., 48, 470–472.

Behroozmand, R. & Larson, C. R. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. BMC Neurosci., 12, 54.

Bell, A. J. & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. Neural Comput., 7, 1129–1159.

Bendixen, A., Scharinger, M., Strauß, A., & Obleser, J. (2014). Prediction in the service of comprehension: modulated early brain responses to omitted speech segments. Cortex, 53, 9–26.

Benjamini, Y. & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Ann. Stat., 29, 1165–1188.

Berg, P. & Scherg, M. (1991). Dipole models of eye movements and blinks. Electroencephalogr. Clin. Neurophysiol., 79, 36–44.

Bishop, D. V. M. & Hardiman, M. J. (2010). Measurement of mismatch negativity in individuals: A study using single-trial analysis.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. Glot. Int., 5, 341–345.

Broca, P. (1861). Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). Bulletin et Memoires de la Societe anatomique de Paris, 6, 330–357.

Brumberg, J. S. & Pitt, K. M. (2019). Motor-Induced suppression of the N100 Event-Related potential during motor imagery control of a speech synthesizer Brain-Computer interface. J. Speech Lang. Hear. Res., 62, 2133–2140.

Burgess, R. C. (2020). Recognizing and correcting MEG artifacts. J. Clin. Neurophysiol., 37, 508–517.

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. J. Acoust. Soc. Am., 103, 3153–3161.

Cao, L., Thut, G., & Gross, J. (2017). The role of brain oscillations in predicting self-generated sounds. Neuroimage, 147, 895–903.

Chang, E. F. (2015). Towards large-scale, human-based, mesoscopic neurotechnologies. Neuron, 86, 68–78.

Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. Proc. Natl. Acad. Sci. U. S. A., 110, 2653–2658.

Chapey, R., Duchan, J. F., Elman, R. J., Garcia, L. J., Kagan, A., Lyon, J. G., & Mackie, N. S. (2000). Life participation approach to aphasia: A statement of values for the future.

Chen, X., Xu, X., Liu, A., Lee, S., Chen, X., Zhang, X., McKeown, M. J., & Wang, Z. J. (2019). Removal of muscle artifacts from the EEG: A review and recommendations. IEEE Sens. J., 19, 5353–5368.

Computing, R. & Others (2013). R: A language and environment for statistical computing. Vienna: R Core Team.

Cooney, C., Folli, R., & Coyle, D. (2018). Neurolinguistics research advancing development of a Direct-Speech Brain-Computer interface. iScience, 8, 103–125.

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The multivariate temporal response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli. Front. Hum. Neurosci., 10, 604.

Daliri, A., Wieland, E. A., Cai, S., Guenther, F. H., & Chang, S.-E. (2018). Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter. Dev. Sci., 21.

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

De Clercq, W., Vergult, A., Vanrumste, B., Van Paesschen, W., & Van Huffel, S. (2006). Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. IEEE Trans. Biomed. Eng., 53, 2583–2587.

Delorme, A. & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods, 134, 9–21.

Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-Frequency cortical entrainment to speech reflects Phoneme-Level processing. Curr. Biol., 25, 2457–2465.

Dimigen, O. (2020). Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments. Neuroimage, 207, 116117.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. Nat. Neurosci., 19, 158–164.

Fargier, R., Bürki, A., Pinet, S., Alario, F.-X., & Laganaro, M. (2018). Word onset phonetic properties and motor artifacts in speech production EEG recordings. Psychophysiology, 55.

Fitzgerald, K. & Todd, J. (2020). Making sense of mismatch negativity. Front. Psychiatry, 11, 468.

Friedman, B. H. & Thayer, J. F. (1991). Facial muscle activity and EEG recordings: redundancy analysis. Electroencephalogr. Clin. Neurophysiol., 79, 358–360.

Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J., & Turner, R. (1996). Movement-Related effects in fMRI time-series.

Ganushchak, L. Y., Christoffels, I. K., & Schiller, N. O. (2011). The use of electroencephalography in language production research: A review.

Gao, J., Zheng, C., & Wang, P. (2009). Automatic removal of ocular artifacts from EEG signals.

Gauvin, H. S., De Baene, W., Brass, M., & Hartsuiker, R. J. (2016). Conflict monitoring in speech processing: An fMRI study of error detection in speech production and perception. Neuroimage, 126, 96–105.

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

Gomez-Herrero, G., De Clercq, W., Anwar, H., Kara, O., Egiazarian, K., Van Huffel, S., & Van Paesschen, W. (2006). Automatic removal of ocular artifacts in the EEG without an EOG reference channel. In Proceedings of the 7th Nordic Signal Processing Symposium - NORSIG 2006, pp. 130–133.

Goncharova, I. I., McFarland, D. J., Vaughan, T. M., & Wolpaw, J. R. (2003). EMG contamination of EEG: spectral and topographical characteristics. Clin. Neurophysiol., 114, 1580–1593.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Parkkonen, L., & Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. Neuroimage, 86, 446–460.

Greenlee, J. D. W., Behroozmand, R., Larson, C. R., Jackson, A. W., Chen, F., Hansen, D. R., Oya, H., Kawasaki, H., & Howard, 3rd, M. A. (2013). Sensory-motor interactions for vocal pitch monitoring in non-primary human auditory cortex. PLoS One, 8, e60783.

Greenlee, J. D. W., Jackson, A. W., Chen, F., Larson, C. R., Oya, H., Kawasaki, H., Chen, H., & Howard, 3rd, M. A. (2011). Human auditory cortical activation during self-vocalization. PLoS One, 6, e14744.

Hamilton, L. S., Edwards, E., & Chang, E. F. (2018). A spatial map of onset and sustained responses to speech in the human superior temporal gyrus. Curr. Biol., 28, 1860–1871.e4.

Hamilton, L. S. & Huth, A. G. (2020). The revolution will not be controlled: natural stimuli in speech neuroscience.

Hashimoto, Y. & Sakai, K. L. (2003). Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: An fMRI study. Hum. Brain Mapp., 20, 22–28.

Hawco, C. S., Jones, J. A., Ferretti, T. R., & Keough, D. (2009). ERP correlates of online monitoring of auditory feedback during vocalization. Psychophysiology, 46, 1216–1225.

Hayes, B. (2011). Introductory Phonology. (John Wiley & Sons).

Heinks-Maldonado, T. H., Mathalon, D. H., Houde, J. F., Gray, M., Faustman, W. O., & Ford, J. M. (2007). Relationship of imprecise corollary discharge

in schizophrenia to auditory hallucinations. Arch. Gen. Psychiatry, 64, 286–296.

Heinks-Maldonado, T. H., Nagarajan, S. S., & Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. Neuroreport, 17, 1375–1379.

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

Hoffman, P. R. (2014). Factors influencing the effects of delayed auditory feedback on dysarthric speech associated with parkinson 's disease.

Houde, J. F. (1998). Sensorimotor adaptation in speech production.

Houde, J. F. & Chang, E. F. (2015). The cortical computations underlying feedback control in vocal production. Curr. Opin. Neurobiol., 33, 174–181.

Houde, J. F. & Nagarajan, S. S. (2011). Speech production as state feedback control. Front. Hum. Neurosci., 5, 82.

Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: an MEG study. J. Cogn. Neurosci., 14, 1125–1138.

Hullett, P. W., Hamilton, L. S., Mesgarani, N., Schreiner, C. E., & Chang, E. F. (2016). Human superior temporal gyrus organization of spectrotemporal modulation tuning derived from speech stimuli. J. Neurosci., 36, 2014–2026.

Ingham, R. J., Ingham, J. C., & Bothe, A. K. (2012). Integrating functional measures with treatment: a tactic for enhancing personally significant change in the treatment of adults and adolescents who stutter. Am. J. Speech. Lang. Pathol., 21, 264–277.

Islam, M. K., Rastegarnia, A., & Yang, Z. (2016). Methods for artifact detection and removal from scalp EEG: A review. Neurophysiol. Clin., 46, 287–305.

Jiang, X., Bian, G.-B., & Tian, Z. (2019). Removal of artifacts from EEG signals: A review. Sensors, 19.

Jones, E., Oliphant, T., Peterson, P., & Others (2001). SciPy: Open source scientific tools for python.

Keough, D., Hawco, C., & Jones, J. A. (2013). Auditory-motor adaptation to frequency-altered auditory feedback occurs when participants ignore feedback. BMC Neurosci., 14, 25.

Keren, A. S., Yuval-Greenberg, S., & Deouell, L. Y. (2010). Saccadic spike potentials in gamma-band EEG: characterization, detection and suppression. Neuroimage, 49, 2248–2263.

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmertest package: Tests in linear mixed effects models. Journal of Statistical Software, Articles, 82, 1–26.

Lester-Smith, R. A., Daliri, A., Enos, N., Abur, D., Lupiani, A. A., Letcher, S., & Stepp, C. E. (2020). The relation of articulatory and vocal Auditory–Motor control in typical speakers. Journal of Speech, Language, and Hearing Research, 63, 3628–3642.

Lightfoot, G. (2016). Summary of the N1-P2 cortical auditory evoked potential to estimate the auditory threshold in adults. Semin. Hear., 37, 1–8.

Lijffijt, M., Lane, S. D., Meier, S. L., Boutros, N. N., Burroughs, S., Steinberg, J. L., Moeller, F. G., & Swann, A. C. (2009). P50, n100, and P200 sensory gating: relationships with behavioral inhibition, attention, and working memory. Psychophysiology, 46, 1059–1068.

Luck, S. J. (2014). An Introduction to the Event-Related Potential Technique.

Luck, Steven J., & Kappenman, Emily S. (2011). The Oxford Handbook of Event-Related Potential Components.

Luthra, S., Guediche, S., Blumstein, S. E., & Myers, E. B. (2019). Neural substrates of subphonemic variation and lexical competition in spoken word recognition. Lang Cogn Neurosci, 34, 151–169.

Makeig, S. & Others (1996). Advances in neural information processing systems 8, eds. d. touretzky, m. mozer and m. hasselmo.

Martikainen, M. H., Kaneko, K.-I., & Hari, R. (2005). Suppressed responses to self-triggered sounds in the human auditory cortex. Cereb. Cortex, 15, 299–302.

Martin, S., Mikutta, C., Leonard, M. K., Hungate, D., Koelsch, S., Shamma, S., Chang, E. F., Millán, J. D. R., Knight, R. T., & Pasley, B. N. (2018). Neural encoding of auditory features during music perception and imagery. Cereb. Cortex, 28, 4222–4233.

Masaki, H., Tanaka, H., Takasawa, N., & Yamazaki, K. (2001). Error-related brain potentials elicited by vocal errors. Neuroreport, 12, 1851–1855.

MATLAB (2017). version 9.3.0.713579 (R2017b). (Natick, Massachusetts: The MathWorks Inc.).

McGuire, P. K., Silbersweig, D. A., Wright, I., Murray, R. M., David, A. S., Frackowiak, R. S., & Frith, C. D. (1995). Abnormal monitoring of inner speech: a physiological basis for auditory hallucinations. Lancet, 346, 596–600.

McMenamin, B. W., Shackman, A. J., Maxwell, J. S., Bachhuber, D. R. W., Koppenhaver, A. M., Greischar, L. L., & Davidson, R. J. (2010). Validation of ICA-based myogenic artifact correction for scalp and source-localized EEG. Neuroimage, 49, 2416–2432.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007).

The essential role of premotor cortex in speech perception. Curr. Biol., 17, 1692–1696.

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. Science, 343, 1006–1010.

Moses, D. A., Mesgarani, N., Leonard, M. K., & Chang, E. F. (2016). Neural speech recognition: continuous phoneme decoding using spatiotemporal representations of human cortical activity. J. Neural Eng., 13, 056004.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin. Neurophysiol., 118, 2544–2590.

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy represent? evidence from speech production. J. Neurosci., 33, 16110–16116.

Okada, K., Matchin, W., & Hickok, G. (2018). Phonological feature repetition suppression in the left inferior frontal gyrus. J. Cogn. Neurosci., 30, 1549–1557.

Parrell, B., Agnew, Z., Nagarajan, S., Houde, J., & Ivry, R. B. (2017). Impaired feedforward control and enhanced feedback control of speech in patients with cerebellar degeneration. J. Neurosci., 37, 9249–9258.

Parrell, B., Ramanarayanan, V., Nagarajan, S., & Houde, J. (2019). The FACTS model of speech motor control: Fusing state estimation and task-based control. PLoS Comput. Biol., 15, e1007321.

Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., Knight, R. T., & Chang, E. F. (2012). Reconstructing speech from human auditory cortex. PLoS Biol., 10, e1001251.

Perkell, J., Matthies, M., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J., & Guiod, P. (1997). Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. Speech Commun., 22, 227–250.

Rastatter, M. & De Jarnette, G. (1984). EMG activity with the jaw fixed of orbicularis oris superior, orbicularis oris inferior and masseter muscles of articulatory disordered children. Percept. Mot. Skills, 58, 286.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. Science, 274, 1926–1928.

Saramaeki, T., Mitra, S. K., & Kaiser, J. F. (1993). Finite impulse response filter design. Handbook for digital signal processing, 4, 155–277.

Shackman, A. J., McMenamin, B. W., Slagter, H. A., Maxwell, J. S., Greischar, L. L., & Davidson, R. J. (2009). Electromyogenic artifacts and electroencephalographic inferences. Brain Topogr., 22, 7–12.

Shuster, L. I. (2003). fMRI and normal speech production.

Singh, T., Phillip, L., Behroozmand, R., Gleichgerrcht, E., Piai, V., Fridriksson, J., & Bonilha, L. (2018). Pre-articulatory electrical activity associated with correct naming in individuals with aphasia. Brain Lang., 177-178, 1–6.

Stepp, C. E. (2012). Surface electromyography for speech and swallowing systems: Measurement, analysis, and interpretation. Journal of Speech, Language, and Hearing Research.

Stokes, E. K. (2011). International classification of functioning, disability and health (ICF).

Tamburro, G., Stone, D. B., & Comani, S. (2019). Automatic removal of cardiac interference (ARCI): A new approach for EEG data. Front. Neurosci., 13, 441.

Tourville, J. A. & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. Lang. Cogn. Process., 26, 952–981.

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. Neuroimage, 39, 1429–1443.

Toyomura, A., Miyashiro, D., Kuriki, S., & Sowman, P. F. (2020). Speech-Induced suppression for delayed auditory feedback in adults who do and do not stutter. Front. Hum. Neurosci., 14, 150.

Tremblay, P. & Dick, A. S. (2016). Broca and wernicke are dead, or moving past the classic model of language neurobiology. Brain Lang., 162, 60–71.

van den Bunt, M. R., Groen, M. A., Ito, T., Francisco, A. A., Gracco, V. L., Pugh, K. R., & Verhoeven, L. (2017). Increased response to altered auditory feedback in dyslexia: A weaker sensorimotor magnet implied in the phonological deficit.

Van Eijden, T. M., Blanksma, N. G., & Brugman, P. (1993). Amplitude and timing of EMG activity in the human masseter muscle during selected motor tasks. J. Dent. Res., 72, 599–606.

Vos, D. M., Riès, S., Vanderperren, K., Vanrumste, B., Alario, F.-X., Huffel, V. S., & Burle, B. (2010). Removal of muscle artifacts from EEG recordings of spoken language production. Neuroinformatics, 8, 135–150.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia, 41, 989–994.

Wernicke, C. (1874). Der aphasische Symptomencomplex: Eine psychologische Studie auf anatomischer Basis. (Cohn.).

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. Nat. Neurosci., 7, 701–702.

Woodruff, P. W., Wright, I. C., Bullmore, E. T., Brammer, M., Howard, R. J., Williams, S. C., Shapleske, J., Rossell, S., David, A. S., McGuire, P. K., &

Murray, R. M. (1997). Auditory hallucinations and the temporal cortical response to speech in schizophrenia: a functional magnetic resonance imaging study. Am. J. Psychiatry, 154, 1676–1682.

Woolson, R. F. (2007). Wilcoxon signed-rank test. Wiley encyclopedia of clinical trials, pp. 1–3.

Working Group on Manual Pure-Tone Threshold Audiometry (2005). Guidelines for manual Pure-Tone threshold audiometry. Tech. rep., Rockville, MD.

Wrench, A. (1999). The MOCHA-TIMIT articulatory database.

Yuan, J. & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. J. Acoust. Soc. Am., 123, 3878.

Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2010). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. J. Cogn. Neurosci., 22, 1770–1781.

Vita

Garret Lynn Kurteff was born in Monterey, California. He received a Bachelor of Arts in Linguistics and Psychology from the University of California at Berkeley in 2015. During his undergraduate education, he received mentorship from Dr. Joseph J. Campos, Dr. Keith Johnson, and Dr. Lev D. Michael. He worked at Bay Area urban legend/record store Rasputin Music post-graduation, but in 2016, returned to academics as a researcher in the laboratory of Dr. Edward F. Chang at UCSF. At UCSF, he studied the effects of direct electrocotical stimulation on speech and language and patterns of recovery from rapid onset aphasia following resective brain surgery. After two years, he applied to the University of Texas at Austin for enrollment in their doctoral program in Speech, Language & Hearing Sciences. He was accepted and started graduate studies in August, 2018.

Address: grt.krtf@gmail.com

This thesis was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.