The Dissertation Committee for G. Lynn Kurteff
certifies that this is the approved version of the following dissertation:

Cortical Suppression of Auditory Feedback during Speech Production and Perception

Committee:

Liberty S. Hamilton, Supervisor

Maya L. Henry

Rosemary A. Lester-Smith

Jun Wang

Stephanie Ries

Cortical Suppression of Auditory Feedback during Speech Production and Perception

by

G. Lynn Kurteff

DISSERTATION

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

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2024

This dissertation is dedicated to Ann "Granny Annie" Kurteff, for always putting forth her best effort to understand my research. I hope being the first "Dr. Kurteff" makes her proud.

Acknowledgments

Writing a dissertation is not something one can accomplish alone, even if many aspects of the process are incredibly solitary. I would like first to thank Dr. Liberty S. Hamilton, my PhD advisor, academic mentor, and friend for teaching me the majority of what I know about the brain at this point. I also would like to thank Dr. Maansi S. Desai for embarking on this crazy journey with me in 2018. The rest of the Hamilton Lab at UT Austin provided me with much-needed support as well. I would also like to thank Dr. Joseph J. Campos, one of my undergraduate mentors, for emphasizing to me the importance of theoretical motivation in a methods-obsessed translational research field; his voice has been echoing in my head for the better part of a decade at this point. I would also like to thank Dr. Neal P. Fox for his mentorship in the early years of my PhD, particularly for his advice to get a dog in grad school, which proved invaluable. It would be remiss of me to not thank my fiancée, Dr. Aubrey T. Rieder, for her unwavering patience and copyediting skills.

Lastly and most importantly, I would like to thank the epilepsy community for trusting researchers like myself with their neural data; without them, none of this would have been possible.

Cortical Suppression of Auditory Feedback during Speech Production and Perception

Publication No. _____

G. Lynn Kurteff, Ph.D.
The University of Texas at Austin, 2024

Supervisor: Liberty S. Hamilton

Speaker-induced suppression (SIS) is a phenomenon where responses to one's own speech are reduced in amplitude compared to the speech of others. It is unknown how this process modulates other phenomena of the auditory system studied in purely perceptual experiments. For example, "onset" responses to the acoustic onset of a speech stimulus are observed in auditory cortex during listening tasks. Onset responses and the N1 EEG component share a high-amplitude and low-latency profile. SIS suppresses the N1, but it is unknown whether SIS modulates onset responses. This dissertation presents two original studies describing how the brain responds to speech stimuli differently during speaking and listening.

In Chapter 2, I recorded scalp EEG while participants read sentences aloud then passively listened to playback, which generated identical acoustics between speaking and listening conditions. A more naturalistic sentence-level

task allowed investigation of phonological feature tuning via temporal receptive field analysis. The results show that during SIS, phonological feature tuning remains relatively stable between perception and production, demonstrating that suppressive feedback control mechanisms during speech production do not affect the more abstract linguistic representations of the auditory system they suppress.

Chapter 2 investigated the interaction of SIS with higher-order linguistic processing in naturalistic speech but could not assess specific brain structures' roles in processing auditory feedback due to EEG's poor signal-to-noise ratio. In Chapter 3, I recruited participants undergoing epilepsy monitoring procedures that involve surgical implantation of intracranial electrodes, a higher-spatial-resolution recording technique. Participants completed the same task described in Chapter 2, and unsupervised clustering of responses to speaking and listening revealed anatomical trends in the neural data. "Onset suppression" electrodes in bilateral auditory cortex showed onset responses at the start of a sentence during speech perception but not during speech production. Other cortical areas were generally selective to either speaking or listening but did not specifically suppress onset responses. However, "dual onset" electrodes in posterior insula exhibited onset responses to both speaking and listening with a similar latency to "onset suppression" electrodes. Similar to Chapter 2, "onset suppression" and "dual onset" regions showed phonological feature tuning that did not differ during perception and production, but due to the higher signal-to-noise ratio of the dataset, individual electrodes

showed stronger tunings to specific classes of phonological features than what was observed via EEG.

The research presented in this dissertation expands on previous studies of SIS by clarifying which aspects of the neural response are suppressed during speech production. The absence of onset responses during speech production suggests they play a role in auditory stimulus orientation, an unnecessary process during speaking due to predictive feedforward control. The insula, which exhibits onset responses during speaking and listening, likely plays a role in the feedback control of speech by integrating feedback from somatosensory and auditory modalities as a part of the feedback monitoring process. Expanding our understanding of how feedback control mechanisms function in the brain can hopefully lead to better assessment and treatment of disorders of speech motor control such as stuttering and apraxia of speech.

Table of Contents

List of Tables

List of Figures

Chapter 1

Background

Speech production is a complex cognitive and sensorimotor process unique to humans (Hauser et al. 2002). It is an intuitive and essential mode of communication that every human universally develops barring extreme individual counterexamples. Despite its ubiquity, the theoretical and biological faculties of speech are not well understood. While it is accepted that the larynx, tongue, and other vocal organs articulate speech and the auditory system and hearing organs comprehend speech, the way that the central nervous system facilitates speech perception and production is of relatively recent study.

Speech motor control exists at the intersection of speech perception and production and refers to a broad series of processes that govern the planning and production of speech "in real time." While preparing to speak, we take an abstract communicative intent and prepare it for articulation by our vocal organs through a series of linguistic transformations into morphological, phonological, and syllabic information. While we speak, we monitor the sensory outcomes of our speech to ensure what we intend to say is actually said. Most contemporary theoretical models of speech motor control divide the process into *feedforward* (or internal, predictive) and *feedback* (or external,

corrective) systems. These dichotomous systems work together in real time to estimate the sensory consequences of speech, monitor ongoing sensory consequences of speech, detect deviances from the estimation, and update motor programs to correct these deviances. The feedforward control system is concerned with generation and maintenance of the motor program and estimation of the sensory consequences of speech. The feedback control system's purpose is to monitor acoustic and somatosensory feedback from our articulators as we speak. The detection and correction of speech errors emerges from interaction between these systems. This dissertation is primarily concerned with feedback control mechanisms; specifically, how our brain processes the sensory consequences of our own speech mechanisms as we speak.

My motivation to expand upon our knowledge of the neural mechanisms of speech motor control comes in part from the relative paucity of speech production research in the cognitive sciences, which is in turn the product of several extenuating circumstances. Firstly, speech motor control is a process unique to humans, which limits its study using animal models to the basic sensory and motor substrates of the process. Second, because *in vivo* neural recordings often require a surgical procedure, the direct study of uniquely human cognitive faculties such as speech motor control has historically been limited to noninvasive imaging of brain activity or to lesion-based studies, as clinical necessity serves as the sole motivator for obtaining invasive recordings of the human brain. Of course, a lack of high-resolution recording techniques for use in humans is a limitation to research of any cognitive faculties of hu-

mans. But, an increased access to invasive human neuroimaging for research studies is beginning to mitigate this limitation (Chang 2015). Third, and more specific to the study of feedback control during speech production, is the presence of movement artifacts in recordings of speech production. Noninvasive techniques such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) can have their recording quality affected by movement fundamentally associated with articulation of speech. EEG in particular is well-suited to study speech production barring this caveat, as it is a direct measure of the synaptic potential of neurons (Ray et al. 2008; Luck 2014) and it is incredibly temporally precise, an important factor considering changes in articulation occur on a rapid timescale that would be temporally smeared in lower-temporal-resolution hemodynamic/metabolic recording techniques such as fMRI and positron emission tomography (PET), respectively. Despite EEG's utility in studying speech production and its existence as a recording technology since the 1920s (Luck 2014), researchers have generally avoided it as a technique for studying speech production due to its tendency to record extraneous electrical activity from the facial muscles involved in articulation as well. Because the scalp muscles are much closer to the electrodes than the neurons researchers intend to record from, EEG recordings taken during articulation often feature large movement artifacts that obfuscate the study of speech production (Vos et al. 2010; Shackman et al. 2009). To circumvent this limitation, researchers have used "imagined" or "covert" speech to either minimize or completely negate the range of motion of articulators

during speech production (Okada et al. 2018; Shuster 2003). Monitoring of somatosensory and auditory feedback from speech production is an essential component of speech motor control (Houde & Nagarajan 2011; Tourville & Guenther 2011), so such covert approaches to speech production research carry a critical limitation, being that somatosensory and auditory feedback are not generated in the same capacity when articulation is restricted. Another historical workaround is the use of single-syllable or single-word stimuli, which aim to minimize movement artifact by reducing the natural coarticulation present in continuous speech. Fortunately, in the last several years, promising artifact correction techniques have yielded the first studies of continuous overt speech production using scalp EEG that maintain a tolerably high signal-to-noise ratio (Ries et al. 2021; Kurteff et al. 2023), opening the door for more noninvasive studies of speech motor control. One of these studies is included as Chapter 2 of this dissertation.

Intracranial recordings, while not completely free of movement artifact (Bush et al. 2022), are relatively spared from artifact compared to their noninvasive counterparts because the electrodes in a surgically implanted device are much, much closer to the neurons they record from. However, the clinical necessity of surgical implantation of electrodes places fundamental limitations on the accessibility of intracranial EEG data to researchers, so the study of speech motor control still stands to benefit from wider access to a noninvasive technique that can record responses during articulation. Fortunately, as surgical procedures involving intracranial recordings become more

commonplace, this type of data is increasingly made available to researchers in the cognitive sciences. Chapter 3, the second of two results chapters in this dissertation, presents original research on feedback control using intracranial stereo-electroencephalography (sEEG) recordings collected from participants undergoing treatment of medically refractory epilepsy.

In addition to increased access to intracranial recordings, there is an ongoing paradigm shift in parallel fields to cognitive science such as machine learning and linguistics that has motivated my dissertation research. The creation and curation of language corpora has expanded the size of datasets available to neurolinguistic researchers. In tandem with advances in recording technologies and analysis techniques, the contemporary neurolinguist can afford to study processes such as speech production in relatively less constrained contexts. Historically, psycholinguistic and neurolinguistic research has used heavily constrained stimuli such as single vowels or syllables to ensure an adequate number of trials to quantify trends in the data. Studying speech and language instead in ecologically valid contexts (i.e., words, sentences, unconstrained conversation) makes the results more generalizable, facilitating the development of clinical interventions for those for whom the faculties of speech and language are disordered (Hamilton & Huth 2020; Matusz et al. 2019).

To summarize, the current state of cognitive-linguistic/speech science research is punctuated by unparalleled access to high fidelity recordings of the human brain made possible by recent technological advancements in recording and computing hardware. Scientists can collect more high-quality data and

can use it to fit more complex statistical models in more naturalistic contexts. Speech production is one specific area of study that serves to benefit from these advancements, as its temporal precision and complex interactions between multiple brain regions have hindered its study with historically available lower resolution methods. In my dissertation, I will describe two experiments on feedback processing during speech production that I designed, collected data for, and analyzed during my tenure as a PhD student. Chapter 2 is a noninvasive scalp EEG study of speaker-induced suppression during a reading and listening task, the results of which are also published in *Journal of Cognitive Neuroscience* (Kurteff et al. 2023). Chapter 3 is an invasive sEEG study collected in patients undergoing intracranial monitoring for epilepsy surgery that uses the same materials as the first study. These results are currently undergoing peer review and are available as a preprint on *bioRxiv* Kurteff et al. (2024). The rest of Chapter 1 will provide a review of the concepts necessary to provide context and motivation for these experiments.

1.1 Speech production and speech motor control

Speech motor control refers to a series of online predictive and corrective cognitive and sensorimotor mechanisms taking place during articulation. This is also referred to in the literature as *speech monitoring* (Gauvin & Hartsuiker 2020). The "speech" in "speech motor control" suggests an extent of domain-specificity, but animal models suggest speech motor control is an emergent process of domain-general motor control mechanisms. For example,

Schneider et al. (2014) demonstrated using intracellular recordings that mice who run on a treadmill that generates a certain tone during locomotion begin to inhibit neural responses to that tone, an example of auditory-motor interaction outside the speech domain and a likely precursor to the suppressive mechanisms at play during speech motor control in humans, namely speaker-induced suppression (§1.2.3.1). To connect general motor control mechanisms to speech motor control, studies like Martikainen et al. (2005) have found a similar inhibition of neural responses to a tone generated by button presses in the human brain. Both studies presented in Chapters 2 and 3 of this dissertation investigate neural responses to self-generated auditory feedback during speaking, putting them in the domain of speech motor control. I am specifically interested at looking at the spatial and temporal dynamics of suppressive mechanisms during speech production, which I will review in §1.2.3.1. First, I will provide a review of the theory of speech motor control, as the barriers to studying speech production with neuroimaging methods have created a deeper literature on speech motor control in purely behavioral psycholinguistic studies.

Willem "Pim" Levelt's work in the 1990s serves as a progenitor for modern theories of speech production and speech motor control(Levelt 1993). Levelt's model of speech production outlined a sequential, multi-step model of the processes required to formulate a thought into an articulated utterance, beginning with retrieval of lexical-semantic information (which he refers to as *lemmas*), followed by a series of morphosyntactic transformations, then phono-

logical encoding. At this point, the speaker knows what words to say, what order to say them in, and what sounds are necessary to produce those words. This linguistic code is whisked away to what Levelt described as "output systems," where articulation would be carried out by the vocal organs and parts of the central nervous system that govern those organs. While these output systems were the final steps in Levelt's model of speech production, they're only the first steps for speech motor control.

A simplified schematic of a theoretical framework for speech motor control is provided in Figure 1.1. At a macroscopic level, models of speech motor control split the process into *feedforward* and *feedback* control systems. The feedforward system utilizes a phonetic program akin to the final component of Levelt's original model of speech production, meaning feedforward control begins prior to articulation with the transformation of an abstract linguistic-phonological code into a series of explicit articulator movements. Articulatory kinematics have corroborated physiological reality in neural representations: Chartier et al. (2018) used linear modeling to show that sensorimotor cortex tracks the kinematics of speech articulators like the upper/lower lip, jaw, larynx, and tongue in real time during articulation.

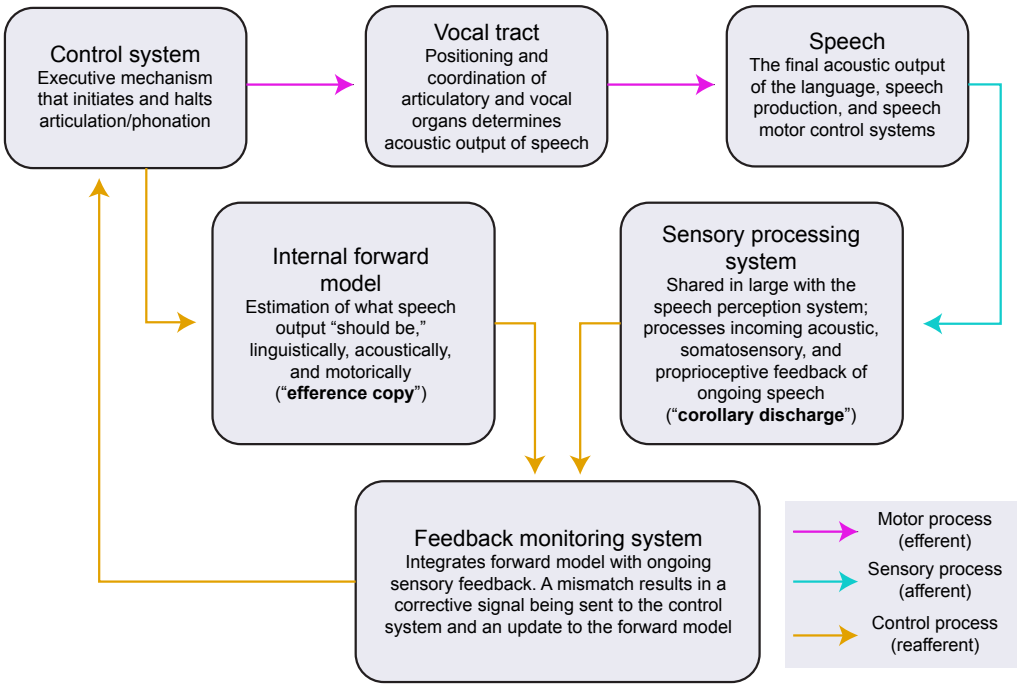

Figure 1.1: **Simplified schematic of speech motor control**. Adapted from models presented in Hickok (2014); Shadmehr & Krakauer (2008); Houde & Nagarajan (2011).

When speech is initiated, the control system sends the articulatory commands to the vocal organs and simultaneously constructs an internal forward model of speech. This feedforward control mechanism is commonly referred to as the *efference copy* (or *internal loop*), an important component of every contemporary model of speech motor control. The efference copy represents a set of sensory expectations about the content of an utterance that is generated during prearticulatory planning of that utterance (Greenlee et al. 2013; Behroozmand & Larson 2011; Zheng et al. 2010; Hawco et al. 2009; Hashimoto

& Sakai 2003). Minimally, the efference copy contains a representation of the expected auditory, somatosensory, and proprioceptive consequences of a given utterance. The rapid timescale of articulation necessitates the existence of a feedforward predictive mechanism like the efference copy. The brain's perceptual system processes changes in auditory feedback with a ∼30-100 millisecond latency, meaning a purely feedback-based system could not correct changes quickly enough for what is observed in natural speech (Houde & Nagarajan 2011). Feedforward control also explains how speakers are able to accurately control speech even in situations where sensory feedback is unavailable or unreliable.

The goal of the feedback control system is to monitor the relevant sensory output (or *corollary discharge*, *outer loop*) generated by the speaker in real time. Ongoing comparisons between feedforward expectations and the feedback sensations are made as the speaker articulates by a feedback monitoring system. The underlying phonetic-to-acoustic transformation that drives feedforward control also generates these expectations about the acoustic and proprioceptive/tactile consequences of speech. Any dissonance between the predictions of the feedforward control system and sensory processing of the feedback control system means the speaker has somehow failed to articulate the intended utterance, which triggers a series of corrective processes and a re-updating of the forward model so that speech can continue onwards as planned. This process is usually split into *error detection* and *error correction*.

1.1.1 Theoretical models of speech motor control

The Directions Into Velocities of Articulators, or DIVA model, is a popular anatomically and computationally explicit model of speech motor control (Tourville & Guenther 2011). At a macroscopic level, the model splits speech motor control into a feedforward and feedback system similarly to the simplified model I presented in Figure 1.1. During pre-articulatory speech preparation, a speech sound map is generated that transforms a phonetic program generated by the language network into explicit velocities of articulators (hence the name of the model). The feedback component of DIVA conceptualizes the efference copy as a series of "auditory and somatosensory target maps" (Tourville & Guenther 2011). Also generated during feedback control are inhibitory "error maps" that, assuming proper accordance between feedback perception and the predictions of the efference copy, suppresses the neural responses to sensory feedback. In this dissertation, I refer to this phenomenon as *speaker-induced suppression*, which I will discuss further below (§1.2.3.1). Error detection is the responsibility of the feedback control system in DIVA; when errors are detected, an error signal is sent to a feedback control region in ventral motor cortex, which converts the perceived errors into corrective motor commands. The corrections are then re-integrated into the feedforward controller.

The hierarchical state-feedback control (HSFC) and Feedback-Aware Control of Tasks in Speech (FACTS) models are an additional family of theoretical models of speech motor control that draw inspiration from feedback per-

turbation research and motor control research in non-speech domains (Houde & Nagarajan 2011; Parrell et al. 2019). In feedback perturbation paradigms, sensory or auditory feedback is manipulated in real time during speech. Speakers are able to rapidly compensate for many different methods of perturbation, suggesting a dynamic role for feedback control. In the HFSC model, the brain first estimates the state of the speech mechanisms, then it generates a set of motor controls based on its estimation. This is in contrast to theories that the full state of the speech mechanisms is available to the motor control network; the authors of the HSFC model argue that such a reality would render any feedforward motor control irrelevant (Houde & Nagarajan 2011). The "feedback" in "state-feedback" refers to the idea that the internal state estimate of the speech mechanisms is updated via sensory feedback. The FACTS model is perhaps best viewed as an update of the HSFC model that assumes the state estimation process present in the HSFC model to be nonlinear in nature to compensate for the influence of the top-down goals of the speaker (i.e., communicative intent; Parrell et al. (2019)).

One criticism of the DIVA[1], HSFC, and FACTS models is that they focus primarily on phonological processing and feedback control without paying enough attention to pre-articulatory speech planning, which likely contains lexical, morphological, and syntactic components as well according to conven-

[1]The updated gradient order DIVA (GODIVA) model Guenther (2016) extends DIVA to account for multisyllabic planning and does contain more explicit theorization of sequence formation than the original DIVA model, but is still ambiguous to how morphosyntactic planning occurs as it is still primarily a speech motor control model.

tional theory (Gauvin & Hartsuiker 2020). For the purposes of my dissertation research, which focuses primarily on feedback control, these criticisms are less noticeable, but I would like to mention two competing theories that pay closer attention to the pre-articulatory and feedforward components of speech motor control. The perceptual loop theory (PLT) posits that both feedforward and feedback control are dependent on perceptual systems (Indefrey & Levelt 2004). PLT has been criticized for assuming that perceiving one's own speech uses the same perceptual processes of perceiving the speech of others, which is contradicted by speaker-induced suppression (§1.2.3.1; Gauvin & Hartsuiker (2020)). Conflict monitoring accounts arose as a response to PLT and infer that feedforward and feedback control are instead dependent on specialized mechanisms of the speech production system (Gauvin et al. 2016). In this model, feedforward control takes place through detection of conflict between potential response options by a domain-general executive system. This is in line with neurophysiological evidence that auditory areas such as the superior temporal gyrus (STG) are silent prior to articulation, but assuming feedback control is strictly within the domain of production systems conflicts with evidence that auditory areas are active (albeit suppressed) during speech production (Cheung et al. 2016).

1.1.2 Neuroanatomy of speech motor control

A broad overview of the neuroanatomy of the speech production, perception, and motor control networks is provided in Figure 1.2, which I will

reference when appropriate throughout this section.

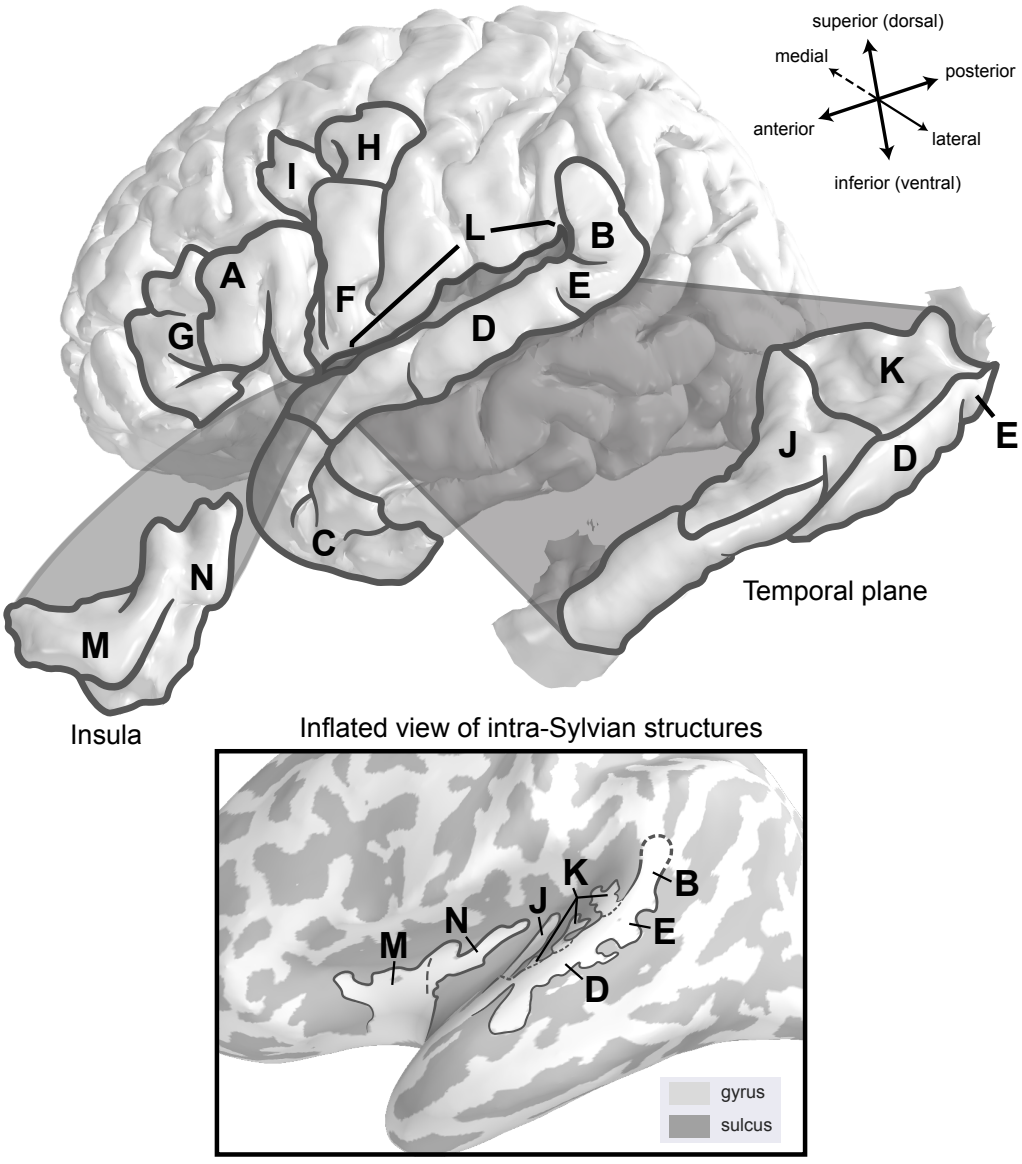

Figure 1.2: **Neuroanatomy of speech perception, speech production, and speech motor control**.
Top: Reconstruction of lateral pial cortex from the `cvs_avg35_inMNI52` template brain. Bottom: Inflated reconstruction of the same template brain. Anatomical boundaries added manually using boundaries from the Destrieux atlas (Destrieux et al. 2010).
(A) Canonical "Broca's area" as popularized by Broca (1865); Geschwind (1970). The definition I endorse here is pars triangularis and pars opercularis of the IFG, although many are utilized in the literature (Tremblay & Dick 2016).
(B) Temporoparietal junction, or "area Spt" (Hickok et al. 2009). Combines with pSTG (E) to form canonical "Wernicke's area."
(C) Anterior temporal lobe / temporal pole.
(D) Middle STG.
(E) Posterior STG. Combines with temporoparietal junction to form canonical "Wernicke's area."
(F) Ventral precentral gyrus. Also referred to as ventral motor cortex. Contains the ventral precentral speech area described by Hickok et al. (2023); its anterior boundary is sometimes confused with Broca's area (Tremblay & Dick 2016). Contains functional regions for laryngeal control (Breshears et al. 2015) and speech arrest (Zhao et al. 2023).
(G) Pars orbitalis of the inferior frontal gyrus.
(H) Middle precentral gyrus. Contains the dorsal precentral speech area described by Hickok et al. (2023). Contains functional regions for laryngeal control (Breshears et al. 2015), speech planning (Silva et al. 2022), and speech arrest (Zhao et al. 2023).
(I) Area 55b, or posterior middle frontal gyrus. Involved in speech planning and recently described in several case studies as an anatomical locus of apraxia of speech (Chang et al. 2020; Levy et al. 2023).
(J) Heschl's gyrus. Sits in the Sylvian fissure atop the STG. Part of primary auditory cortex with PT (K).
(K) Planum temporale. Sits in the Sylvian fissure atop the pSTG. Part of primary auditory cortex with Heschl's gyrus (J).
(L) The Sylvian fissure. Expanded views of intra-Sylvian structures (insula, temporal plane) are provided and also labeled in inflated space.
(M) Anterior insula (short gyrus). Historically an anatomical locus of apraxia of speech (Dronkers 1996).
(N) Posterior insula (long gyrus). Location of the insular auditory field (§4.2).

Neurobiologically, speech production was originally localized to the posterior inferior frontal gyrus (IFG), commonly referred to as "Broca's Area" (Broca 1865) (Figure 1.2A). This view of speech production pervaded research throughout most of the 19th century as early modern theories of language congealed (Geschwind 1970). A three-part model, consisting of Broca's area for speech production, Wernicke's area for speech perception (Figure 1.2B), and the arcuate fasciculus as a white matter pathway linking the two exists almost monolithically in medical textbooks even half a century later in the present day (Tremblay & Dick 2016). Contemporary theories of the neurobiology of language have attempted to update the "Classic Model" described above in a fashion that is in line with modern empirical research. The most popular of these, the Dual Stream Model, takes heavy inspiration from vision research, which has its own dual stream model that it was able to construct due to a several-decade head-start on language research, as vision can be studied in animal models while language cannot (Hickok & Poeppel 2007). Perception and production are not dichotomized to the same extent in the Dual Stream model as they are in the Classic Model, but production is mostly confined to the dorsal "how" stream while perception is mostly confined to the ventral "what" stream. The dorsal stream is conceptualized as the transformation of the abstract intent of speech into concrete articulatory-motor programs through a series of phonological representations. For example of how this dichotomy differs from the production-versus-perception dichotomy of the Classic Model, electrical stimulation of temporal cortex during awake brain surgery can result

in errors of speech production such as naming and phonological errors (Miozzo et al. 2017). These errors are theorized to arise from difficulties with lexical-semantic processing localized to the anterior temporal lobe (Figure 1.2C) and phonological processing localized to the superior temporal gyrus (Figure 1.2D, E), respectively (Jackson et al. 2016; Mesgarani et al. 2014), two forms of representation that are part of the ventral "what" stream but still able to impact speech production despite their localization to temporal cortex.

1.1.2.1 Feedforward control

The neuroanatomy of speech motor control is usually anatomically separated by feedforward and feedback control. The DIVA model is the most anatomically explicit model of speech motor control and prescribes every stage of the process to specific brain regions (Tourville & Guenther 2011). DIVA localizes the anatomical seat of feedforward control to ventral motor cortex (Figure 1.2F) and inferior frontal cortex (Figure 1.2A,G). Ventral sensorimotor cortex is close to laryngeal motor control, speech planning, and speech arrest regions of interest that were described after the 2011 DIVA update with high-resolution intracranial neuroimaging (Breshears et al. 2015; Silva et al. 2022; Zhao et al. 2023). Higher-order phonetic information enters ventral motor cortex through premotor cortex, a loosely defined anatomical region in the posterior inferior frontal gyrus and precentral gyrus (Figure 1.2 A,F,H). Here is where linguistic/phonetic information is converted into an articulatory motor program which tells primary motor cortex (i.e., the precentral gyrus; Figure

1.2 F,H) when and how to move each articulator. Stimulation to the premotor cortex results in *speech arrest*, a commonly observed phenomenon during neurosurgical mapping where continuous speech is halted without any evident motor interference (Zhao et al. 2023). This emphasizes the role of premotor cortex in DIVA as a phonetic-to-articulatory encoding region, as speech arrest can be conceptualized as the prevention of translation of phonetic-linguistic information into an articulatory motor program (Tate et al. 2014). Alternatively, speech arrest in premotor cortex might index speech inhibition rather than phonetic-to-articulatory encoding. Zhao et al. (2023) combined cortical stimulation with intracranial electrocorticography (ECoG) to show stimulation-induced speech arrest sites anatomically overlap with a speech inhibition task, which suggests ventral premotor cortex plays an inhibitory role in speech motor control. To relate this to a conceptual framework of speech motor control, corrective signals issued by the feedback control system are sent back to premotor cortex in the DIVA and HSFC models to adjust articulatory programs and re-update the efference copy for continuing error monitoring post-correction. An inhibitory speech-stopping mechanism localized to premotor cortex would thus assist in halting speech to issue corrections.

While the ventral precentral gyrus (Figure 1.2F) is the seat of articulatory-kinematic representations, the just-anterior inferior frontal gyrus (Figure 1.2A; canonically "Broca's area') is the seat of phonetic representations that serve as the input for the process of feedforward control. Flinker et al. (2015) supports this pre-articulatory role for Broca's area by showing that Broca's area is

silent during articulation, as at the time of articulation, the phonetic representations would have already been shipped off to motor cortex and transformed into an articulatory program. This distinctly pre-articulatory role for Broca's area is contrary to classic views of Broca's area as a speech production region (Geschwind 1970).

A more recent proposal outlines a functional region in the sensorimotor cortex described by Silva et al. (2022) as the middle precentral gyrus and by Hickok et al. (2023) as the "dorsal precentral speech area," putting it dorsal of the primary motor areas used in speech motor control posited by DIVA (Figure 1.2H). This region encompasses the dorsal laryngeal motor control areas documented in Dichter et al. (2018) and serves a myriad of roles in speech production. Interestingly, Silva et al. (2022) proposes a role for middle precentral gyrus in motor coordination, an important component of speech motor control. This is a notable departure from historical accounts of sensorimotor cortex as the proposed role of motor coordination is not a "primary" motor role, meaning implicating the middle precentral gyrus in motor coordination implies the region may perform more higher-order computations than previously believed. Indeed, this classification of middle precentral gyrus is more in line with how speech motor control literature discusses premotor cortex, which is anterior and ventral to this region (Figure 1.2I). Several case studies report that lesions to the neighboring area 55b (part of the middle frontal gyrus) result in apraxia of speech, a disorder of motor coordination, corroborating Silva et al. (2022)'s claims that the middle precentral gyrus is implicated in

motor coordination (Levy et al. 2023; Chang et al. 2020).

The cerebellum also plays a notable role in several aspects of speech motor control, which at first glance appears explainable by the structure's more general involvement in motor control (Manto et al. 2012). However, individuals with cerebellar degeneration directly associate the region with feedforward speech motor control, as these individuals exhibit less of an anticipatory response to feedback perturbation, followed post-perturbation by an increased corrective response Parrell et al. (2017, 2021). This demonstrates that individuals with cerebellar degeneration over-rely on feedback control due to an impaired feedforward control system.

1.1.2.2 Feedback control

While feedforward control is commonly prescribed to sensorimotor and inferior frontal cortex, feedback processing mechanisms are commonly mapped to the superior temporal gyrus (Figure 1.2E). The temporoparietal junction, or area Spt (Figure 1.2B), is also posited as a seat for auditory-motor integration during the feedback control process within the Dual Stream model (Hickok & Poeppel 2007; Hickok et al. 2009).

Error detection and correction, which involve comparing internal predictions about sensory feedback with ongoing sensory feedback, are localized to ventral sensorimotor cortex in DIVA (Figure 1.2F). In ventral motor cortex, perceived errors are converted into corrective motor commands, which are re-integrated into the feedforward controller. The feedforward control sys-

tem likely contains anatomically distinct error detection and correction loci, given errors can be corrected incredibly early in an utterance, before the auditory system would have access to auditory feedback (Gauvin & Hartsuiker 2020; Houde & Nagarajan 2011). Gauvin et al. (2016) used fMRI to localize feedforward error correction to the dorsal anterior cingulate. The authors characterize this region as a domain-general executive center.

1.2 Speech perception

Although speech motor control is a process that occurs during speech production, the focus of my dissertation primarily on feedback control necessitates a discussion of the mechanisms of speech perception and the auditory system. The neurobiology of speech perception is more well-understood than speech production as there are less methodological considerations for passive listening tasks than there are for speech production tasks: articulator movement artifact during speech production can render many types of neural signal too noisy to analyze (Vos et al. 2010; Shackman et al. 2009; Burgess 2020; Bush et al. 2022; Shuster 2003). While this may be construed as a limitation of speech production research, it is also an advantage of feedback control research: there is a large body of adjacent speech perception research to draw upon in a similar vein to how adjacent vision research informed contemporary theories of speech perception and production (Hickok & Poeppel 2007). For example, an abundance of speech perception research has revealed that the auditory system processes speech differently during feedback control when

compared to passive listening, which I will discuss in detail (§1.2.3). But first, I will provide an overview of how speech perception functions in the brain to give context for its modulation during speech production.

1.2.1 Organization of the auditory system

For the cerebral cortex, audition begins in the temporal lobe, which contains the primary auditory cortex (A1). A1 is located in Heschl's gyrus (HG) (Figure 1.2J) and planum temporale (PT) (Figure 1.2K) on the superior surface of the lobe and tucked into the Sylvian fissure (Figure 1.2L). A1 receives ascending auditory input from the subcortical auditory nuclei (Moore et al. 2010). However, the temporal lobe is not a monolithic primary auditory area, as it plays a role in many higher-order aspects of speech perception. This is evidenced in part by direct electrical stimulation of the temporal cortex during neurosurgery, which results in a multitude of perceptual effects, including auditory hallucinations when A1 is stimulated and scaling up the linguistic hierarchy into higher-order language comprehension errors when non-primary areas of the temporal lobe are stimulated (Hamberger 2007; Hamilton et al. 2021). Invasive electrophysiology (sEEG and ECoG) has also demonstrated that higher order auditory areas, primarily STG, are involved with abstraction of acoustic sensory information into higher-order linguistic features such as phonemes (Mesgarani et al. (2014); §1.2.2). Clinical observations that most cases of aphasia caused by temporal lobe injury have a primarily comprehension-based impairment profile (Manasco 2013) further corroborates

the multiple roles of the temporal lobe in speech perception.

The parietal lobe is, upon initial inspection, absent from the Classic Model, but is regardless critical for speech perception. Cortical stimulation accounts from several neurosurgeons report that the functional localization of Wernicke's area extends into the parietal lobe (Penfield & Roberts 1959; Lewandowsky 1912), although there is still a lack of consensus among cognitive linguists about the specific localization of Wernicke's area, with a survey presented in Tremblay & Dick (2016) reporting that only 26% of survey respondents endorsing a definition of the region that included portions of the parietal lobe. The parietal lobe plays a more prominent role in the Dual Stream model as it houses a region of the planum temporale near the temporo-parietal junction commonly referred to as "area Spt[2]," (Figure 1.2B) which serves as a sensorimotor interface for speech by transforming sensory input into motor commands for the articulatory system (Hickok & Poeppel 2007; Hickok et al. 2009). However, Spt is not thought to be a domain-specific speech or language region in the same way as the frontal and temporal regions of the language network: Spt has clear homologues in non-human primates (Cui & Andersen 2007), and is active during nonspeech vocalizations (Hickok et al. 2003). Most of the dorsal and posterior parietal lobe is not widely believed to play an active role in speech perception, instead supporting the process through more domain-general functions such as the frontoparietal attention network

[2]Area Spt is not formally localized to an anatomical structure and is instead functionally defined; however, most functional localizations place it around the temporo-parietal junction.

(Fedorenko et al. 2013; Germann & Petrides 2020; Meyyappan et al. 2021).

The involvement of the frontal lobe directly in speech perception (cf. indirectly through attentional mechanisms; §1.2.3.3) is currently a topic of debate, as research in the past ten years has almost conclusively determined that Broca's area as specified by the Classic Model is not a useful label for a locus of speech production (Flinker et al. 2015; Fedorenko & Blank 2020; Tremblay & Dick 2016). The role of the frontal lobe is further complicated by the motor theory of speech perception (Liberman et al. 1967), a historically influential theory of speech perception that is now widely discredited (Liberman et al. 1967; Massaro & Chen 2008). The motor theory states that frontal motor systems are recruited for perceiving speech through gesture perception. While the motor theory does have some contemporary supporters (Galantucci et al. 2006; D'Ausilio et al. 2009), there is an abundance of evidence in the contrary that uses modern neuroimaging to show that while the motor cortex does encode speech features during passive listening, these representations are based on auditory information and not perceived gestures, meaning the auditory system is still supraordinately important for speech perception over any motor system involvement (Arsenault & Buchsbaum 2016; Cheung et al. 2016; Du et al. 2014). The results I present in Chapter 3 include purely perceptual responses localized to the IFG (Figure 1.2A, G); however, these responses were confined to a single subject and could not be replicated elsewhere in my dataset (Figure B.2).

Areas of the brain beyond lateral cortex are also important for speech

perception. These are of particular interest for sEEG (Chapter 3), as many methods of neuroimaging have limitations that prevent high-fidelity recording of subcortical regions that sEEG bypasses by directly penetrating the cerebrum. For example, the thalamus is a subcortical structure that plays an important role as a general sensory relay to the cortex. For speech, this means the thalamus projects directly to primary auditory cortex (Dick et al. 2012). Perhaps more surprisingly, it is hypothesized that the thalamus also directly projects to speech perception areas in non-primary auditory cortex, suggesting the thalamus may be specialized for speech beyond its role as a more general sensory relay (Hamilton et al. 2021). Additionally, the amygdala forms a functional network with anterior insula (Figure 1.2M) that is active during processing of suprasegmental emotional components of speech (Zhang et al. 2018). The insular cortex, a lobe sequestered within the Sylvian fissure which divides the frontal and temporal lobes, is of particular interest to my dissertation results. The insula is historically underrepresented in neurobiological models of speech and language, in part due to how difficult it is to obtain *in vivo* recordings from the human insula: its placement underneath the pial cortical surface means that traditional ECoG grids are insufficient to reach it, making dissection of the Sylvian fissure necessary (Remedios et al. 2009). For the same reason, direct cortical stimulation of the insula is also rare (Zhang et al. 2018). The relatively recent advent of sEEG depth electrodes has made recording from the insula much easier, as sEEG can penetrate through pial cortex into the insula with minimal surgical preparation (Youngerman et al.

2019). The original sEEG research I present in this dissertation includes an unexpected finding in the insula (§3.5), which I personally attribute to the relative nascency of high-resolution insular recordings in humans. Among the large swath of brain regions I recorded from in my study, the posterior insula (Figure 1.2N) uniquely showed low-latency "onset" responses (§1.2.2.1) to speech production *and* speech perception stimuli. This insular response profile is most similar to primary auditory areas in temporal cortex that show low-latency responses to speech perception. The timing of these responses resembles primary and secondary auditory responses, leading me to theorize that the insula receives parallel thalamic sensory input to the primary auditory cortex. This means that the insula processes auditory stimuli in parallel to primary sensory areas during speech production, similar to the putative parallel thalamic input supplied to the posterior STG in Hamilton et al. (2021). If my theory is true, this could implicate the insula directly in speech motor control, as rapid attention to auditory feedback during speech production is a critical component of feedback control mechanisms present in models of speech motor control (Houde & Nagarajan 2011).

1.2.2 Linguistic abstraction

Linguistic abstraction is the process by which variable acoustic information is perceived as the invariable linguistic content. For example, if the same sentence is spoken by a frail grandparent with a delicate voice and a towering basketball player with a booming voice, our perceptual systems can

somehow extract the same communicative meaning from the variant frequency, amplitude, and timbre. In a sense, humans can understand the meaning of speech in a fashion seemingly separated from of the acoustic properties of the waveform. Most, if not all, conventional models of the neurobiology of language agree that linguistic abstraction takes place during speech perception as a fundamental precursor to language comprehension. Many models extend this claim by specifying the linguistic substrates used in abstraction of sensory information. This is inspired by a seminal work from the 1990s that set the stage for modern research into the neurobiology of language by introducing the "lack of invariance problem," which describes a gap in our understanding of cognitive neuroscience concerning the transformation of continuous, topologically organized perceptual stimuli (e.g., frequency) and discrete abstract (or "invariant") categories of representation (e.g., words; Appelbaum (1996)). The specifics of how the brain performs this task are of interest to my dissertation research.

Despite several robust hypotheses, there is no current consensus for how the brain performs the task of linguistic abstraction and the "lack of invariance problem" is very much a topic of debate in modern cognitive linguistics. A common theory still being explored today could provide the physiological reality that linguists seek for their theoretical constructs: linguistic units such as phonological features (Mesgarani et al. 2014), syllables (Sun & Poeppel 2023), and morphemes (Khanna et al. 2024) are theorized intermediate abstract representations of low-level acoustic stimuli. The hierarchical

organization and anatomical localization of these intermediate representations are an unresolved topic in the field. While the brain's intermediate representations have conventionally been conceptualized as a sequential process that transforms low-level sensory stimuli into serially more abstract representations, recent intracranial research suggests the process may be more parallel than previously suspected. Hamilton et al. (2021) used surgically implanted ECoG grids to look at how primary and non-primary auditory areas encode information during passive listening to sentences and showed that the flow of information between these areas was not sequential but instead parallel and/or reciprocal. That is, primary and non-primary areas exhibited intertangled response latencies and feature representations that did not support the conventional theory of step-by-step serial processing of information from primary to non-primary auditory cortex. The analysis methods present in this study are similar to what I utilize in Chapter 3, where I find a similar non-serial auditory processing in primary/non-primary auditory cortices as well as the posterior insula, replicating and expanding on this work to better characterize how the brain forms abstract representations of speech and language from continuous acoustic stimuli.

1.2.2.1 Onset and sustained responses

Phonemes in particular are a popular candidate for invariant representational unit during speech production and perception (Mesgarani et al. 2014; Cheung et al. 2016; Khanna et al. 2024), but their organization within auditory

cortex is modulated by suprasegmental characteristics of speech. In what is unquestionably the largest source of inspiration for my dissertation research, Hamilton et al. (2018) used an unsupervised clustering technique on ECoG responses during a passive listening task to identify two anatomically distinct response profiles in the auditory cortex. The first, labeled "onset" responses, were characterized by a high-amplitude, low-latency transient response at the beginning of the acoustic onset of a sentence or phrase (i.e., after $>$ 200 milliseconds of silence). "Sustained" responses, on the other hand, were relatively delayed when compared to onset responses and could happen throughout the timecourse of the sentence rather than exclusively at the onset. Notably, a breadth of phonological feature representations were observed separately in the onset and sustained regions, meaning that electrodes that prefer specific classes of phonological features were present in both regions. This suggests that onset and sustained response profiles are not subordinate to phonological feature representations but rather a supraordinate organizational feature of auditory cortex. Lastly, Hamilton et al. (2018) posited that onset responses in particular may serve as a temporal "reference frame" in continuous speech given that they only appear at boundaries, supported by the result that onset responses could be used in a decoding framework to predict sentence boundaries. Segmentation of speech is regarded as a fundamental characteristic of linguistic abstraction as one of the earliest transformations of acoustic information into invariant perceptual representations (Appelbaum 1996). The combination of onset and sustained responses' ability to modulate the phonological

feature tuning of an individual electrode and the notion that onset responses serves a role in temporal landmark detection during speech perception led me to wonder if these organizing features could differ during speech production, as robust feedforward expectations about upcoming utterance content (i.e., the efference copy) could negate the need for a segmentational cue and could explain the amplitude reduction observed in auditory responses to self-generated speech (§1.2.3.1).

1.2.3 Auditory feedback processing

The auditory system processes internally generated speech differently than externally generated speech through the direct activation of sensory systems by the motor system, a process known as *corollary discharge* (Schneider et al. 2014; Khalilian-Gourtani et al. 2022). Corollary discharge is the primary mechanism through which feedback control is performed and, in the case of speech production, often leads to a suppression of auditory cortical activity to internally generated speech (§1.2.3.1). This is not always the case, however: several studies have identified antithetical *enhancement* to internally generated speech in the auditory cortex, yet in these studies enhancement effects are usually a secondary result to a primary suppression effect (Greenlee et al. 2011; Chang et al. 2013). Auditory feedback processing can also be impaired in schizophrenia: a prevailing theory for the source of auditory hallucinations present in the disorder is an inability for the auditory system to distinguish internally generated and externally generated speech (Heinks-Maldonado et al.

2007; McGuire et al. 1995; Johns et al. 2001). Abnormal auditory feedback processing has also been associated with dyslexia (van den Bunt et al. 2017) and cerebellar degeneration (Parrell et al. 2017).

1.2.3.1 Speaker-induced suppression

Speaker-induced suppression (SIS) is the primary neural biomarker of auditory feedback processing (Niziolek et al. 2013; Houde & Chang 2015). SIS refers to a phenomenon where neural responses to self-generated speech are reduced in amplitude (suppressed) compared to the speech of others. Related phenomena regarding suppression of self-produced nonspeech sounds can be observed in animals (Schneider et al. 2014) and in humans triggering sounds via button press (Martikainen et al. 2005). Even though perception of internal state and, even more explicitly, suppression of self-generated sensation are common mechanisms in general models of motor control (Houde & Nagarajan 2011; Parrell et al. 2019), SIS is not considered a domain-general mechanism, nor does it reflect a general suppression of neural activity. Houde et al. (2002), one of the earliest studies demonstrating SIS in neural data, formulated the theory that SIS arises from a comparison of sensory feedback with the forward efference copy through a series of magnetoencephalography (MEG) experiments. This idea is well-supported in the literature and eventually made it into the DIVA model through the presence of inhibitory error maps as part of the feedback control system (Tourville & Guenther 2011).

At first it may seem unintuitive that an active monitoring process would

result in a *reduction* of neural response, but studies have shown that SIS is a trace of the feedback monitoring system by demonstrating that the amount of cortical suppression is reduced during a speech error (Ozker et al. 2022, 2024; Houde et al. 2002; Behroozmand & Larson 2011). Studies like these directly associates SIS with feedback comparison mechanisms and show that SIS can be modulated by the adherence (or deviation) of corollary discharge to the efference copy. This establishes SIS as a neural trace of the state estimation systems present in the HSFC and FACTS models (Houde & Nagarajan 2011; Parrell et al. 2019). Of particular importance to my dissertation is a section in the discussion of Niziolek et al. (2013), which writes about the neural representations present in sensory cortex during SIS[3]. To connect the dots, Niziolek et al. (2013) wrote that SIS is sensitive to subphonemic variation, or acoustic changes within a phonemic category (e.g., differences in voice onset time between two utterances of the sequence /ba/). In this study, the feedback monitoring system's response to individual productions of a given vowel depended on the acoustic proximity of an individual utterance to the average production of that vowel throughout the task. This demonstrates that the efference copy is not a literal series of articulatory commands sent to primary motor cortex; if that were the case, the subphonemic variations of individual trials would be present in the efference copy and the degree of mismatch between feedforward and feedback systems would never change for non-error trials.

[3]As a reminder, the nature of abstract representations during speech perception (and production) is a question of great interest to cognitive scientists (often called the "lack of invariance problem;" §1.2.2; Appelbaum (1996)).

So, because the mechanisms of the feedback control system are dependent on subphonemic variation despite that variation being absent from the efference copy, the feedforward system must contain some degree of abstraction away from the actual articulator kinematics. This is in line with the FACTS model's statements about how speech is a goal-oriented behavior: the efference copy represents a sensory goal, not an explicit sensory target. Niziolek et al. (2013) conclude this discussion section by speculating on the nature of abstraction present in the efference copy. Does abstraction take place when the efference copy is generated in premotor cortex, or are the rough edges of individual trial variation "smoothed off" via abstraction later on in sensory cortex (for example, a process akin to the phonological feature abstraction observed in Mesgarani et al. (2014))? The authors used a comparison to error literature to conclude that the efference copy itself is likely abstract before it reaches feedback control systems in auditory cortex. The decreased SIS observed in trials farther away from the average vowel production is reminiscent of the decreased SIS observed in speech errors (Houde et al. 2002; Behroozmand & Larson 2011), even though the "less good" trials were not consciously realized as errors by the speaker (i.e., there was no corrective behavior). This suggests the efference copy contains representational information as even though there was sensory mismatch in the feedback, there was no mismatch in the higher-order representational space (e.g., phonemes) used by efference copy, which would have triggered an error correction signal. The specific nature of the abstract representations in the efference copy posited by Niziolek et al.

(2013) are unspecified by the authors and a major source of inspiration for both studies presented in my dissertation (Chapters 2 and 3).

1.2.3.2 Interaction with onset responses

To recap, SIS is a neurophysiological trace of the error detection/correction component of the speech motor control network that engages in linguistic abstraction like the perceptual system, but with an unspecified mechanism. The less neural activity during SIS, the greater the adherence of the sensory feedback to the feedforward expectation. Interestingly, the N1 component of the neural response documented by Niziolek et al. (2013) and others as being suppressed during SIS bears a spectrotemporal similarity to onset responses (§1.2.2.1; Hamilton et al. (2018)) in terms of its high-amplitude and low-latency nature (Behroozmand & Larson 2011; Heinks-Maldonado et al. 2007; Martikainen et al. 2005). While I am not attempting to conflate onset responses with the N1, both have been theorized to index speech segmentation: onset responses in Hamilton et al. (2018) discussed above and the N1 in studies such as Sanders et al. (2002), where researchers identified N1 amplitude increased in word-initial syllables relative to word-medial syllables during a pseudoword learning task.

The goal of my dissertation is to study the physiology of onset responses during speech production in an attempt to link EEG and MEG literature on SIS to high-resolution intracranial electrophysiology studies on the mechanisms of auditory perception. Because onset responses can modulate phonological

feature representation, a follow-up goal of this aim is to examine if speaker-induced suppression influences phonological feature tuning as well.

1.2.3.3 Interaction with other cognitive systems

The top-down, goal-oriented nature of the speech motor control system demonstrated by studies like Niziolek et al. (2013) and models like Parrell et al. (2019) show that the speech motor control system can be modulated by top-down information[4]. In fact, both the speech production and the speech perception systems can be manipulated by top-down effects from other cognitive systems such as attention and predictive processing. For example, behavioral studies of altered auditory feedback in which self-generated speech is acoustically perturbed in real time through manipulation of pitch or other acoustic qualities of speech (e.g., timing) show that consistent perturbations of feedback may elicit larger corrective responses than inconsistent ones (Lester-Smith et al. 2020), suggesting that top-down anticipations of feedback coming from some domain-general predictive processing system (potentially a conflict monitoring network localized to dorsal anterior cingulate described by Gauvin et al. (2016)) may influence how the speech motor control system responds to feedback.

Studies that show an expectancy effect in EEG identify differences in later components such as the N400 and P600 (Goregliad Fjaellingsdal et al.

[4]DIVA, HSFC, and FACTS all localize top-down modulation on the speech motor control system as originating anatomically from frontal cortex (Tourville & Guenther 2011; Houde & Nagarajan 2011; Parrell et al. 2019).

2020) as well as earlier components such as mismatch negativity (Bendixen et al. 2014; Hawco et al. 2009; Näätänen et al. 2007) and the N1 (Astheimer & Sanders 2011), the latter of which has been posited as a neural biomarker of the efference copy (§1.1.1; Behroozmand & Larson (2011); Heinks-Maldonado et al. (2007); Martikainen et al. (2005)). The amplitude of the N1 is also modulated by SIS (Niziolek et al. 2013; Kurteff et al. 2023).

Attentional networks can also influence responses during perception tasks. In the EEG literature, amplitude of the N1 component is reduced during active listening when compared to passive listening (Brumberg & Pitt 2019; Houde et al. 2002). The susceptibility of the N1, a biomarker of speaker-induced suppression, to top-down modulations of other cognitive systems such as attention and expectancy suggests that these cognitive systems can modulate the speech motor control system to an extent. This theoretical link motivated the consistent/inconsistent feedback modulations present in the experimental designs of Chapters 2 and 3 of this dissertation (§2.4; 3.4).

1.3 Aims

The top-level objective of this dissertation is to investigate the neural mechanisms of feedback control using electrophysiology. Specifically, there is a gap in the literature between speech motor control and neurolinguistics. Speech motor control research is informed by engineering principles (Houde & Nagarajan 2011), while research on the brain's language network draws much from conventional theoretical linguistics (Appelbaum 1996; Mesgarani et al.

2014). Speech perception research has developed relatively more sophisticated techniques and theories due to its ease of study when compared to speech production and, therefore, speech motor control. This means that while we may have principles of how the auditory system operates during speech perception, how many of these principles remain in place during processing of our own auditory feedback during speech production, a process we know behaves differently than passive listening due to phenomena like speaker-induced suppression (§1.2.3.1), is unknown. The extent of *how* differently feedback control operates in comparison to passive speech perception is what I aim to investigate in this dissertation.

In Chapter 2, I present the results of a noninvasive scalp EEG study where participants speak aloud then passively listen to playback of their own voice. The speaking condition indexes ongoing feedback control while the passive listening condition indexes speech perception in general. Any difference in neural response between the two can then be attributed to the mechanisms of the feedback control system. In Chapter 3, I present the results of the same speaking-playback paradigm but using intracranial recordings from epilepsy patients. While this method yields smaller sample sizes, it affords a much higher signal-to-noise ratio and allows investigation of the role of specific brain structures in feedback control, something impractical in EEG given the coarse spatial resolution. In both results chapters, my analysis consists of a mixture of conventional event-related potential analysis (Luck (2014); §2.4; §3.4) and modern linear modeling techniques that investigate the relationship between

stimulus characteristics and the neural response (Di Liberto et al. 2015; Theunissen et al. 2000). The goal of the latter is to investigate how phonological feature tuning, previously observed in auditory (Mesgarani et al. 2014) and motor (Cheung et al. 2016) areas during passive listening, are affected during the cortical suppression of auditory feedback during speech production. I hypothesize that higher-order phonological feature representations are unaffected by feedback control of self-generated speech.

What I *do* hypothesize to be affected during feedback control are onset responses (§1.2.2.1). Onset responses are a response profile identified in Hamilton et al. (2018) using sentence perception stimuli. An onset response is defined as a transient high-amplitude spike of neural activity at the acoustic edge of a stimulus; in this study and in my dissertation, that means the beginning of a sentence. The reason I hypothesize onset responses to be affected during self-generated speech is due to a theorized role for onset responses in segmentation of continuous acoustic stimuli into discrete abstract representational units (such as phonemes), a fundamental component of speech perception (§1.2.2; Appelbaum (1996)). If onset responses are present in perception to identify the representational "edges" of a stimulus, the presence of a feedforward expectation about the sensory content of self-generated speech may nullify the need for onset responses during speech production. Such a result would also link onset responses to speaker-induced suppression, as the absence of onset responses would explain the amplitude reduction observed in auditory processing of one's own speech.

1.3.1 Clinical populations

I wanted to dedicate a section at the end of Chapter 1 to discuss the most common speech and language pathologies applicable to the study of the neurobiology of speech motor control. People with disordered speech and language are the ones who serve to benefit most from basic science research of the neurobiological mechanisms behind speech and language. Reciprocally, without consenting participants from these clinical populations, research like mine would not be possible. Scientists like myself have a commitment to disseminate results in a way that can inform development of better diagnostic and therapeutic interventions for people with disordered communication.

The bulk of neurogenic communication disorders by volume consists of aphasia and apraxia of speech caused by middle cerebral artery stroke, the main supply of blood to most of the lateral surface of the cerebral cortex (Manasco 2013; Caviness et al. 2002). The most common aphasia classification system, the Boston classification system (Goodglass & Kaplan 1972), divides aphasia into categories determined by the presence or absence of deficits in three domains: fluency, speech comprehension, and word repetition. Of particular importance to my dissertation research are the transcortical sensory/motor aphasias, as these aphasias involve interactions between the perception and production systems that are not well understood (Boatman et al. 2000; Ardila 2010). A more fine-grained understanding of how specific brain regions process auditory feedback during speech using intracranial recordings (i.e., Chapter 3) has the potential to lead to better diagnosis of these rarer

aphasia types. Conduction aphasia, described by the Boston classification system as a selective deficit in word repetition, has impairment of auditory-motor integration as a core deficit and thus also serves to benefit from my research (Buchsbaum et al. 2011).

A concomitant pathology to aphasia that is much more directly relevant to the study of speech motor control is acquired apraxia of speech (AOS). AOS is a motor speech disorder that involves difficulty in motor speech coordination and is thus fundamentally linked to feedforward control deficits (Duffy 2019). Notably, AOS does not involve muscle weakness like its sister disorder, dysarthria, meaning AOS is purely conceptualized as a coordination and planning deficit. In foundational aphasiology research, AOS is considered a breakdown between phonological planning and motor execution (Darley et al. 1975), but despite this, expressive ("Broca's") aphasia and AOS very commonly co-occur to an extent that many clinicians experience difficulty distinguishing between the two disorders (Patidar et al. 2013; Kobayashi & Ugawa 2013). Furthermore, AOS is much more rarely diagnosed than the expressive aphasias (even moreso in isolation), limiting the number of studies on the disorder. Many studies on the neurobiology of AOS are case studies as a product of this limitation (Patidar et al. 2013; Chang et al. 2020; Levy et al. 2023). Because of the difficulties surrounding extensively researching AOS due to its rare and often misdiagnosed nature, basic research on speech motor control, such as the results I present in my dissertation, is important foundational research that can lead to better diagnosis of apraxia of speech.

Stuttering, while more commonly a developmental pathology than a neurogenic one, is a fluency disorder that likely involves a malfunctioning speech motor control system. Early neuroimaging research on stuttering theorized that people who stutter have an impairment with feedback control processing, as evidenced by decreased activation of auditory areas during speech production (Fox et al. 1996). Delayed auditory feedback, which does not improve fluency in AOS (Jacks & Haley 2015), has a fluency-increasing effect on people who stutter, which has led to the use of delayed auditory feedback as a therapeutic technique in people who stutter[5] (Kalinowski & Stuart 1996). The feedback control hypothesis of stuttering is strengthened by studies of speaker-induced suppression showing relatively less cortical suppression to self-generated speech in people who stutter as compared to healthy controls (Toyomura et al. 2020). It is possible that the feedforward control system is also impaired in people who stutter: Max & Daliri (2019) looked at another neurophysiological correlate of speech motor control, pre-speech auditory modulation, and found that people who stutter have reduced pre-articulatory cortical evoked potential compared to healthy controls. The authors reconcile this observation with feedback control theories of stuttering by suggesting that reduced pre-speech auditory modulation may reflect the feedforward controller's failure to properly prepare feedback control systems for upcoming auditory feedback. Following this framework, fluency errors caused by stuttering are

[5]Although, the use of delayed auditory feedback as a therapeutic intervention in stuttering is a topic of controversy among clinicians. See Laiho et al. (2022) for a review.

the result of an over-corrective feedback control system trying to erroneously correct ongoing articulation.

The pathologies of aphasia, AOS, and stuttering show that the assessment and treatment of communication disorders that involve the speech motor control system have a lot to gain from a better understanding of the neural mechanisms underlying the process. The research I present in Chapters 2 and 3 specifically investigates how auditory feedback processing is modulated during speech, an important mechanism for the feedback control of speech, as evidenced by its dysfunction in stuttering. A concrete example of how my dissertation could benefit these clinical populations is via informing the surgical placement of a hypothetical neural prosthetic that would modulate the feedback control system in moments of malfunction through electrical stimulation, similar to the deep brain stimulators already in use in clinical populations such as Parkinson's disease (Benabid 2003).

Medically refractory epilepsy, meaning epilepsy that is deemed unmanageable using medication alone (Englot & Chang 2014), is not primarily a communication disorder but is an important clinical population for my dissertation research. Most sEEG data used in research, including the data I present, are recorded from epilepsy patients undergoing monitoring procedures for surgical intervention (Hamberger 2007). sEEG data is a rare and valuable resource for neuroscientific research due to its high spatial and temporal resolution and ability to record from subcortical structures (Chang 2015). Much is owed by researchers such as myself to the goodwill of consenting epilepsy pa-

tients. Although severe epilepsy cases have the potential to affect the structure and function of the human brain (Möddel et al. 2009), and thus can impede the generalization of many findings using this research paradigm, the sparsity of intracranial recordings of the human brain with this caliber of spatiotemporal resolution means it is an unavoidable limitation. As such, noninvasive recordings of healthy human brains form an essential complement of intracranial research, as they can corroborate findings without the potential influences of epilepsy on the brain. This is why I present data from both invasive and noninvasive recordings in this dissertation.

Chapter 2

EEG Results: Speaker-induced suppression during a naturalistic reading and listening task

2.1 Preface

This chapter of my dissertation contains original research designed, collected, analyzed, interpreted, and written by me. The entirety of this chapter was adapted from either my Masters thesis (Kurteff 2020) or from a peer-reviewed publication in *Journal of Cognitive Neuroscience* detailing the same experiment (Kurteff et al. 2023). This work formed the bulk of research conducted in my first several years in my combined Masters-PhD program and underwent several major revisions between its initial presentation in my Masters thesis and its eventual publication. This chapter will most closely resemble the version published in *JoCN*, but with the Introduction and Discussions abridged to minimize redundancy with the Introduction and Discussion sections of this dissertation. While my Masters thesis contained a large focus on removal of electromyographic artifact from EEG activity, this dissertation chapter (and the corresponding *JoCN* article) focus primarily on the experimental results of the task. Information about artifact correction is available in Appendix A.

Data presented in this chapter are publicly available for those wishing to replicate my results or conduct original analyses. The data are hosted as an OSF repository, while the code to replicate the analyses is available as a GitHub repository.

I am grateful to my co-authors on the *JoCN* manuscript for their assistance in all steps of preparing this chapter. My (at-the-time) undergraduate research volunteers: Amanda Martinez, Nicole Currens, Jade Holder, Cassandra Villarreal, Valerie R. Mercado, Christopher Truong, Claire Huber, and Paranjaya Pokharel; without them, the arduous task of transcribing and phoneme-aligning this dataset wouldn't have been possible. I would also like to thank lab members Maansi Desai and Ian Griffith for their feedback during all stages of this research as well as undergraduate mentee Tasha Anslyn for collaborating with me on a supplemental analysis of speech errors that did not make it into the final manuscript. PhD committee member and co-author Rosemary A. Lester-Smith provided invaluable feedback during experimental design and the peer review process. Fellow PhD committee member Stephanie Ries, while not an author on this manuscript, provided helpful commentary during data collection that steered the early direction of my analysis. Lastly, my PhD supervisor Liberty S. Hamilton made this project possible, providing much-needed assistance throughout every step of the research process.

2.2 Abstract

Speaking elicits a suppressed neural response when compared with listening to others' speech, a phenomenon known as speaker-induced suppression (SIS). Previous research has focused on investigating SIS at constrained levels of linguistic representation, such as the individual phoneme and word level. Here, I present scalp EEG data from a dual speech perception and production task where participants read sentences aloud then listened to playback of themselves reading those sentences. Playback was separated into immediate repetition of the previous trial and randomized repetition of a former trial to investigate if forward modeling of responses during passive listening suppresses the neural response. Concurrent electromyography (EMG) was recorded to control for movement artifact during speech production. In line with previous research, event-related potential (ERP) analyses at the sentence level demonstrated suppression of early auditory components of the EEG for production compared with perception. To evaluate whether linguistic abstractions (in the form of phonological feature tuning) are suppressed during speech production alongside lower-level acoustic information, I fit linear encoding models that predicted scalp EEG based on phonological features, EMG activity, and task condition. I found that phonological features were encoded similarly between production and perception. However, this similarity was only observed when controlling for movement by using the EMG response as an additional regressor. My results suggest that SIS operates at a sensory representational level and is dissociated from higher order cognitive and linguistic processing

that takes place during speech perception and production. I also detail some important considerations when analyzing EEG during continuous speech production.

2.3 Introduction

2.3.1 Speaker-induced suppression and speech motor control

Speech production and speech perception are frequently studied separately in research, yet the two processes have a robust, interactive theoretical link (§1.1; Skipper et al. (2017); Houde & Nagarajan (2011); Tourville & Guenther (2011); Zheng et al. (2010); Watkins et al. (2003)). Models of the neurobiology of speech production universally include the sensorimotor control of speech, a mechanism by which speakers can detect errors via auditory and somatosensory feedback and subsequently correct those errors (§1.1.1; Parrell et al. (2019); Houde & Chang (2015); Tourville & Guenther (2011); Perkell et al. (1997)). For the feedback control component of speech motor control, speaker-induced suppression (SIS) is among the most-well documented neural biomarkers, in which neural responses to (errorless) self-generated sounds are suppressed in relation to externally generated sounds (§1.2.3.1; Brumberg & Pitt (2019); Niziolek et al. (2013); Martikainen et al. (2005); Houde et al. (2002)).

2.3.2 Linguistic abstraction

While SIS and other aspects of speech motor control such as the efference copy have previously identified robust biomarkers (§1.1.1; Behroozmand & Larson (2011); Heinks-Maldonado et al. (2007); Martikainen et al. (2005)), the interaction of SIS with other cognitive and linguistic processes ongoing during speech perception and production is not well studied. It is widely accepted that the brain uses some sort of intermediate representations when processing language from its constituent acoustic signal (§1.2.2; Mesgarani et al. (2014); Appelbaum (1996)), and specific representations of the perceptual response may be deemed unnecessary during production (e.g., phonological features, acoustic properties of the speech signal) because of more complete information about auditory stimuli during speech production. Previous work has investigated linguistic feature representation during speech production (Cheung et al. 2016), but this work did not establish how linguistic feature representations *differ* during perception and production, meaning the nature of feature representation preservation during SIS is still an open question. In addition, Cheung et al. (2016) mainly addressed changes to feature tuning in the motor cortex itself, rather than changes to tuning in sensory speech areas of the auditory cortex. Research by another group has speculated some form of invariance is utilized in the efference copy, which is itself a feedforward cortical signal that is suppressed during speech production as part of SIS (Niziolek et al. 2013). Niziolek et al. (2013) did not identify an explicit source/nature of the speculated invariant representations present in this signal, but this pa-

per's observation that the efference copy does not predict all variability in the sensory feedback of speech lends itself to two hypotheses: either the efference copy itself represents an invariant motor plan (cf. the explicit outgoing motor commands of speech), or it represents precise encoding of motor commands that lose their precision (i.e., become invariant) in sensory cortex. The association between SIS's sensitivity to subphonemic variation and the invariant encoding of phonological features in sensory cortex is unclear. An approach that could successfully demonstrate differential encoding of phonological features between speaking and listening could help establish the proposed form of invariance in feedforward control of speech.

2.3.3 Forward expectations during speech perception and production

A separate question concerns how expectations about utterance content may influence feedforward speech motor control. While speech perception does involve feedforward processing (Poeppel & Monahan 2011), forward models present during production are much more complete due to expectations about utterance content being internally generated during utterance planning. Therefore, the presence of a robust forward model during speech production represents a fundamental difference between conditions where speech is suppressed and where speech is not suppressed. Potentially, the feedback monitoring mechanisms at work during SIS could serve a role in general predictive processing; alternatively, the mechanisms of SIS could be specific to speech motor control and any predictability-based modulations could reflect domain-

general prediction mechanisms exerting a top-down influence on SIS, which is primarily theorized as a bottom-up sensory process comparing a forward model of motoric goals with auditory/sensorimotor feedback (Niziolek et al. 2013).

2.3.4 The importance of naturalistic stimuli in EEG speech production experiments

The last question of interest for this study is whether or not SIS, a neurophysiological phenomenon previously studied in highly constrained "laboratory speech" environments, is observable in more generalized "naturalistic speech." Shifting EEG studies toward language as it occurs in natural settings compared with the heavily constrained single word or syllable-level studies of the past facilitates generalization to clinical applications and reinforces the interdisciplinary drive to use more ecologically valid stimuli in studies of the neural representation of speech and language (Hamilton & Huth 2020). Studies that expand beyond using evoked stimuli and incorporate naturalistic stimuli (e.g., sentences) raise the ecological validity of the research while also providing a window of analysis for the feedforward and feedback processes that link perception and production (Kearney & Guenther 2019; Houde & Nagarajan 2011; Poeppel & Monahan 2011; Casserly & Pisoni 2010).

2.3.5 Aims

In this study, I aim to investigate differences in EEG responses between sentence-level speech perception and production, as well as speech perception

in consistent and inconsistent contexts to define the neural representations underpinning SIS more precisely. Deviance from a motoric goal has been previously demonstrated to modulate SIS, suggesting the process takes place at a sensory representational level. However, it is unclear whether the suppression of sensory representations during speech production affects linguistic abstractions that are generated from sensory processing. If linguistic representations were suppressed in conjunction with SIS, I would expect to see differential phonological tuning to specific speech features (Desai et al. 2021; Khalighinejad et al. 2017; Di Liberto et al. 2015). In addition, I opted to structure my task such that participants can anticipate when playback of auditory responses is inconsistent with the preceding speaking trial. This allowed me to determine if there is a link between predictive processing in passive listening and the consistency or inconsistency of feedback (Lester-Smith et al. 2020). To investigate these questions, I designed an experiment that used identical acoustic stimuli in separate speech perception and production conditions then compared the difference in event-related potentials (ERPs) as well as in tuning of phonological features across conditions. I hypothesized that, although speech production will be suppressed relative to perception in this study, phonological feature tuning would remain stable between modalities of speech. In addition, I expected a similar trend in inconsistent perceptual stimuli, such that phonological feature representations will remain stable but reduced in amplitude in comparison to consistent perceptual stimuli. If responses to consistent versus inconsistent playback show a similar pattern of

suppression to that observed during speaking versus listening, we may conclude that these processes involve similar underlying computations. On the other hand, if we see differences in these patterns, this suggests that SIS is computationally distinct from other phenomena that involve forward modeling or expectations of upcoming speech. In addition, observing differences in linguistic abstraction of acoustics into phonological features can help contextualize the phenomenon in relation to other cognitive and linguistic processes operating during speech perception and production.

2.4 Methods

2.4.1 Subject details

Twenty-one participants (11 women, age 24.4±3.9) were recruited from The University of Texas at Austin. This is in line with sample sizes of recent EEG studies of speech production (Ries et al. 2021; Goregliad Fjaellingsdal et al. 2020; Zhao & Rudzicz 2015). All participants were native speakers of English with typical hearing as assessed through pure-tone audiometry and a speech-in-noise hearing test (QuickSIN, Interacoustics). Participants provided written consent for participation in the study and were compensated at a rate of $15/hr with an average session length of 2 hr (1 hr for setup, 1 hr for recording EEG). One participant was excluded because of a recording error, leaving 20 participants in the final analysis. All experimental procedures were approved by the institutional review board at The University of Texas at Austin.

2.4.2 Perception-production task

The task was designed using a dual perception-production block paradigm, where trials consisted of a dyad of sentence production followed by sentence perception. In each trial, participants overtly read a sentence and then listened to a recording of themselves reading the produced sentence. Perception trials were divided into blocks of consistent and inconsistent stimuli. *Consistent* stimuli consisted of immediate playback of the production trial, whereas *inconsistent* stimuli consisted of a randomly selected production trial from the previous block. Consistent and inconsistent playback trials were presented in a block paradigm to avoid eliciting an "oddball" response, a commonly observed ERP component that elicits a response to randomly deviant perceptual stimuli (Barry et al. 2000). A schematic is provided in Figure 2.1. The generation of perception trials from the production aspect of the task allowed stimulus acoustics to be functionally identical across conditions.

Sentences were taken from the MultiCHannel Articulatory (MOCHA) database, a corpus of 460 sentences that include a wide distribution of phonemes and phonological processes typically found in spoken English (Wrench 1999). These sentences have been used previously in intracranial studies of speech production (Chartier et al. 2018). A subset of 50 sentences (100 for the first two participants) from MOCHA were chosen at random for the stimuli in the present study; however, before random selection, I manually removed 61 sentences for either containing offensive semantic content or being difficult for an average reader to produce to reduce extraneous cognitive effects and error pro-

duction, respectively. I changed the sentence set from 100 to 50 sentences after the first two participants because of concerns about participant fatigue during the task. Participants completed six blocks of the task for 300 perception and 300 production trials per participant (400 for the first two participants). Sentences had a median length of 2.9 sec. A broadband click tone was played in between trials as an additional cue to assess the effect of EMG correction on low-level auditory responses (see Appendix A).

Stimuli were presented in a dimly lit sound-attenuated booth on an Apple iPad Air 2 using custom interactive software developed in Swift (Apple XCode Version 9.4.1). Auditory stimuli were presented at a comfortable listening level via foam-tipped insert earbuds (3M, E-A-Rtone Gold 10Ω). Visual stimuli were presented in a white font on a black background after a 1000 millisecond fixation cross to minimize visual artifact in the EEG signal (Figure 2.1). Accurate stimulus presentation timing was controlled by synchronizing events to the refresh rate of the screen. The iPad was placed on a table over the participants' lap so they could advance trials during the task with minimal arm movement. Participants were instructed to complete the task at a comfortable pace and were familiarized with the task before recording began. Trial information, including onset and offset of each trial, transcriptions of produced and heard sentences, trial type, trial number, and block number were collected by an automatically generated log file to assist in data processing.

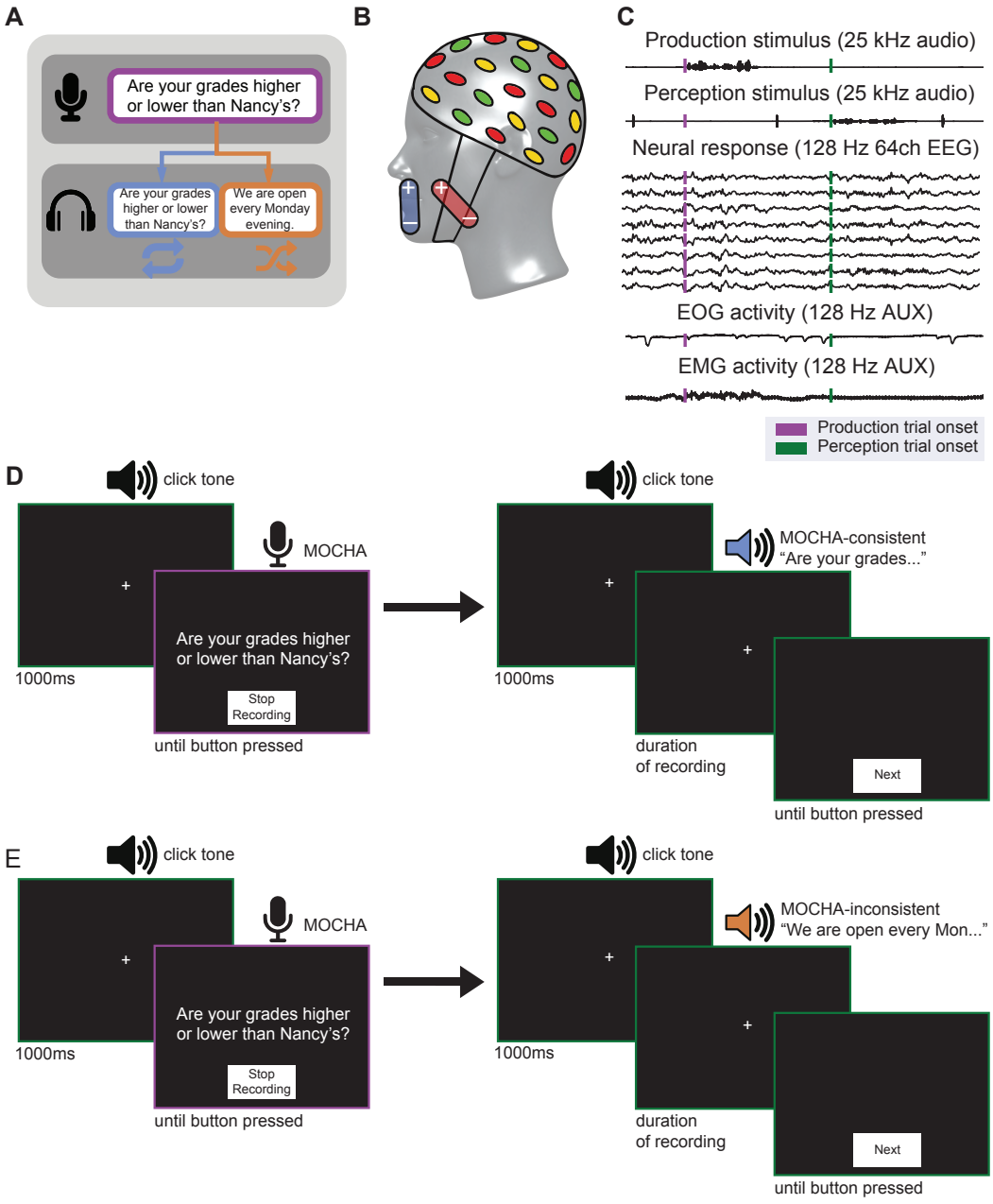
A
Are your grades higher or lower than Nancy's?
Are your grades higher or lower than Nancy's?
We are open every Monday evening.
B
C
Production stimulus (25 kHz audio)
Perception stimulus (25 kHz audio)
Neural response (128 Hz 64ch EEG)
EOG activity (128 Hz AUX)
EMG activity (128 Hz AUX)
Production trial onset
Perception trial onset
D
click tone
MOCHA
Are your grades higher or lower than Nancy's?
Stop Recording
1000ms
until button pressed
click tone
MOCHA-consistent
"Are your grades..."
1000ms
duration of recording
Next
until button pressed
E
click tone
MOCHA
Are your grades higher or lower than Nancy's?
Stop Recording
1000ms
until button pressed
click tone
MOCHA-inconsistent
"We are open every Mon..."
1000ms
duration of recording
Next
until button pressed

Figure 2.1: **Dual perception-production task and EEG data collection schematic**.
(A) Schematic of trial types in the task. The participant first reads a sentence aloud (purple) then hears playback of the same audio (yellow, consistent playback condition) or audio from a different random trial (light blue, inconsistent playback).
(B) Schematic of auxiliary EMG electrode placement on orbicularis oris (blue) and masseter (red).
(C) Visualization of all signals recorded during task, including produced audio (speech), perceived audio (clicks and speech), and EOG and EMG channels. Only eight EEG channels are visualized here, but 64 were recorded and used in analysis. Vertical lines denote the onset of a production (purple) or perception (green) trial (i.e., the acoustic onset of the first phoneme of the sentence). Blinks are observed as deflections in the EOG channel; muscle activation during production is notable as high activity in the EMG channel.
(D, E) Outline of trial procedure for consistent (yellow) and inconsistent (light blue) blocks.

2.4.3 EEG and EMG acquisition

Sixty-four-channel scalp EEG and audio were recorded continuously via BrainVision actiChamp amplifier (Brain Products) with active electrodes at 25 kHz. A high sampling rate was used to synchronize task audio and EEG, which were recorded using the same amplifier. Conductive gel (SuperVisc, EASYCAP) was applied to the scalp at each electrode, and impedance at each electrode was kept below 15 kΩ throughout the recording. Audio signals from both the insert earphones (presented audio, 3M E-A-Rtone Gold 10Ω earphones) and microphone (produced audio, Audio Technica U853rw cardiod condenser microphone) were captured as additional EEG auxiliary channels (also at 25 kHz) and were aligned with neural data via a StimTrak processor

(Brain Products). Vertical electrooculography (vEOG) was captured via auxiliary electrodes above and below the left eye in line with the pupil. Auxiliary electrodes were also used to capture facial EMG activity (Figure 2.1); these electrodes were placed on the orbicularis oris and mandible in the majority of participants ($n = 11$), but on other muscles important to articulation (masseter ($n = 6$), submental triangle ($n = 2$)) in several participants (Stepp 2012; Van Eijden et al. 1993; Rastatter & De Jarnette 1984). Multiple placements were utilized because of issues with electrode adherence caused by participant facial hair. All placements were trialed on a participant who consented to additional time during setup. A reference electrode for all auxiliary electrodes was placed on the left earlobe. Auxiliary EMG placement was not required for preprocessing but provided validation that EMG artifact was removed during preprocessing. EMG activity associated with the onset of articulation, which caused the largest artifacts in the temporal window of interest for ERP analysis of speech production, was automatically detected and epoched from auxiliary EMG channel activity. All recorded signals timed according to stimulus onsets are visualized in Figure 2.1. The first two participants did not have auxiliary electrode placement because of unavailability of recording hardware, so EMG activity was corrected based on EEG channels only.

2.4.4 Data preprocessing

All EEG processing was performed offline using custom Python scripts and functions from the `mne` software package (Gramfort et al. 2014). EEG,

EOG, and EMG data were downsampled from 25 kHz to 128 Hz before analysis. EEG data were referenced to the linked mastoid electrodes (the average of the TP9 and TP10 channels) and notch filtered at 60 Hz to remove line noise. For one participant (OP17), one reference electrode was a bad channel and was interpolated before re-referencing. The data were next filtered from 1–30 Hz (Hamming window, 0.0194 passband ripple with 53-dB stopband attenuation, 6 dB/octave). Bad channels and segments were manually rejected, then Independent Component Analysis (ICA) was performed to correct for EOG and electrocardiographic artifact with the number of components equal to the number of good channels. ICA components related to vEOG, hEOG, and electrocardiographic artifact were manually identified and removed via scalp topography and epoching component activity to vEOG activity (obtained via `mne.create_eog_epochs()`). The selected ICA components were next removed from the unfiltered data. After ICA, data were filtered at 0.16 Hz and corrected for EMG artifact via blind source separation algorithm based on canonical correlation analysis (CCA; De Clercq et al. (2006)), a technique that has been previously demonstrated to correct for EMG artifact in speech production EEG tasks (Ries et al. 2021; Riès et al. 2013; Vos et al. 2010). In line with these studies, CCA was performed 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. CCA was performed using the Automatic Artifact Removal plugin for `eeglab` (Gómez-Herrero 2007). Validity of CCA artifact correction for the removal of EMG

from continuous speech production data is not discussed further in this chapter; however, an additional verification of the technique can be found in Appendix A. After CCA and before analysis, bad channels were interpolated and data were bandpass filtered between 1 and 30 Hz.

2.4.5 Event-related potential (ERP) analysis

Accurate timing information for words, phonemes, and sentences was generated to allow epoching of EEG data to multiple levels of linguistic representation. Log files generated by the task application were used to identify the timing of individual sentences in the task, which were then made temporally precise using a modified version of the Penn Phonetics Forced Aligner (Yuan & Liberman 2008), which automatically generated Praat TextGrids (Boersma 2002). Automatically generated TextGrids were checked for accuracy at the sentence, word, and phoneme level by the undergraduate research volunteers who are authors on the *JoCN* manuscript that this chapter is adapted from (A.M., N.C., J.H., C.V., V.M., C.T., C.H., and P.P). I supervised the transcription process and checked the final TextGrids for accuracy before generating event files used in the analyses. Event files containing start and stop times for each phoneme, word, and sentence, as well as information about trial type (perception vs. production; consistent vs. inconsistent playback), were created using the log files and TextGrids. A second set of event files corresponding to the intertrial click sound were generated via a match filter process where the audio signal of the click was convolved with the EEG audio signal to find exact

timing matches (Turin 1960). To examine the differences between perception and production at the sentence level, sentence-level event files were used to epoch the neural response between −1.5 seconds and +3 seconds relative to sentence onset, which I quantified as the acoustic onset of the first phoneme of the sentence (Ozker et al. 2022). Epochs ±10 *SD*s from the within-subject mean were rejected.

2.4.6 Linear mixed-effects (LME) modeling

Linear mixed-effects (LME) models were created and assessed using the `lmertest` package (Kuznetsova et al. 2017) in R to determine statistical differences between different task conditions within relevant time windows, specifically the N1 (80–150 msec) and P2 (150–250 msec). The peak amplitudes and latencies of these windows, as well as the peak-to-peak amplitude of the N1 and P2 components, were used as response variables. Latency was calculated as the time at which the largest peak within a time window of interest occurred. LME models were specified using Equation 2.1:

$$y = X\beta + Zu + \epsilon \tag{2.1}$$

where β represents fixed-effects parameters, u represents random effects, and ϵ represents residual error. X and Z are matrices of shape $(n * p)$, where n is the number of observations of each parameter and p is the value of parameter at observation n. In all models, the fixed effect was the response variable of interest (i.e., N1 & P2 amplitude & latency; peak-to-peak amplitude) and subject

was used as a random effect. F tests were calculated using Kenward–Roger approximation with n degrees of freedom specified (Kenward & Roger 1997).

2.4.7 Multivariate temporal receptive field (mTRF) modeling

Linear encoding models (also referred to as spectrotemporal or multivariate temporal receptive field models in previous literature) were fit to describe the selectivity of the EEG responses to phonological features corresponding to place and manner of articulation (Desai et al. 2021; Hamilton et al. 2018; Crosse et al. 2016; Di Liberto et al. 2015; Mesgarani et al. 2014). This model takes the form of Equation 2.2:

$$\hat{y}_n(t) = \sum_{f} \sum_{\tau=-0.3}^{\tau=0.5} w(f,\tau) S(f, t-\tau) + \epsilon \tag{2.2}$$

where $\hat{y}_n(t)$ represents the estimated EEG signal for electrode n at time t. The stimulus matrix S consists of behavioral information regarding features (f) for each time point $t-\tau$, where τ is the time delay between the stimulus and neural activity in seconds. Features included combinations of binary features for perception, production, consistent playback, and inconsistent playback trials, as well as continuous, normalized EMG activity recorded from auxiliary electrodes, and binary features for the presence of phonological features at each time point (as in Desai et al. (2021); Hamilton et al. (2018); Mesgarani et al. (2014)). The "full" model stimulus matrix contained 14 phonological features as well as four binary features encoding trial information (perception, production, consistent playback, inconsistent playback) and normalized EMG

activity from facial electrodes for 19 features. These phonological features for place and manner of articulation were identical to those used in previous work (Desai et al. 2021; Hamilton et al. 2021; Mesgarani et al. 2014) and included sonorant, obstruent, voiced, nasal, syllabic, fricative, plosive, back, low, front, high, labial, coronal, and dorsal. Phonemes were coded in a binary matrix where a 1 indicated the onset of a phoneme's articulation via timing information obtained from the TextGrids. I fit separate models to predict the EEG response in each channel using time delays of -0.3 seconds to $+0.5$ seconds, relative to the acoustic onset of the phoneme. This delay range encompassed the temporal integration times to similar responses found in previous research (Hamilton et al. 2018) but with an added negative delay to encompass potential pre-articulatory neural activity (Chartier et al. 2018). Data were split 80-20 into training and validation sets. To avoid overfitting, the data were segmented along sentence boundaries, such that the training and validation sets would not contain information from the same sentence. These segments were then randomly combined into the 80/20 training/validation sets. Weights for each feature and time delay $w(f, \tau)$ were fit using ridge regression on the training set and a regularization parameter chosen by 10 bootstrap iterations, fitting on subsets of the training set. The ridge parameter was selected at the value that provided the highest average correlation performance across all bootstraps. Ridge parameters between 10^{-5} and 10^{5} were tested in 20 logarithmically scaled intervals. Model performance was assessed using correlations between the EEG response predicted by the model and the

true EEG response. Significance of these correlations was obtained through a bootstrap procedure with 100 iterations in which the training data were shuffled in chunks to remove the relationship between the stimulus and response but preserve temporal correlations within the EEG signal. Visual inspection of the data revealed two participants (OP4 and OP17) for whom responses showed no discernible receptive field structure even after greatly expanding the range of ridge parameters, motivating their exclusion from the analysis. To investigate the relationship between encoding of phonological features during perception and production, a "task-specific" model was fit (Figures 2.3; 2.5), which contained three sets of phonological features: those that occurred exclusively during production trials, those that occurred exclusively during perception trials, and a combined perception-plus-production set of phonological features identical to those included in the "full" model described above. After fitting this model, I calculated a feature-by-feature correlation for the production-specific and perception-specific feature weights (e.g., correlation of fricative-production with fricative-perception) to investigate how representations of phonological features change between modes of speech (Figure 2.5). I also used the production-specific and perception-specific model weights to fit separate predictions of the held-out validation set EEG activity, which were then averaged relative to sentence onset to facilitate comparison to the ERP analysis (Figure 2.5).

2.5 Results

Topographic inspection of sentence-level ERP activity revealed a frontocentral ROI of nine channels that elicited the strongest response to sentence onset during speech perception and production (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, and C2). This ROI is used in the ERP results, but linear encoding models were fit on all channels for all participants.

2.5.1 Speaker-induced suppression observed at the sentence level

After verifying the integrity of the dataset, I wished to understand whether and how responses to continuous speech differ for production versus perception and for the consistent and inconsistent playback conditions. Sentence-level ERPs for both perception and production were epoched to the acoustic onset of sentence articulation (the first phoneme in the trial sentence). These ERPs demonstrated a relative suppression of EEG activity in production trials compared with perception trials (Figure 2.2). The N1 and P2 components are present at the sentence level in both perception and production conditions but reduced in amplitude for the production trials. I fit LME models (Equation 2.1) comparing perception and production in windows of interest (`windowed_amplitude` $\sim$ `Condition + (1|Subject)` and `windowed_latency` $\sim$ `Condition + (1|Subject)`). I report the estimated marginal mean (EMM) and standard error of the contrasts between perception and production responses here. I found significantly lower amplitudes for N1 ($EMM_{\text{perception}-\text{production}} = -2.31 \pm 0.15\ \mu V$; $p < .001$) and significantly higher

amplitudes for P2 ($EMM_{\text{perception}-\text{production}} = 1.72 \pm 0.15\ \mu V$; $p < .001$) during perception compared with production. This was also in line with increased peak-to-peak amplitude ($EMM_{\text{perception}-\text{production}} = 3.96 \pm 0.15\ \mu V$; $p < .001$) in perception compared with production. In addition, N1 latency was decreased in production compared with perception ($EMM_{\text{perception}-\text{production}} = 1.64 \pm 0.47$ msec; $p < .001$), and similar results were seen for P2 latency ($EMM_{\text{perception}-\text{production}} = 2.75 \pm 0.66$ msec; $p < .001$). Suppression during speech production relative to perception in this task highlights differences in processing internally and externally generated speech.

Next, differences in consistent and inconsistent playback trials were assessed to evaluate the presence or absence of a suppression similar to the one observed between perception and production. Although differences were significant between perception and production trials, the differences between consistent and inconsistent speech perception were less pronounced: LME modeling (`Window` $\sim$ `Condition + (1|Subject)`) did not reveal a significant difference in N100 ($EMM_{\text{consistent}-\text{inconsistent}} = 0.31 \pm 0.20\ \mu V$; $p = .12$) and P200 ($EMM_{\text{consistent}-\text{inconsistent}} = -0.20 \pm 0.22\ \mu V$; $p = .37$) amplitudes across this contrast. However, peak-to-peak amplitude ($EMM_{\text{consistent}-\text{inconsistent}} = -0.52 \pm 0.24\ \mu V$; $p = .03$) and N100 latency ($EMM_{\text{consistent}-\text{inconsistent}} = 1.50 \pm 0.66$ msec; $p = .02$) differed significantly between consistent and inconsistent trials, with an earlier response to inconsistent compared with consistent playback. P2 latency did not differ significantly ($EMM_{\text{consistent}-\text{inconsistent}} = 0.18 \pm 0.91$ msec; $p = .84$). Because consistent and inconsistent perception tri-

als were split into blocks during the task, an oddball response was not elicited for the inconsistent stimuli. To further investigate the significance of peak-to-peak amplitude and N1 latency between consistent and inconsistent perceptual stimuli, a series of Wilcoxon signed-ranks tests with Benjamini-Yekutieli correction (Benjamini & Yekutieli 2001) comparing N1-P2 peak-to-peak amplitude and N1 latency on a within-subject basis were performed. These significance tests revealed only three individual participants that demonstrated a significant suppression between consistent and inconsistent speech perception (OP1 $p = .02$; OP7 $p < .001$; OP21 $p = .0002$), and only two participants with a significant difference in N1 latencies (OP1 $p = .004$; OP19 $p = .04$). This within-subject analysis suggests the significance of peak-to-peak amplitude and N1 latency observed in the LME results is caused by outlier participants rather than a generalizable effect. Overall, differences within consistent and inconsistent perception trials were less pronounced than the differences between perception and production trials. These minor differences between expected and unexpected speech perception suggest that SIS is not fundamentally linked to general forward modeling of speech production. In other words, feedforward processing of speech perception and feedforward processing of speech production reflect different neural mechanisms.

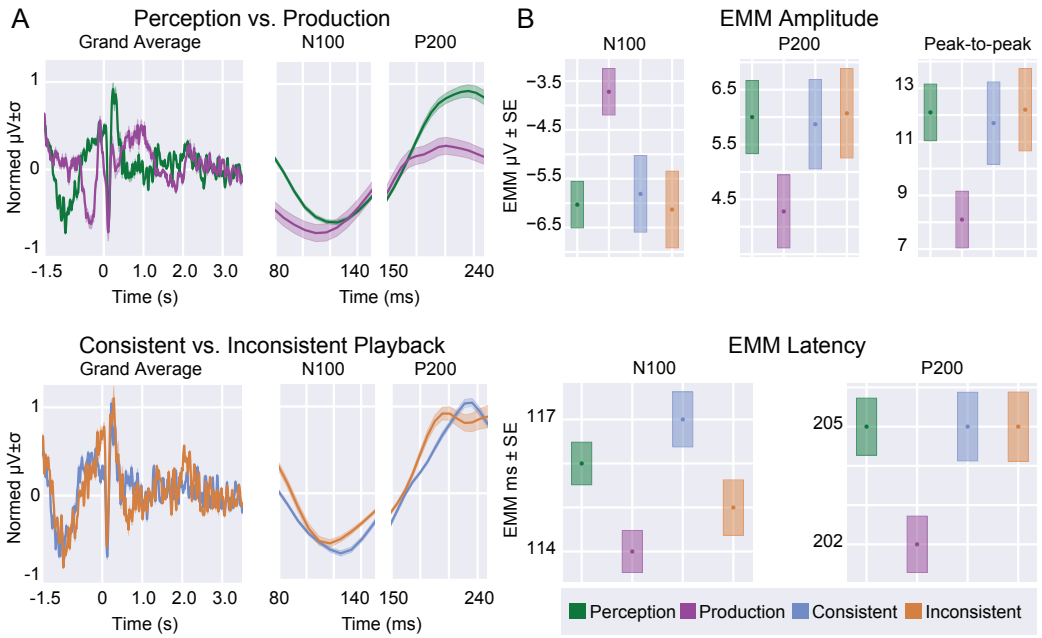

Figure 2.2: **ERPs to sentence onset demonstrate suppression of N1-P2 during speech production**.
Speech production (purple) is suppressed relative to perception (green), but no such difference is observable for consistent (yellow) versus inconsistent (light blue) speech perception.
(A) Grand average ERPs and N1/P2-windowed ERPs comparing speech production and speech perception (top) and consistent and inconsistent speech perception (bottom).
(B) LME model EMMs for the four experimental conditions' amplitudes (top) and latencies (bottom). Shaded area represents standard error.

2.5.2 Suppression of phonological feature tuning during speech production

While the ERP results provide insight into the timing and magnitude of differences in responses during perception and production, they do not pro-

vide information regarding any potential differences in responses to specific speech features or content. Furthermore, ERP analyses are constrained by the need to average many trials that are time-locked to a particular event (Luck 2014). Thus, ERP analyses may not be as sensitive to uncovering differences outside of the onset of the sentence, or for specific phonological features within continuous speech. To address this limitation, I performed additional analyses where I fit mTRF models for continuous production and perception (Equation 2.2). These analyses are powerful in that they allow for investigation of continuous, natural speech without the need for trial averaging. Although I could perform a phoneme-by-phoneme ERP analysis to show task-related differences across the sentence, such an analysis would suffer from an inability to account for coarticulation or other temporal correlations of activity. The mTRF model regression weights are calculated for multiple time delays simultaneously, allowing the model to account for activity in response to combinations of features across time (Theunissen et al. 2000). They also allow me to further probe specific differences (or lack thereof) in tuning across my different task conditions.

Model performance was evaluated by calculating the linear correlation coefficients (r) between the EEG response predicted by the model and the actual response for held out data not used to train the model. I also probed the importance of individual features on model performance by ablating specific features from the stimulus matrix S and observing the change in correlation coefficients between ablated and full models. Similar variance partitioning

methods have been used to uncover the unique variance explained by particular features (Hamilton et al. 2021; de Heer et al. 2017). For example, if a model that omitted normalized EMG predicted the neural response less accurately, the interpretation is that EMG contains important information for accurately modeling EEG activity. For each task-related feature in the "full" model (14 phonological + 4 task features; Figure 2.3), I fit a separate model omitting that feature. Lastly, one model had two additional sets of phonological features (i.e., 14 phonological features during production + 14 phonological features during perception + 14 phonological features in either condition + 4 task features; Figure 2.3). These were split by modality to observe if phonological feature tuning changed between perception and production. I call this model the "task-specific" encoding model, which is in comparison to the "identical" encoding model in which phonological feature tuning is assumed to be the same across all conditions, with only a baseline change fit by the two condition features (perception and production). The *EMM*s of the contrast in correlation coefficients between models were evaluated via LME modeling with subject and channel location as random effects ($r \sim$ `Model + (1|Subject) + (1|Channel)`). Separating phonological tuning by the modality of speech (i.e., perception vs. production) had a significant effect on model performance ($EMM_{\text{identical}-\text{separate}}$ $r = -0.014 \pm 0.003$; $p < .001$), such that separating phonological feature tuning during production from phonological feature tuning during perception improved the model's ability to predict the held-out neural response (Figure 2.3). This result, which was contrary to my initial

hypothesis, suggested that phonological feature encoding differs during speech perception and production. However, because of the influence of EMG artifact during speech production, speech perception in this task is a combination of sensory and motor responses, whereas speech perception in this task is purely sensory, which may explain the difference in the models presented in Figure 2.3.

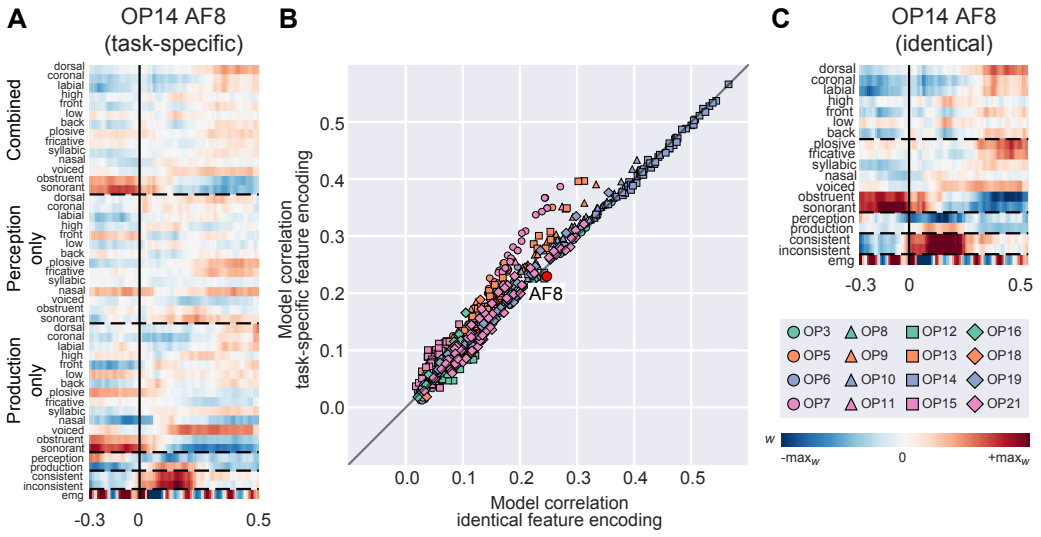

Figure 2.3: **Separating phonological feature encoding by modality of speech improves model performance**.
(A) Temporal receptive field for an individual electrode with stimulus characteristics divided by task condition (i.e., perception vs. production).
(B) Scatter plot of channel-by-channel correlation coefficients between two compared models. Color and markers are used to denote individual participants. Diagonal black line represents unity (equal model performance).
(C) Temporal receptive field for an individual electrode with stimulus characteristics identical across task condition.

Although I utilized methods to correct for EMG artifact that have been

previously demonstrated in the literature to be successful (Ries et al. 2021; Chen et al. 2019; Vos et al. 2010), there is no definitive way to rule out residual EMG given the lack of ground truth in the sources that contribute to the electroencephalogram. As a result, I further explored the influence of EMG artifact on model performance by fitting mTRFs that included normalized EMG activity recorded from auxiliary facial electrodes in tandem with the EEG as a regressor. Models that include or exclude the auxiliary EMG but are otherwise identical in their stimulus matrices were compared in an ablation-based approach to explore the contribution of specific features to model performance (Ivanova et al. 2021a). Linear correlation coefficients were compared using an LME model identical to the model used for comparing the "identical" versus "task-specific" models described above. The inclusion or exclusion of normalized EMG in the stimulus matrix significantly affected model performance regardless of whether phonological features were task specific ($p < .001$) or identical ($p < .001$). Including information about normal- ized EMG activity recorded from auxiliary facial electrodes improved model performance (Figure 2.4) as shown by the greater number of channels below the unity line. On an individual participant basis, all but two participants (OP6 and OP16) showed a significant difference in model performance across the inclusion or omission of normalized EMG activity as a stimulus feature as assessed by Wilcoxon signed-ranks test. When comparing the relative difference between "identical" and "task-specific" models (Figure 2.3) in the presence or absence of an EMG regressor, models including an EMG regressor showed less of a difference

in performance between methods of phonological feature encoding, suggesting that residual EMG decreases the stability of phonological feature tuning across modalities of speech (Figure 2.4). A verification of artifact removal in the context of the ERP results reported above is provided in Appendix A.

Although the linear encoding model results occur on a phoneme-by-phoneme timescale (cf. the sentence-level ERPs presented in Figure 2.2), the EEG data used to fit the models were collected during a task that elicited SIS. Thus, I sought to identify any reduction in phonological feature response between production-specific and perception-specific feature weights in an effort to link the linear encoding model results to the ERP analysis presented earlier in this chapter. Feature-by-feature, I calculated the correlation coefficient (r) of the production-specific and perception-specific weights of the task-specific mTRF (Figure 2.3) and observed a strong negative correlation between the production and perception-specific weights (Figure 2.5).

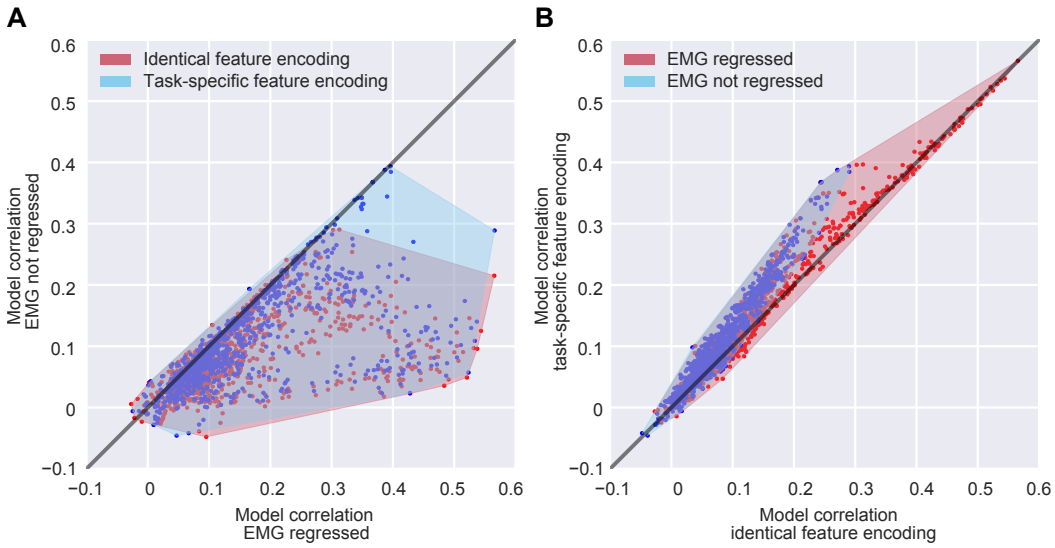

Figure 2.4: **Including EMG as an encoded feature in linear models greatly improves their performance, as well as the stability of phonological feature encoding between perception and production**.
(A) Individual electrodes' correlation coefficients with held-out neural response within models that do contain an EMG regressor (x axis) and those that do not (y axis), for models that separate phonological feature tuning by task modality (blue) and models that do not (red). Diagonal black line represents unity. Shaded area is the convex hull of points within each group to show overall trends.
(B) Individual electrodes' correlation coefficients with held-out neural response within models that differentially encode phonological features according to modality of speech (y axis) and those that do not (x axis), in the presence (red) or absence (blue) of information about normalized EMG activity recorded from auxiliary facial electrodes. Diagonal black line represents unity. Shaded area same as (A). When EMG was regressed, more points lie along the unity line, indicating similar phonological feature tuning and that EMG may be captured in the different phonological features when it is not available as a regressor.

Trial-specific stimulus features were also ablated to assess their contribution to model performance. Omitting trial modality (i.e., whether a

phoneme was produced or perceived) did not significantly affect the mTRF model's ability to predict the held-out neural response ($p = .65$). Similarly, ablating information about whether the perception trials were consistent or inconsistent with their preceding production trials did not affect model performance ($p = .56$). If the EMG regressor is removed in conjunction with trial-specific features, the differences in model performance when trial modality is included or ablated are less profound but still nonsignificant ($p = .23$). When ablating playback consistency, no changes are observed in significance between inclusion ($p = .56$) and omission ($p = .54$) of an EMG regressor, which is expected considering this contrast is constrained to perception trials where EMG associated with articulation is absent from the response. The ablation of consistency contrast not affecting model performance is in line with the ERP results presented above (Figure 2.2). However, ablating trial modality (i.e., perception vs. production) not affecting model performance is incongruent with the ERP results, for which a stark contrast between perception and production were observed. The difference in time frame between the ERP analysis (sentence level) and the linear encoding model analysis (phoneme level) may explain the difference between the ERP and mTRF results. In other words, sentence-level processing of speech perception and production may involve different neural mechanisms, but at an individual phoneme level, the mechanisms are shared between perception and production. Alternatively, the incorporation of the EMG regressor may be delineating perception and production in the model, making explicit information about trial modality, effectively marking

the explicit inclusion of trial type in the stimulus matrix redundant. This explanation is supported by the observation that omission of an EMG regressor substantially impacted model performance.

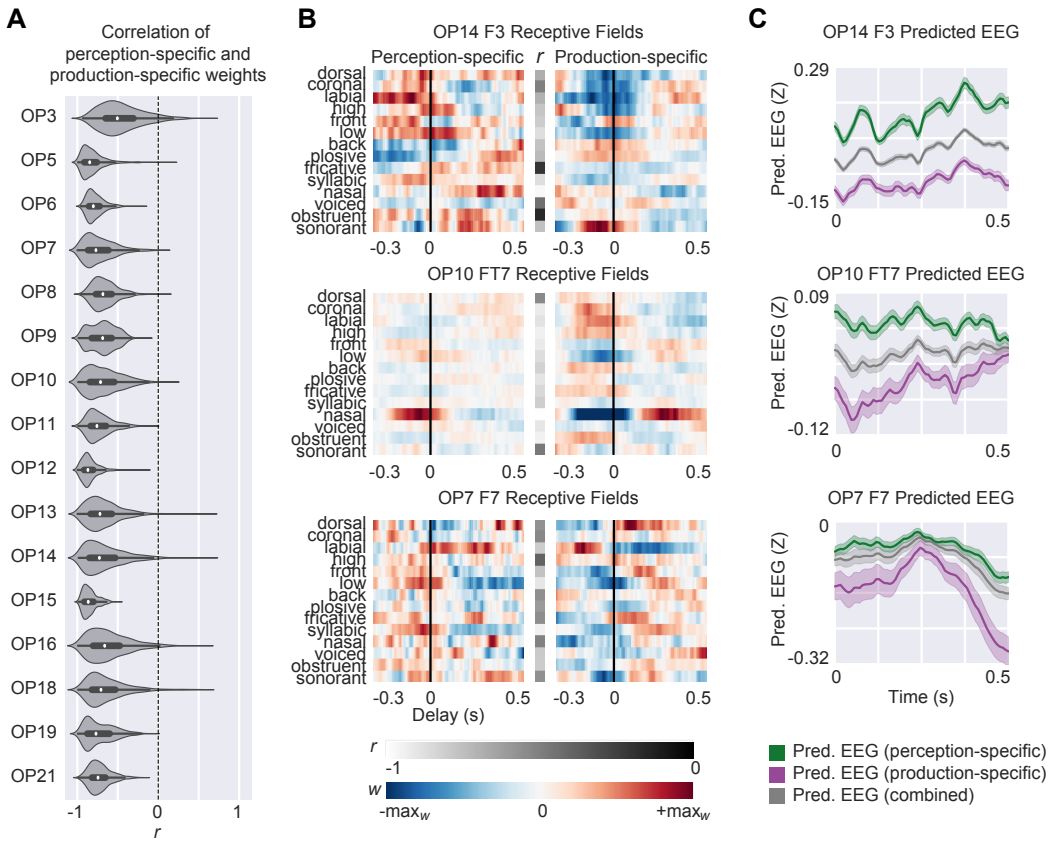

Figure 2.5: **Production-specific and perception-specific phonological feature weights are strongly negatively correlated with each other, suggesting a suppressive relationship**.
(A) Violin plot showing the distribution of channel-by-channel, feature-by-feature correlation coefficients between phonological features specific to perception and phonological features specific to production, separated by individual participant. Thick line in violin interior represents range of Quartiles 1–3. Density of plot (violin width) scaled by individual participant.
(B) Temporal receptive fields for three individual electrodes. Phonological feature weights taken from task-specific model separated into perception-specific (left) and production-specific (right) receptive fields. Center grayscale column represents the correlation of each row of weights between the two receptive fields.
(C) Predicted EEG activity for the held-out validation set as predicted by the perception-specific (green), production-specific (purple), or combined (gray) phonological feature weights. Electrodes OP14 F3 (predicted vs. actual EEG $r = .53$; $p < .001$) and OP10 FT7 (predicted vs. actual EEG $r = .42$; $p < .001$) exhibit similar model performance between task-specific and identical phonological feature encoding models (i.e., lie along unity line of Figure 2.3), whereas electrode OP7 F7 (predicted vs. actual EEG $r_{\text{task-specific}} = 0.37$; $r_{\text{identical}} = 0.24$; $p < .001$) exhibits diverging model performance between models. Overall, predicted EEG based on the production-specific weights was lower in amplitude than predicted EEG based on the combined or perception-specific weights. All mTRFs presented in this figure are from the task-specific model that included an EMG regressor.

Taken together, the mTRF results suggest that linguistic abstractions remain invariant during speaking and listening. Similar encoding of phonological features between these modes of speech after EMG regression suggests the amplitude reduction corresponding to SIS is not explicable by a difference in linguistic abstraction, constraining it to endogenic sensorimotor processes. However, a feature-by-feature correlation shows an inverse relationship between the encoding of phonological features during speaking and listening,

demonstrating a suppressive relationship between speaking and listening can be observed on an individual phoneme timescale throughout sentences that exhibit SIS at sentence onset. Methodologically, the mTRF results show that regressing EMG activity recorded from auxiliary electrodes during the task is an informative characteristic of the stimulus in the context of modeling neural responses to speech. Including information about trial type (perception vs. production, consistent vs. inconsistent playback) was less informative when EMG was included as a regressor, potentially the result of fundamental differences in expected residual EMG between articulating speech and passively listening. EMG regression also reduces phonological feature tuning changes across modality of speech, suggesting residual EMG artifact in the postprocessed signal is responsible for changes in phonological feature tuning, as well as motivating auxiliary EMG recordings as a safeguard against residual EMG in the postprocessed response.

2.6 Conclusion

The results presented in this chapter demonstrate a difference in EEG responses to perceiving and producing naturalistic stimuli. At the sentence level, a suppression of early auditory components N1 and P2 was observed in speech production relative to perception. These findings are in line with previous literature on SIS and auditory processing more generally, which have identified that internally produced stimuli generate less of a change in neural activity than externally produced stimuli. This study sought to replicate the

phenomenon of SIS in a more naturalistic setting, as many studies of SIS use low-level acoustic stimuli such as pure tones (Martikainen et al. 2005) and single vowels (Niziolek et al. 2013; Heinks-Maldonado et al. 2006; Houde et al. 2002), whereas many neurolinguistic studies now use more naturalistic stimuli such as podcasts (Goldstein et al. 2022; Huth et al. 2016), audiobooks (Herff et al. 2015), and movie trailers (Desai et al. 2021) in an effort to better capture how speech and language are used in daily life (Hamilton & Huth 2020). I was able to demonstrate SIS at the sentence level, which is comparatively much more naturalistic than the lower-level characteristics of speech used in previous studies of SIS. In contrast with the suppression observed between perception and production trials, differences between EEG responses to consistent and inconsistent perceptual trials were minor.

To investigate whether linguistic abstraction into phonological features persists during SIS, in neural responses to these two modes of speech, I fit linear encoding models describing neural activity as a function of different stimulus features. These features allowed us to test different hypotheses about changes in phonological tuning at the individual feature level versus overall baseline changes during perception and production. Performance of these models were evaluated by how well the mTRF weights correlated with held-out EEG response. Differentially encoding phonological features during perception and production in the stimulus matrix yielded higher model performance; however, residual EMG artifact may be driving performance improvements in the differential phonological features model, considering the inclusion of normal-

ized EMG recorded from facial electrodes substantially improved model performance. As EMG activity is expected to disproportionately affect speech production because of articulatory movement, residual EMG that is unaccounted for with an additional regressor may be providing the model with a clear contrast between the perception and production conditions that is spuriously encoded in the separation of phonological features across modes of speech. Accordingly, the inclusion of the EMG regressor reduces the variance in phonological feature encoding between perception and production by accounting for uncorrected EMG artifact in a separate regressor. Thus, I conclude that phonological feature encoding is a shared representation during speaking and listening. Despite similar ability (after regressing EMG) to predict held-out EEG as models that do not separate phonological features into whether they occurred during perception or production, models that do separate phonological features show a strong negative correlation between task-specific phonological features (Figure 2.5).

The ERP and mTRF analyses presented in this chapter extend our understanding of SIS by scaling the phenomenon into a more naturalistic contex and exploring the interaction of SIS and higher-order linguistic abstractions that take place during speech perception and production. This research hopes to illuminate differences in electrophysiological responses to perception and production and motivate future study of naturalistic speech production with noninvasive EEG.

Chapter 3

sEEG Results: Processing of auditory feedback in perisylvian and insular cortex

3.1 Preface

This chapter of my dissertation contains original research designed, collected, analyzed, interpreted, and written by me. The entirety of this chapter was adapted from a journal article that is currently under peer review. A preprint is available on *bioRxiv* (Kurteff et al. 2024). Data for this chapter are not publicly available because they could compromise research participant privacy and consent. You may contact my doctoral advisor to request access by emailing liberty.hamilton@austin.utexas.edu. The accompanying code, however, is publicly available as a GitHub repository.

I am grateful to my co-authors on the submitted article for their assistance in all steps in preparation of the manuscript and this chapter. The clinical teams at Dell Children's Medical Center (Elizabeth C. Tyler-Kabara, Dave Clarke), Texas Children's Hospital (Howard L. Weiner, Anne E. Anderson, Andrew Watrous), and Dell Seton Medical Center (Robert J. Buchanan, Pradeep N. Modur) made collection of these data possible, while Saman Asghar and Alyssa M. Field in the Hamilton Lab helped with data collection

and preprocessing. I am also grateful to lab members Maansi Desai and Elise Rickert for their assistance in data collection and their feedback on the project as it developed. My PhD supervisor Liberty S. Hamilton provided assistance with every step of the project.

Lastly, I would like to thank the patient participants at our three recording sites (Dell Children's Medical Center, Texas Children's Hospital, Dell Seton Medical Center) for volunteering time during their arduous hospital stay to participate in this research. Listening to your own voice isn't very fun, and I imagine having epilepsy surgery isn't much fun either, so I am incredibly grateful for the opportunity to do this research.

3.2 Abstract

When we speak, we not only make movements with our mouth, lips, and tongue, but we also hear the sound of our own voice. Thus, speech production in the brain involves not only controlling the movements we make, but also auditory and sensory feedback. Auditory responses are typically suppressed during speech production compared to perception, but how this manifests across space and time is unclear. Here I recorded intracranial EEG in seventeen pediatric, adolescent, and adult patients with medication-resistant epilepsy who performed a reading/listening task to investigate how other auditory responses are modulated during speech production. I identified onset and sustained responses to speech in bilateral auditory cortex, with a selective suppression of onset responses during speech production. Onset responses

provide a temporal landmark during speech perception that is redundant with forward prediction during speech production. Phonological feature tuning in these "onset suppression" electrodes remained stable between perception and production. Notably, the posterior insula responded at sentence onset for both perception and production, suggesting a role in multisensory integration during feedback control.

3.3 Introduction

A key component of speaking is the integration of ongoing sensory information from the auditory, tactile, and proprioceptive domains (Hickok 2014; Tourville et al. 2008). When we read a sentence out loud, our brain must convert visual information into a motor program for moving our articulators (lips, jaw, tongue, larynx) to create sounds. The brain then processes these sounds as they are uttered, so the talker can hear if they have made a mistake. Auditory information is processed differently during speaking compared to listening (Cogan et al. 2014; Creutzfeldt et al. 1989; Houde et al. 2002; Nourski et al. 2021; Towle et al. 2008). A prime example is speaker-induced suppression (SIS), a phenomenon in which self-generated speech generates a lower amplitude neural response than externally generated speech (§1.2.3.1; Behroozmand & Larson (2011); Martikainen et al. (2005); Flinker et al. (2010)). SIS and related phenomena are components of the speech motor control system, the purpose of which is to ensure ongoing sensory feedback is in line with feedforward expectations generated prior to articulation (Guenther 2016; Houde & Nagarajan

2011; Tourville & Guenther 2011). This link is established by studies that correlate the extent of cortical suppression with the accuracy of the utterance: both speech errors and subphonemic changes in utterance acoustics can result in decreased cortical suppression, indicative of a feedback control system ready to adjust the motor program in real time (Niziolek et al. 2013; Ozker et al. 2022). While feedback control has primarily been studied using noninvasive techniques with a lower signal-to-noise ratio (e.g., EEG, MEG; (Chang 2015; Houde et al. 2002; Okada et al. 2018)), intracranial recordings allow for more precise investigation of this process (Chang 2015; Hamilton 2024; Mercier et al. 2022; Lachaux et al. 2012). This can potentially illuminate the spatiotemporal specificity of feedback suppression mechanisms like SIS. In addition, we can investigate how speech production affects other aspects of the perceptual system, such as linguistic abstraction and neural response timing.

3.3.1 Organization of speech cortex during listening and speaking

Transformation of low-level acoustics into some form of intermediate linguistic representation is a necessary component of speech perception (Appelbaum 1996). In several studies, this abstraction is organized according to place and manner of articulation, motivated by linguistic feature theory. Place of articulation describes the location of constriction in the vocal tract (e.g., a bilabial /b/ sound is produced by closing the lips). Manner of articulation, on the other hand, describes the degree of constriction and airflow through the vocal tract. Mesgarani and colleagues used electrocorticography (ECoG)

to observe tuning of electrode populations within the superior temporal gyrus (STG) that preferentially responded to specific classes of phonological features (namely manner of speech) during passive listening (Mesgarani et al. 2014). For example, the same intracranial electrode might respond selectively to plosive phonemes such as /b/, /d/, /g/, /p/, /t/, and /k/, while not responding to fricatives such as /f/, /v/, /s/, /ʃ/. In more recent work, the same level of representation was observed at the single neuron level (Lakretz et al. 2021; Leonard et al. 2023). The same group that identified STG electrodes tuned to specific classes of phonological features later expanded on this result using a speech production task to demonstrate feature tuning changes during speech production in the motor cortex (Cheung et al. 2016). Notably, they observed that motor cortex was organized according to place of articulation during speech production, as would be expected from somatotopic representations (Bouchard et al. 2013), but organized according to manner of articulation during passive listening. However, this manuscript did not report on responses in STG during speech production, nor was a direct comparison of phonological tuning made between perception and production.

A more recent insight about how the auditory system is organized comes from research on temporal response profiles in the STG (Hamilton et al. 2018). The STG contains two such profiles: first, an "onset" response region localized to posterior STG with high temporal modulation selectivity (Hullett et al. 2016) that transiently responds to the acoustic onset of a stimulus. These onset responses are useful for segmenting continuous acoustic information into

discrete linguistic units, such as phrases and sentences. Second, a "sustained" response region localized to middle STG with a longer temporal integration window that does not show the same strongly adapting responses following sentence onset. Onset and sustained response profiles are a globally organizing feature of speech-responsive cortex, and responses to all phonological features are seen across both (Hamilton et al. 2018). If responses to phonological information can be modified by the acoustic context of a sound, it is possible they could also be modulated by feedback suppression during speech production. Other top-down cognitive processes can affect speech perception as well, such as expectations about upcoming stimuli evidenced in both speech production (Goregliad Fjaellingsdal et al. 2020; Lester-Smith et al. 2020; Scheerer & Jones 2014) and speech perception (Astheimer & Sanders 2011; Bendixen et al. 2014; Caucheteux et al. 2023). In general, auditory stimuli that are consistent with the listener's expectations generate less of a response than inconsistent stimuli (Chao et al. 2018; Forseth et al. 2020). While consistency effects are also a component of the motor system (Gonzalez Castro et al. 2014; Shadmehr & Krakauer 2008), the link between speaker-induced suppression and more general top-down expectation is not well established.

3.3.2 Speaker-induced suppression in noninvasive recordings

Recent research from my group (presented in Chapter 2) used scalp EEG recordings to demonstrate that responses to continuous sentences are suppressed during production compared to perception of those same sentences

while phonological tuning remains unchanged (Kurteff et al. 2023). However, such conclusions may be tempered by the low spatial resolution of scalp recordings, motivating the use of high-resolution intracranial stereo EEG (sEEG) recordings. When we plan to speak, the motor efference copy contains expectations about upcoming auditory feedback and may contain information about temporal/linguistic landmarks in that feedback (Levelt 1993; Niziolek et al. 2013; Schneider et al. 2014). Onset responses, which encode the temporal landmarks of speech, may then be suppressed as a redundant processing component during speech production. This is corroborated by scalp EEG/MEG research showing that SIS occurs primarily within the N1/M1 components. That is, the N1 and M1 neural responses are suppressed during speaking as compared to playback. The N1/M1 component is an early-onset neural response that is observed at acoustic edges with high temporal modulation (Luck 2014), making these components share characteristics with onset responses observed using invasive recordings.

3.3.3 The role of the insula in speech perception and production

The use of sEEG as a recording methodology affords an additional advantage to the current study: the ability to record from deeper structures in the cortex. One such structure is the insula, a multifunctional region that is theorized to be involved in sensory, motor, and cognitive aspects of speech (Kurth et al. 2010). Recent work using sEEG reported the insula to be more active for self-generated speech when compared to externally generated speech,

an opposite trend to the cortical suppression of self-generated speech observed in auditory cortex (Woolnough et al. 2019). The insula is difficult to record from using several popular neuroimaging techniques due to its placement deep in the Sylvian fissure (Chang 2015; Remedios et al. 2009). In speech, the insula conventionally plays a role in pre-articulatory motor coordination (Dronkers 1996). Because of the proximity of the insula to the temporal plane and hippocampus, insular coverage is rather common in sEEG epilepsy monitoring cases (Nguyen et al. 2022). I aim to expand upon the functional role of the insula in speech perception and production by directly comparing auditory feedback processing and phonological feature encoding during speaking and listening while recording from the region in high resolution.

3.3.4 Aims

To address how cortical suppression during speech production interacts with documented organizational phenomena during speech perception such as linguistic abstraction and onset/sustained response profiles, I used high-resolution sEEG recordings of neural activity from electrodes implanted in the cortex as part of surgical epilepsy monitoring (Guenot et al. 2001). These participants completed a dual speech production-perception task where they first read sentences aloud, then passively listened to playback of their reading to identify potential changes in local field potential recorded by the implanted electrodes. My first goal was to identify if previously identified onset and sustained response profiles in auditory cortex (Hamilton et al. 2018) were also

present during speech production. Additionally, I varied the playback condition between a consistent playback of the preceding production trial and a randomly selected playback inconsistent with the preceding trial to assess the spatial and temporal similarity of a more general perceptual expectancy effect with feedback suppression during speech production. Lastly, I investigated how linguistic feature tuning changes at individual electrodes during speech production vs. perception and how this is modulated by expectation. My results have implications for understanding important auditory-motor interactions during natural human communication.

3.4 Methods

3.4.1 Subject details

17 individuals (sex: 9F; age: 16.6±6.4, range 8-37; race/ethnicity: 8 Hispanic/Latino, 6 White, 1 Asian, 2 multi-racial) undergoing intracranial monitoring of seizure activity via sEEG for medically intractable epilepsy were recruited from three hospitals: Dell Children's Medical Center in Austin, Texas ($n = 13$); Texas Children's Hospital in Houston ($n = 3$), Texas; and Dell Seton Medical Center in Austin, Texas ($n = 1$). Demographic and relevant clinical information is provided in Table D.2. Participants (and for minors, their guardians) received informed consent and provided written consent for participation in the study. All experimental procedures were approved by the Institutional Review Board at the University of Texas at Austin.

3.4.2 Neural data acquisition

Intracranial sEEG and ECoG data from a total of 2044 electrodes across subjects were recorded continuously via the epilepsy monitoring teams using a Natus Quantum headbox (Natus Medical Incorporated, San Carlos, CA, USA). At Texas Children's Hospital, sEEG depths (AdTech Spencer Probe Depth electrodes, 5mm spacing, 0.86mm diameter, 4-16 contacts per device), strip electrodes (AdTech) and grids (AdTech custom order, 5mm spacing, 8x8 contacts per device) were implanted in the brain by the neurosurgeon in brain areas that are determined via clinical need. At Dell Children's Medical Center and Dell Seton Medical Center, sEEG depths (PMT Depthalon, 0.8mm diameter, 3.5mm spacing, 4-16 contacts per device) were used. A TDT S-Box splitter was used at Dell Children's Medical Center to connect the data stream to a TDT PZ5 amplifier, which then recorded the local field potential from the sEEG electrodes onto a research computer running TDT Synapse via a TDT RZ2 digital signal processor (Tucker Davis Technologies, Alachua, FL, USA). Speaker (perceived; GENELEC 1029A powered monitor speakers) and microphone (produced; Audio Technica U853A cardioid condenser microphone connected to an RDL STM-2 preamp) audio were also recorded via RZ2 at 22 kHZ to circumvent downsampling of audio by the clinical recording system. At the other two recording locations, use of a dedicated research recording system was not possible due to clinical constraints; instead, the auditory stimuli from the iPad were simultaneously played to the participant through speakers and recorded directly on the clinical system using an audio

splitter cable, allowing synchronization of stimulus events with neural activity. Because audio recorded on the clinical system could only be collected at 3 kHz, simultaneous high-resolution audio was recorded for both speaking and playback using an external microphone (Sennheiser MD 42) and a second splitter cable from the iPad both plugged into a MOTU M4 USB audio interface (MOTU, Cambridge, MA, USA) plugged into the research computer running Audacity recording software. After the recording session, a match filter was used to synchronize high-resolution audio from the external recording system to the neural data recorded on the clinical system (Turin 1960). Intracranial data were recorded at 3 kHz and downsampled to 512 Hz before analysis for all sites.

3.4.3 Data preprocessing

Data were preprocessed offline using a combination of custom MATLAB scripts and custom Python scripts built off the `mne` software package (Gramfort et al. 2014). First, data were notch filtered at 60/120/180 Hz to remove line noise, then channels were manually inspected and rejected. Bad channels were identified through visible noise or excessive artifact in the signal. Next, a common average reference was applied across all non-bad channels. The high gamma analytic amplitude response (Lachaux et al. 2012), which has been shown to strongly correlate with speech (Kunii et al. 2013) and serves as a proxy for multi-unit neuronal firing (Ray & Maunsell 2011), was extracted via Hilbert transform (8 bands, log spaced, Gaussian kernel, 70-150 Hz). Lastly,

the 8-band Hilbert transform response was Z-scored relative to the mean activity of the individual recording block. All preprocessing and subsequent analyses were performed on a research computer with the following specifications: Ubuntu 20.04, AMD Ryzen 7 3700X, 64GB DDR4 RAM, Nvidia RTX 2060.

3.4.4 Electrode localization

Electrodes' locations were registered in the three-dimensional Montreal Neurological Institute (MNI) coordinate space (Evans et al. 1993). Electrodes were localized through coregistration of an individual subject's T1 MRI scan with their CT scan using the Python package `img_pipe` (Hamilton et al. 2017). Three-dimensional reconstructions of the pial surface were created using an individual subject's T1 MRI scan in Freesurfer and anatomical regions of interest for each electrode were labeled using the Destrieux parcellation atlas (Dale et al. 1999; Destrieux et al. 2010). These reconstructions were then inflated for better visualization of intra-Sylvian structures such as the insula and Heschl's gyrus via Freesurfer. To visualize electrodes on the new inflated mesh, electrodes were projected to the surface vertices of the inflated mesh, which maintained the same number of vertices as the default pial reconstruction. To preserve electrode location using inflated visualization, each electrode was projected to a mesh of its individual Freesurfer region of interest (ROI) before projection to inflated space. Additionally, any depth electrodes greater than 4 millimeters from the cortical surface ($n = 691$) were not visualized on inflated

surfaces due to a previously identified spatial falloff in high gamma frequency bands for electrodes greater than 4 millimeters apart from each other (Muller et al. 2016). Electrodes greater than 4 millimeters from the cortical surface, while excluded from visualization, were included in analyses if they contained a robust response ($p < 0.05$ for bootstrap procedure, $r \geq 0.1$ for TRF modeling) to any task stimuli. To visualize electrodes across subjects, electrodes were nonlinearly warped to the `cvs_avg35_inMNI152` template reconstruction (Dale et al. 1999) using procedures detailed in Hamilton et al. (2017). While nonlinear warping ensures individual electrodes remain in the same anatomical region of interest as they were in native space, it does not preserve the geometry of individual devices (depth electrodes or grids). For inflated visualization in warped space, an identical ROI-mesh-to-inflated-surface projection method as described above was utilized, but the ROI and inflated meshes were generated from the template brain instead. Anatomical regions of interest were always derived from the electrodes in the original participant's native space.

3.4.5 Overt reading and playback task

The task used is identical to the one described in Chapter 2. The task was designed using a dual perception-production block paradigm, where trials consisted of a dyad of sentence production followed by sentence perception. Both perception and production trials were preceded by a fixation cross and broadband click tone (Figure 3.2A). Production trials consisted of participants overtly reading a sentence, then the trial dyad was completed by

participants listening to a recording of themselves reading that produced sentence. Playback of this recording was divided into two blocks of consistent and inconsistent perceptual stimuli: consistent playback matched the immediately preceding production trial, while inconsistent playback stimuli were instead randomly selected from the previous block's production trials. The generation of perception trials from the production aspect of the task allowed stimulus acoustics to be functionally identical across conditions.

Sentences were taken from the MultiCHannel Articulatory (MOCHA) database, a corpus of 460 sentences that include a wide distribution of phonemes and phonological processes typically found in spoken English (Wrench 1999). A subset of 100 sentences from MOCHA were chosen at random for the stimuli in the present study; however, before random selection, 61 sentences were manually removed for either containing offensive semantic content or being difficult for an average reader to produce to reduce extraneous cognitive effects and error production, respectively.

Unlike in Chapter 2, a modified version of the task optimized for participants with a lower reading level was created so that pediatric participants could perform the task as close to errorless as possible. This version took the randomly selected MOCHA sentences from the main task and shortened the length and utilized higher-frequency vocabulary that still encompassed the range of phonemes and phonological processes found in the initial dataset. Seven of the seventeen participants (TC1, TC3, DC10, DC12, DC13, DC16, DC17) completed the easy-reading version of the task. Participants completed

the task in blocks of 20 sentences (25 sentences for the easy-reading version) produced and subsequently perceived for a total of 40 (50) trials per block. Participants produced (and listened to subsequent playback of) an average of 142±61 trials. A broadband click tone was played in between trials.

Stimuli were presented in the participant's hospital room on Apple iPad Air 2 using custom interactive software developed in Swift (Apple). Auditory stimuli were presented at a comfortable listening level via external speakers. Insert earbuds and/or other methods of sound attenuation (e.g., soundproofing) were not possible given the clinical constraints of the participant population. Visual stimuli were presented in a white font on a black background after a 1000 millisecond fixation cross. Accurate stimulus presentation timing was controlled by synchronizing events to the refresh rate of the screen. The iPad was placed on an overbed table and trials were advanced by the researcher using a Bluetooth keyboard. Participants were instructed to complete the task at a comfortable pace and were familiarized with the task before recording began. Timing information was collected by an automatically generated log file to assist in data processing.

As mentioned above, electrodes >4 millimeters from the cortical surface were automatically excluded from visualization. However, electrodes identified as outside the brain or its pial surface via manual inspection of the subject's native imaging were excluded from all analyses. Electrodes in a ventricle or in a lesion were excluded using the same method. Adjacent electrodes that displayed a similar response profile to outside-brain electrodes were also excluded;

conversely, electrodes on the lateral end of a device that displayed a markedly different response profile than medially adjacent electrodes were determined to be outside the brain and thus excluded. As an additional measure of manual artifact rejection, channels that displayed high trial-to-trial variability were excluded from analysis. Lastly, while data were common average referenced in analysis, the data were re-preprocessed using a bipolar reference and any electrodes with a markedly different response when the referencing method was changed were excluded from analysis. All electrodes rejected through manual inspection of imaging were discussed and agreed upon by me and two of my coauthors (Alyssa M. Field and Liberty S. Hamilton).

3.4.5.1 Speech motor control task

A subset of six participants (TC6, DC7, DC10, DC13, DC16, DC17) completed a supplementary task with the goal of obtaining nonspeech oral motor movements to use as a control comparison for any electrodes that were production-selective to determine if they were speech-specific or not. Stimuli for this task consisted of written instructions accompanying a "go" signal on the iPad screen to prompt the participant to follow the instructions. The nine possible instructions, presented in a random order, were: "smile," "puff your cheeks," "open and close your mouth," "stick your tongue out," "move your tongue left and right," "tongue up (tongue to nose)," "tongue down (tongue to chin)," and "say 'aaaa'," "say 'oo-ee-oo-ee'." These instructions were chosen as a subset of movements evaluated during typical oral mechanism exams

conducted by speech-language pathologists (St. Louis & Ruscello 1981). Each movement was repeated 3 times.

For the nonspeech oral motor control task, except for the last two instructions (say "aa" or "oo-ee-oo-ee"), oral motor movements did not include an acoustic component. Thus, instead of being epoched to the acoustic onset of the trial like the primary task, responses were instead epoched to the display of the instruction text before the "go" signal, which was accompanied by the same broadband click tone as the main task. A match filter, identical to the one described above used to align high-resolution task audio with clinical recordings, identified the timing of these clicks and assisted in generation of the event files.

3.4.6 Event-related potential (ERP) analysis

I annotated accurate timing information for words, phonemes, and sentences to epoch data to differing levels of linguistic representation. A modified version of the Penn Phonetics Forced Aligner (Yuan & Liberman 2008) was used to automatically generate Praat TextGrids (Boersma 2002) using a transcript generated by the iPad log file. I checked the automatically generated TextGrids for accuracy myself, unlike in Chapter 2 where I had the assistance of undergraduate research volunteers. Event files containing start and stop times for each phoneme, word, and sentence, as well as information about trial type (perception vs. production), were created using the iPad log file and accuracy-checked TextGrids. These event files were then used to average

Z-scored high gamma across trials relative to sentence onset. For both production and perception, the onset of the sentence was treated as the acoustic onset of the first phoneme in the sentence as identified from the spectrogram. Responses were epoched between -0.5 and +2.0 seconds relative to sentence onset, with the negative window of interest intending to capture any pre-articulatory activity related to speech production (Chartier et al. 2018).

Electrode significance was determined by bootstrap t-test with 1000 iterations comparing activity during the stimulus to randomly selected inter-stimulus-interval activity; bootstrapped significance for perception and production activity were calculated separately as to identify electrodes that may be selectively responsive to either perceptual or production stimuli. A schematic is provided in 3.1. For the bootstrap procedure, I averaged activity 5-550 milliseconds after sentence onset and compared that to average activity during a silent 400-600 milliseconds after the inter-trial click as a control (Figure 3.1A). The control time window was selected as to not include potential evoked responses from the click sound but still be in the 1000 millisecond window between the click sound and stimulus presentation. A similar procedure was used to calculate significance for the consistent-inconsistent playback contrast (same time windows used). Bootstrap significance for the speech motor control task used activity 500-1000 milliseconds after the click sound played when text instructions were displayed to avoid including evoked responses to the click sound itself in the procedure (Figure 3.1B). Because there were no inter-trial click sounds in the speech motor control task with the click instead marking

the display of instructions, activity -500 to 0 milliseconds prior to the click sound was used as the control interval.

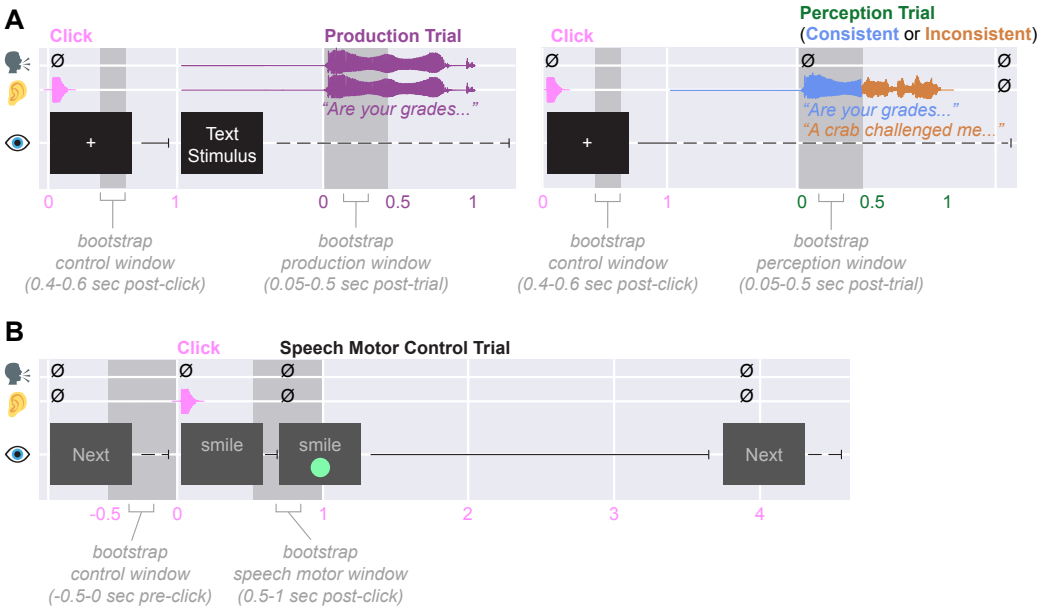

Figure 3.1: **Schematic of time windows for bootstrap t-tests.** (A) Schematic for bootstrapping during the speaking and listening task. Colors of X-axis values indicate time (in seconds) relative to the click sound (pink), production trial (purple), or perception trial (green). Rows represent information seen, heard, and spoken by the participant over the course of a trial. Shaded gray areas indicate time windows of high gamma activity compared during the bootstrap procedure. The perception trial waveform is split into two colors to indicate that the same windows of activity are used to calculate bootstrap significance for the consistent/inconsistent playback manipulation. (B) Schematic for bootstrapping during the speech motor control task. Different time windows are used for the bootstrap procedure due to the lack of inter-trial click sounds in the speech motor control task. The speech motor window is calculated relative to the click sound to capture any potential preparatory motor activity before the go signal (green circle).

In addition to suppression, I was interested to see how onset responses change between speaking and listening. To quantify the presence of an onset response at a particular electrode, I looked in the first 300 milliseconds of response relative to sentence onset for activity >1.5 SD above the mean response for the electrode's activity epoched to sentence onset. The time window of the onset response was defined as the range of contiguous samples of activity >1.5 SD above the mean, with the peak amplitude of the onset response being the greatest activity within the onset window. Onset latency was calculated as the maximum rate of change (differential) in the rising slope of the onset response. While I required an onset response to begin in the first 300 milliseconds of activity after sentence onset, I did not specify a time window in which one must end. Onset responses were quantified separately for the average production response and average perception response of each electrode. Electrodes that exhibited an onset response during speech perception and production were classified as "dual onset," while electrodes that exhibited an onset response during speech perception only were classified as "onset suppression."

3.4.7 Convex non-negative matrix factorization (cNMF)

To uncover patterns of evoked activity for speech production, speech perception, and auditory (click) perception that were consistent across participants, I employed convex non-negative matrix factorization (cNMF; Ding et al. (2010)). This is an unsupervised clustering technique that reveals underlying statistical structure of datasets and has previously been used by my

research group to discover profiles of neural response without explicitly specifying the feature represented by the response nor the anatomical location of the electrodes (Hamilton et al. 2018, 2021). I use a similar approach to these papers, summarized by the following equations:

$$X \approx \hat{X} = FG^{\top} \tag{3.1}$$

$$X_{p,n} \approx \frac{1}{t} \sum_{n=-1}^{\widehat{n=2}} H\gamma_{p,n} = FG^{\top} \tag{3.2}$$

where X is high gamma time series of shape (n samples, p electrodes) averaged across t epochs, and $F = XW$, where W is a matrix of shape (p electrodes, k clusters) and represents the cluster weights applied to the neural time series, and G is a matrix of shape (p electrodes, k clusters) and represents the weighting of an individual electrode within a cluster. cNMF was applied using this method to a concatenation of Z-scored evoked responses across subjects to sentences. Epochs consisted of a temporal range of -1 to +2 seconds relative to sentence onset. Epochs t were averaged within their response type then concatenated; possible response types were production onset, perception (playback) onset, and inter-trial click onset. This method of performing cNMF on averaged epochs across different types of trials has been utilized in prior intracranial studies of speech (Leonard et al. 2019). In a supplemental analysis, I concatenated additional epoch averages corresponding to presentation of

visual cues (e.g., text prior to reading, fixation cross) and a subdivision of playback onsets into consistent and inconsistent playback, but these manipulations did not significantly alter the clusters observed. I concatenated ERPs based on the response to production onset, perception (playback) onset, and click onset. I also incorporated information about expected vs. unexpected playback as well as presentation of the visual cue in separate supplemental analyses, but these did not significantly alter the clusters observed. The final concatenation resulted in a matrix X of $n * 3$ samples (production epochs, perception epochs, click epochs) by p electrodes. The number of basis functions to include was determined by two primary factors: first, the identification of a threshold such that adding additional clusters resulted in diminishing increases in percent variance explained; second, identifying a point at which adding additional clusters resulted in redundant average responses across clusters. I calculated percent variance as the coefficient of determination (R^2; Wright (1921)). This threshold was reached at k=9 clusters and 86% of the variance in the data explained. The average response for each of the k=9 clusters is provided in Figure C.1.

Inclusion in the cNMF analysis was determined using the bootstrap t-test described in 3.4.6. Electrodes above the significance threshold ($p > 0.05$) for both perception and production were excluded from cNMF clustering if the electrode also had a low correlation ($r < 0.1$) during the mTRF modeling procedure (§3.4.10). In other words: electrodes without a significant perception or production response to sentence onset nor a moderate performance during

mTRF model fitting were excluded from cNMF.

3.4.8 Suppression index (*SI*)

Within the sentence-onset epochs, a further window of interest was defined to calculate the degree of suppression between task conditions. The window of interest for onset responses was defined as 0 to 1 seconds after sentence onset. Window sizes were determined by previous research on onset and sustained responses (Hamilton et al. 2018) as well as preliminary results of the unsupervised clustering technique shown in Figure 3.4. The suppression index (SI), or degree of suppression during speaking as compared to listening, was quantified at each electrode as the ratio of high gamma activity between two separate conditions averaged across all epochs for the task condition occurring at that electrode. This is formalized as:

$$SI = \frac{H\gamma_L - H\gamma_S}{H\gamma_L + H\gamma_S} \tag{3.3}$$

where SI of electrode n is the difference of high gamma activity during speaking ($H\gamma_S$) subtracted from high gamma activity during listening ($H\gamma_L$) divided by the sum of high gamma activity during speaking and listening in the first 1 second after the acoustic onset of the sentence. A positive SI means that activity was greater during listening as compared to speaking, whereas a negative SI means activity was greater during speaking compared to listening. An SI of zero would reflect no difference between conditions.

3.4.9 Linear mixed-effects (LME) modeling

Linear mixed-effects (LME) models were fit using the package `lmertest` (Kuznetsova et al. 2017) in R at several points in analysis to quantify trends in the data. The approach bears similarities to the LME models fit in Chapter 2, but there are differences in their construction as the high spatial resolution of sEEG affords additional research questions which may be investigated with the technique. As with EEG, I chose LME as my statistical testing framework due to its ability to regress across within- and between-subject variability, facilitating generalization across subjects. The general equation takes the form, identical to Equation 2.1:

$$y = X\beta + Zu + \epsilon \tag{3.4}$$

where β represents fixed-effects parameters, u represents random effects, and ϵ error. Contrast significance for all LMEs described below is calculated using F tests with Kenward-Roger approximation with n degrees of freedom specified, where n is the length of matrix X (Kenward & Roger 1997).

The first LME reported in this chapter is used to quantify differences between suppression observed in onset and sustained responses. Suppression index (see §3.4.8) was used as the response variable with window of interest (two-way categorical: onset or sustained) and ROI as fixed effects and subject as a random effect (in R: `si` $\sim$ `window + roi + (1|subject)`). SI was calculated separately in the onset and sustained windows for this analysis,

unlike the SI formulation described in §3.4.8: onset SI was calculated between 0 and 750 milliseconds and sustained SI was calculated between 1000 and 1750 milliseconds after sentence onset. I chose these windows based on the average duration of the onset response across all electrodes and chose to make the sustained time window non-contiguous with the onset window to prevent extraneous activity from longer onset responses erroneously being factored as sustained activity in the model. I report the contrast in estimated marginal mean (EMM) SI of the two windows. I then used post-hoc Wilcoxon signed-rank tests with Benjamini-Yekutieli correction to calculate significant differences in SI between the onset and sustained responses within each ROI (Benjamini & Yekutieli 2001).

The second LME I report in this paper is used to quantify response latency within three regions of interest: primary auditory (HG, PT), non-primary auditory (STG, STS), and insular auditory (posterior + inferior insula). Peak latency values for the onset response (§3.4.6) are used as the response variable with ROI (three-way categorical) as a fixed effect and subject as a random effect (in R: `peak_latency` $\sim$ `roi + (1|subject))`. I report the EMM peak latencies of the three ROIs as well as their contrasts.

The third LME reported in this paper is used to quantify the mTRF (see §3.4.10) ablation analysis, a causal probing technique where specific stimulus features are added or removed from an encoding model and differences in performance are recorded (Ivanova et al. 2021a). For this LME model, the linear correlation coefficients (r) between $\widehat{H\gamma}$ and $H\gamma$ are used as the re-

sponse variable with model features (i.e., full vs. ablated) as a fixed effect and subject and channel as a random effect (in R: `r ~ model + (1|subject) + (1|channel)`). I chose to include channel as a random effect here as I did not have a specific hypothesis for anatomical differences in ablated model performance; additionally, including channel as a fixed effect instead would have resulted in an uninterpretable amount of pairwise comparisons and introduce multiple comparisons bias into the analysis. I report the *EMM* r values of the four models (base, ablate perception/production contrast, ablate consistent/inconsistent contrast, task-specific phonological feature encoding) as well as their contrasts.

3.4.10 Multivariate temporal receptive field (mTRF) modeling

Similar to Chapter 2, multivariate temporal receptive field (mTRF) models were fit to describe the selectivity of the high gamma response to different sets of stimulus features (Aertsen & Johannesma 1981; Crosse et al. 2016; Di Liberto et al. 2015; Theunissen et al. 2000). These models take the form of the equation below:

$$\hat{y}_n(t) = \sum_f \sum_{\tau=-0.3}^{\tau=0.5} w(f,\tau) S(f, t-\tau) + \epsilon \tag{3.5}$$

This equation is identical to Equation 2.2, but I will summarize it again here for the reader's convenience. $\hat{y}_n t$ represents the estimated high gamma signal at electrode n at time t. The stimulus matrix S consists of behavioral information regarding features (f) for each time point $t - \tau$, where τ is the

time delay between the stimulus and neural activity. I fit separate models to predict the high gamma response in each channel using time delays of -0.3 sec to 0.5 sec. This delay range encompasses the temporal integration times to similar responses found in previous research (Hamilton et al. 2018), but with an added negative delay to encompass potential pre-articulatory neural activity (Chartier et al. 2018; Kurteff et al. 2023). Data were split 80-20 into training and validation sets. To avoid overfitting, the data were segmented along sentence boundaries, such that the training and validation sets would not contain information from the same sentence. These segments were then randomly combined into the 80/20 training/validation sets. Weights for each feature and time delay $w(f, \tau)$ were fit using ridge regression on the training set and a regularization parameter chosen by 10 bootstrap iterations. The ridge parameter was selected at the value that provided the highest average correlation performance across all bootstraps. Ridge parameters between 10^2 and 10^8 were tested in 20 logarithmically scaled intervals. Model performance was assessed using correlations between the high gamma response predicted by the model and the true high gamma response. Significance of these correlations was obtained through a bootstrap t-test procedure with 100 iterations in which the training data were shuffled in chunks to remove the relationship between the stimulus and response.

3.5 Results

3.5.1 Onset responses are selectively suppressed during speech production

To examine potential differences in neural processing during speech production and perception, I acquired data from 17 pediatric, adolescent, and adult participants (9F, age 16.6 ± 6.4, range 8 to 37 years; Table D.2) surgically implanted with intracranial stereo-electroencephalography (sEEG) depth electrodes and pial electrocorticography (ECoG) grids for epilepsy monitoring. These patients performed a task where they read aloud naturalistic sentence stimuli then passively listened to playback of their reading (Figure 3.2A). For all analyses, I extracted the high gamma analytic amplitude of the local field potentials (Lachaux et al. 2012), which has been shown to correlate with single- and multi-unit neuronal firing (Ray & Maunsell 2011) and tracks both acoustic and phonological characteristics of speech (Mesgarani et al. 2014; Oganian et al. 2023). Based on prior work from my advisor, I expected to observe strong onset and sustained responses during sentence playback (Hamilton et al. 2018, 2021), as well as sensorimotor responses during the production portions of the task that would reflect articulatory control (Bouchard & Chang 2014; Chartier et al. 2018). Additionally, my task design allowed me to investigate the role of auditory-motor feedback during speech production by comparing neural responses to auditory feedback in real time to passive listening to an acoustically matched playback of each trial.

I recorded from a total of 2044 sEEG depth electrodes implanted in

perisylvian cortex and insula. This included coverage of speech responsive areas of the lateral superior temporal gyrus, but also within the depths of the superior temporal sulcus (STS), primary auditory cortex, and surrounding regions of the temporal plane. Within- and across-subject visualizations of electrode coverage are available as supplemental figures (Figures C.2, C.3). To examine differences between speech perception and production on individual electrodes, I plotted event-related high gamma responses for speech perception and production trials relative to the beginning of the acoustic onset of the sentence. I identified 144 electrodes with significant responses to perceptual stimuli, 350 electrodes with significant responses to production stimuli, and 110 electrodes with significant responses to both perceptual and production stimuli (Figure 3.2B; bootstrap t-test, $p < 0.05$). I quantified individual electrodes' selectivity to speech production or perception by calculating a suppression index (SI, Equation 3.3). An $SI > 0$ reflects higher activity during listening compared to speaking, and $SI < 0$ reflects higher activity during speaking compared to listening (Figure 3.2C).

Single-electrode responses can be visualized on an interactive 3D brain at `https://hamiltonlabut.github.io/kurteff2024/`[1]. I observed single electrodes with selective responses to speech perception in bilateral Heschl's gyrus and STG (Figure 3.2D). 51.4% of electrodes in STG ($n = 70$) and 100% of electrodes in Heschl's gyrus ($n = 13$) responded significantly to speech per-

[1]In the event this URL is taken offline in the future, please contact me and I will provide you with access.

ception stimuli. Response profiles of electrodes in this region consisted of a mixture of transient onset responses and lower-amplitude sustained responses during passive listening, consistent with previous research (Hamilton et al. 2018, 2021). In primary and non-primary auditory cortex, onset responses were notably absent during speech production, while sustained responses remained relatively un-suppressed ($Estimated\ marginal\ mean_{onset-sustained}\ SI = 0.153$; $p < .001$). Electrodes in primary sensorimotor cortex were typically more production-selective, in line with conventional localization of sensorimotor control of speech (Bouchard et al. 2013; Guenther 2016; Penfield & Roberts 1959). This pattern of responses demonstrates selective suppression of onset responses during speech production in primary and secondary auditory regions of the human brain. This result supports prior research that posits onset responses play a role in temporal parcellation of speech, a process unnecessary during speech production due to the speaker's knowledge of upcoming auditory information (Houde & Nagarajan 2011; Tourville & Guenther 2011).

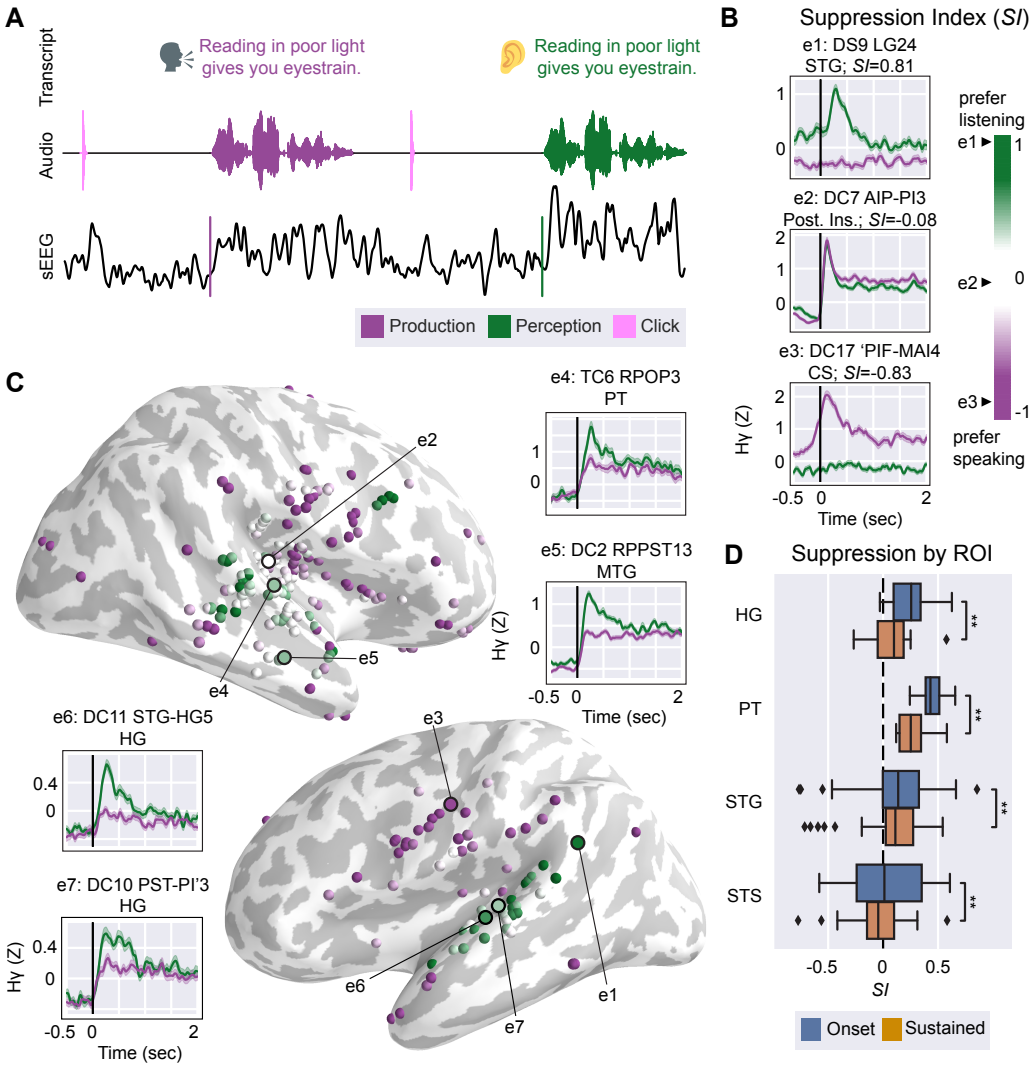
A
Transcript
Reading in poor light gives you eyestrain.
Reading in poor light gives you eyestrain.
Audio
sEEG
Production
Perception
Click
B
Suppression Index (SI)
e1: DS9 LG24
STG; SI=0.81
prefer listening
e1
e2: DC7 AIP-PI3
Post. Ins.; SI=-0.08
e2
e3: DC17 'PIF-MAI4
CS; SI=-0.83
e3
prefer speaking
Hγ (Z)
Time (sec)
C
e4: TC6 RPOP3
PT
e5: DC2 RPPST13
MTG
e6: DC11 STG-HG5
HG
e7: DC10 PST-PI'3
HG
e1
e2
e3
e4
e5
e6
e7
D
Suppression by ROI
HG
PT
STG
STS
SI
Onset
Sustained

Figure 3.2: **Auditory onset responses are suppressed during speech production**.
(A) Schematic of reading and listening task. Participants read a sentence aloud (purple) then passively listened to playback of themselves reading the sentence (green). Pink spikes in the beginning and middle of the audio waveform indicate inter-trial click tones, used as a cue and an auditory control.
(B) Single-electrode plots showing different profiles of response selectivity across the cortex. Color gradient represents normalized SI values. A more positive SI indicates an electrode is more responsive to speech perception stimuli (e1) while a more negative SI means an electrode is more responsive to production stimuli (e3). e2 and e3 are examples of response profiles described in subsequent figures (Figures 3.3 and 3.4, respectively). Example electrodes' SI are indicated on the gradient. Subplot titles reflect the participant ID and electrode name from the clinical montage.
(C) Whole-brain and single-electrode visualizations of perception and production selectivity (SI). Electrodes are plotted on a template brain with an inflated cortical surface; dark gray indicates sulci while light gray indicates gyri. Single-electrode plots of high-gamma activity demonstrate suppression of onset response relative to the acoustic onset of the sentence (vertical black line).
(D) Box plot of suppression index during onset (blue) and sustained (orange) time windows separated by anatomical region of interest in primary and non-primary auditory cortex. Brackets indicate significance (* = $p < 0.05$; ** = $p < 0.01$).
Abbreviations: HG: Heschl's gyrus; PT: planum temporale; STG: superior temporal gyrus; STS: superior temporal sulcus; MTG: middle temporal gyrus; CS: central sulcus; Post. Ins.: posterior insula.

3.5.2 The posterior insula uniquely exhibits onset responses to speaking and listening

The ability of sEEG to obtain high-resolution recordings of human insula is a unique strength, as other intracranial approaches such as ECoG grids and electrocortical stimulation cannot be applied to the insula without prior

dissection of the Sylvian fissure, an involved and rarely performed surgical procedure (Remedios et al. 2009; Zhang et al. 2018). Similarly, hemodynamic and lesion-based analyses may suffer from vasculature-related confounds in isolating insular responses (Hillis et al. 2004). Here I present high spatiotemporal resolution recordings from human insula and identify a functional response profile localized to this region.

While onset responses to speech perception were mostly confined to auditory cortex, a functional region of interest in the posterior insula demonstrated a different morphology of onset responses. Across participants, electrodes in the posterior insula showed robust onset responses to perceptual stimuli in similar fashion to auditory electrodes. Unlike auditory electrodes, however, posterior insular electrodes also showed robust onset responses during speech production (Figure 3.3D). Out of all posterior insula electrodes ($n = 47$), 23.4% responded significantly to speech perception and 31.9% responded significantly to speech production. These posterior insula onset electrodes responded similarly to stimuli regardless of whether they were spoken or heard (Figure 3.3).

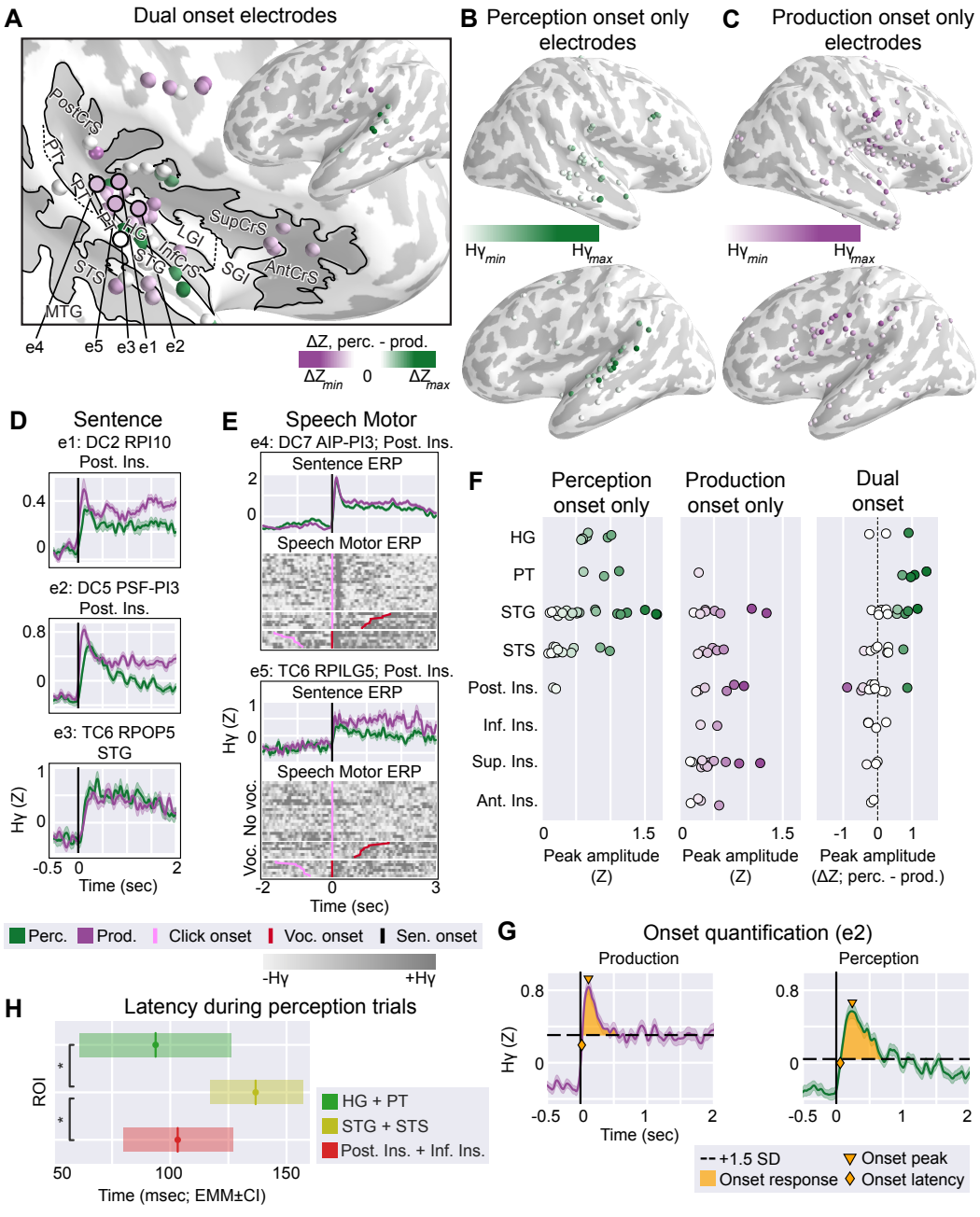
A
Dual onset electrodes
PostCrS
PT
HG
LGI
SupCrS
InfCrS
SGI
AntCrS
STS
STG
MTG
e4
e5
e3
e1
e2
ΔZ, perc. - prod.
ΔZmin
0
ΔZmax
B Perception onset only electrodes
C Production onset only electrodes
Hγmin
Hγmax
D Sentence
e1: DC2 RPI10 Post. Ins.
e2: DC5 PSF-PI3 Post. Ins.
e3: TC6 RPOP5 STG
Hγ (Z)
Time (sec)
E Speech Motor
e4: DC7 AIP-PI3; Post. Ins.
Sentence ERP
Speech Motor ERP
e5: TC6 RPILG5; Post. Ins.
Voc. No voc.
Time (sec)
F
Perception onset only
Production onset only
Dual onset
HG
PT
STG
STS
Post. Ins.
Inf. Ins.
Sup. Ins.
Ant. Ins.
Peak amplitude (Z)
Peak amplitude (ΔZ; perc. - prod.)
Perc.
Prod.
Click onset
Voc. onset
Sen. onset
-Hγ
+Hγ
G
Onset quantification (e2)
Production
Perception
Hγ (Z)
Time (sec)
+1.5 SD
Onset response
Onset peak
Onset latency
H
Latency during perception trials
ROI
HG + PT
STG + STS
Post. Ins. + Inf. Ins.
Time (msec; EMM±CI)

Figure 3.3: **A functional region of interest in posterior insula shows onset responses to both speaking and listening**.
(A) Whole-brain and visualization of dual onset electrodes. Electrodes are plotted on a template brain with an inflated cortical surface; dark gray indicates sulci while light gray indicates gyri. Black outline on template brain highlights functional region of interest in posterior insula with anatomical structures labeled. Electrode color indicates the difference in Z-scored high gamma peaks during the speaking and listening conditions (ΔZ). Right hemisphere is cropped to emphasize insula ROI, while left hemisphere is shown in entirety due to lower number of electrodes.
(B) Whole-brain visualization of electrodes with onset responses only during speech perception. Electrode color indicates the peak high gamma amplitude during the onset response.
(C) Whole-brain visualization of electrodes with onset responses only during speech production. Electrode color indicates the peak high gamma amplitude during the onset response.
(D) Single electrode activity from posterior insular electrodes highlighting dual onset responses during speech production and perception. Vertical black line indicates acoustic onset of sentence. Subplot titles reflect the participant ID, electrode name from the clinical montage, and anatomical ROI.
(E) Grayscale heatmaps of single-trial electrode activity during a nonspeech motor control task, separated by no vocalization (e.g., "stick your tongue out") and vocalization (e.g., "say 'aaaa' "). For vocalization trials, onset of acoustic activity is visualized relative to the click accompanying the presentation of instructions (pink) and the onset of vocalization (red).
(F) Strip plot showing the distribution of channel-by-channel onset response peak amplitudes separated by anatomical region of interest and whether onset responses occur only during perception (left), only during production (center), or occur during perception and production (right). Electrodes are colored according to the colormaps of (A), (B), and (C).
(G) Schematic of quantification of onset response for an example electrode (e2, DC5 PSF-PI3). The first contiguous peak of activity > 1.5 SD above the mean response constitutes the onset response and is shaded in orange. Peak amplitude values displayed in (B), (C) and (G) are indicated.
Caption continued on next page.

Figure 3.3: (H) Bar plot showing the estimated marginal mean (*EMM*) latency of the onset response in three regions of interest: auditory primary (HG + PT), auditory non-primary (STG + STS), and posterior + inferior insular. Insular onset latency is comparable to primary auditory latency. Brackets indicate significance (* = $p < 0.05$; ** = $p < 0.01$).
Abbreviations: HG: Heschl's gyrus; STG: superior temporal gyrus; STS: superior temporal sulcus; MTG: middle temporal gyrus; Inf/Sup/Ant/Post/ CrS: inferior/superior/anterior/posterior circular sulcus of the insula; LGI: long gyrus of the insula; SGI: short gyrus of the insula; PT: planum temporale.

I hypothesized that such responses might reflect a relationship to articulatory motor control or somatosensory processes, which prompted me to trial a nonspeech motor control task in a subset of the participants ($n = 6$; §3.4.5.1, Table D.2). The purpose of this task was to determine if such "dual onset" responses were speech-specific or whether they could be elicited by simpler, speech-related movements. In this task, participants were instructed to follow instructions displayed on screen when a "go" signal was given; the instructions consisted of a variety of nonspeech oral-motor tasks taken from a typical battery used by speech-language pathologists during oral mechanism evaluations (St. Louis & Ruscello 1981). The "go" signal contained both a visual (green circle) and an auditory cue (click), after which the participant would perform the task. Some tasks required vocalization (e.g., "say 'aaaa' ") while others did not (e.g., "stick your tongue out"). While a few insular electrodes did exhibit responses during the speech motor control task, they were not consistently responsive to the speech motor control task except for trials that involved auditory feedback (Figure 3.3E). I interpret these as responses to the click sound when instructions are displayed to the participant

or to the subjects' own vocalizations rather than an index of sensorimotor activity related to the motor movements. When significance is calculated in a time window that excludes the click sound (500-100 milliseconds post-click), only 2% of insula electrodes ($n = 49$) significantly responded to the speech motor control task. By comparison, 25.7% of sensorimotor cortex electrodes ($n = 35$) significantly responded, demonstrating that the speech motor control task was sensitive to sensorimotor activity. Additionally, posterior insular electrodes that were responsive to the speech motor control task and all dual onset insular electrodes in the main task were only active after the onset of articulation. This later response suggests that these electrodes were involved in sensory feedback processing and not direct motor control. The posterior insula region of interest was the only anatomical area in my dataset that was equally responsive to acoustic onsets during both production and perception. While electrodes with dual onset responses during speaking and listening were seen in both primary/secondary auditory areas (22.7% of dual onset electrodes) and the insula (28.8% of dual onset electrodes), electrodes with similar amplitudes for speaking and listening were most common in posterior insula (Figure 3.3A, F). In other words, while temporal electrodes did sometimes demonstrate dual onset responses, the amplitudes of these responses were larger for speech perception compared to production. I quantified this restriction of "dual onset" electrodes to posterior insula by taking the peak amplitude in the first 300 milliseconds of activity prior to sentence onset greater than 1.5 SD above the epoch mean as a measure of the onset response (Figure 3.3G).

The response latencies of different anatomical regions can provide a proxy for understanding how information flows from one region to another, or where in the pathway a certain response may occur. For example, my advisor's prior work showed similar latencies between the pSTG and posteromedial Heschl's gyrus, indicating a potential parallel pathway (Hamilton et al. 2021). Here, the dual onset electrodes in posterior insula responded with comparable latency to the speech perception onset response electrodes observed in primary (HG & PT) and non-primary auditory cortex (STG & STS), in some cases responding earlier relative to sentence onset than the auditory cortex electrodes (EMM_{A1} peak latency = 93.7 ± 16.2 msec; $EMM_{Aud.\ non-primary}$ peak latency = 136.7 ± 9.4 msec; $EMM_{insular}$ peak latency = 103.2 ± 11.7 msec; *A1-Aud. non-primary* $p = 0.03$; *A1-insular* $p = 0.85$; *Aud. non-primary-insular* $p = 0.03$; Figure 3.3H). This does not suggest a conventionally proposed serial cascade of information from primary auditory cortex and is instead indicative of a parallel information flow to primary auditory cortex and the posterior insula, potentially from the terminus of the ascending auditory pathway. The similar latency of posterior insular dual onset electrodes and primary auditory onset suppression electrodes alongside the tendency of posterior insular electrodes to also show low-latency onset responses during speech production leads me to speculate that the posterior insula receives a parallel thalamic input and serves as a sensory integration hub for the purposes of feedback processing during speech.

3.5.3 Unsupervised identification of "onset suppression" and "dual onset" functional response profiles

Visualization of individual electrodes' responses to the onset of perceived and produced sentences allows for manual identification of response profiles in the data but is subject to *a priori* bias by the investigators. Data driven methods such as convex non-negative matrix factorization (cNMF) allow identification of patterns in the data without access to spatial information or the acoustic content of the stimuli (§3.4.7; Ding et al. (2010)). This method was used to identify onset and sustained responses in STG (Hamilton et al. 2018). Here, I used cNMF to identify response profiles in the data in an unsupervised fashion using average evoked responses as the input to the factorization. A solution with $k = 9$ clusters explained 86% of the variance in the data (Figure 3.4A). I chose this threshold as increasing the number of clusters in the factorization beyond $k = 9$ resulted in redundant clusters. Single-electrode responses to spoken sentences, perceived sentences, and an inter-trial click tone were used as inputs to the factorization such that responses to each of these conditions were jointly considered for defining a "cluster." The average responses of all top-weighted electrodes within cluster for the $k = 9$ factorization is available as a supplemental figure (Figure C.1). Visualization of the average response across sentences of the top-weighted electrodes within each cluster identifies two primary response profiles in correspondence with manually identified response profiles: (c1) an "onset suppression" cluster localized to bilateral STG and Heschl's gyrus characterized by evoked responses

to speech production and speech perception but an absence of onset responses during speech production; and (c2) a "dual onset" cluster localized to the posterior insula/circular sulcus characterized by evoked responses to the onset of perceived and produced sentences (Figure 3.4B, C). An additional cluster (c3) was localized to ventral sensorimotor cortex and showed selectivity to speech production trials, particularly prior to articulation. This cluster is located in ventral sensorimotor cortex, and likely reflects motor control of speech articulators (Bouchard et al. 2013; Breshears et al. 2015; Dichter et al. 2018).

Because the onset suppression and dual onset clusters are relatively close to each other anatomically, I quantified their functional separation by examining whether individual electrodes contributed strong weighting to both clusters. I observed that despite the spatial proximity of the clusters[2], the majority of electrodes in both onset suppression and dual onset clusters were only strongly weighted within a single cluster (Figure 3.4D). The top 50 electrodes of the onset suppression contributed 86.5% of their weighting to the onset suppression cluster and 13.5% to the dual onset cluster, while the top 50 electrodes of the dual onset cluster contributed 88.8% to the dual onset cluster and 11.2% to the onset suppression cluster (Figure 3.4E). This suggests that despite anatomical proximity, the onset responses in posterior insular electrodes are not the result of spatial spread of activity from nearby primary auditory electrodes in Heschl's gyrus and planum temporale. Taken

[2]Notably, cNMF does not have information about anatomical location of these electrodes when clustering the responses.

together, the supervised and unsupervised analyses suggest auditory feedback is processed differently by two regions in temporal and insular cortex. Auditory cortex suppresses responses to self-generated speech through attenuation of the onset response, while the posterior insula uniquely responds to onsets of auditory feedback regardless of whether the stimulus was self-generated or passively perceived.

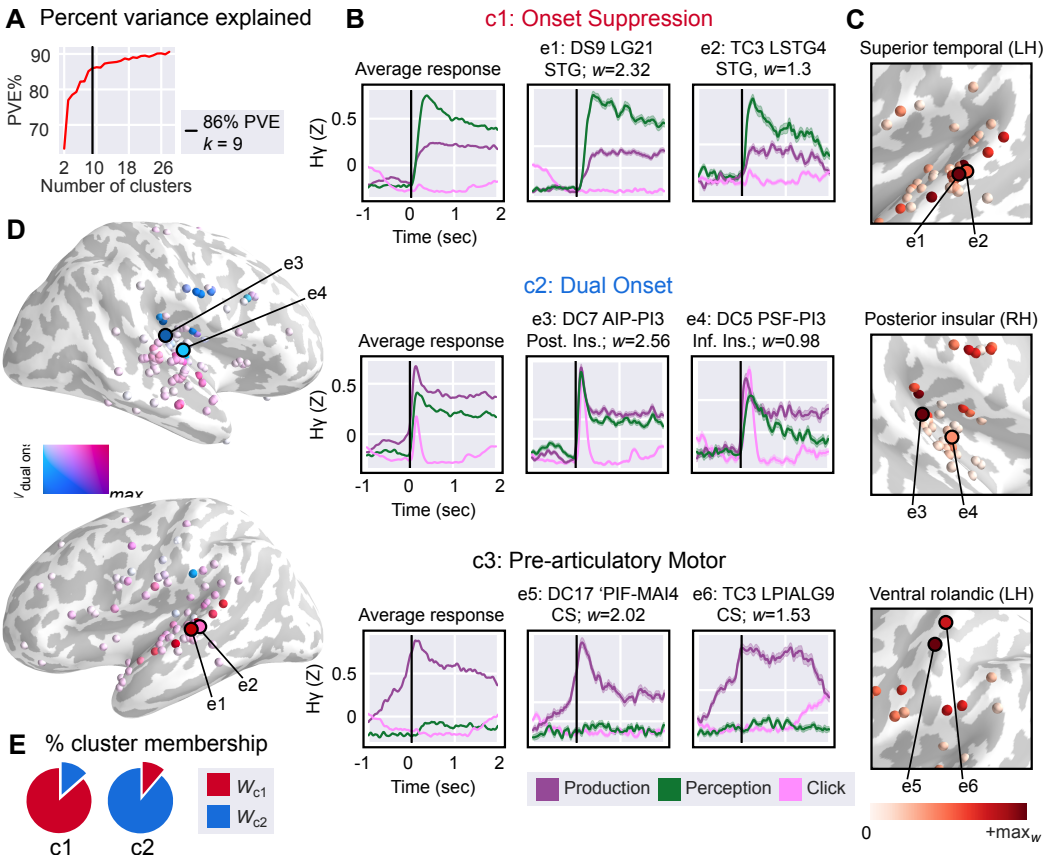

Figure 3.4: **Anatomically distinct onset suppression and dual onset clusters represent a subclass of response profiles to continuous speech production and perception**.
(A) Percent variance explained by cNMF as a function of total number of clusters in factorization. Threshold of $k = 9$ factorization plotted as vertical black line.
(B) cNMF identifies three response profiles of interest: (c1) onset suppression electrodes, characterized by a suppression of onset responses during speech production and localized to STG/HG; (c2) dual onset electrodes, characterized by the presence of onset responses during perception and production and localized to posterior insula; (c3) pre-articulatory motor electrodes, characterized by activity prior to acoustic onset of stimulus during speech production and localized to ventral sensorimotor cortex. Left: Cluster basis functions for speaking sentences (purple), listening to sentences (green), and inter-trial click (pink) for c1, c2, and c3. Center, right: Two example electrodes from the top 16 weighted electrodes. Subplot titles reflect the participant ID and electrode name from the clinical montage.
(C) Cropped template brain showing top 50 weighted electrodes for individual clusters (c1, c2, c3). A darker red electrode indicates higher within-cluster weight.
(D) Individual electrode contribution to dual onset and onset suppression cNMF clusters in both hemispheres. Top 50 weighted electrodes for each cluster are plotted on a template brain with an inflated cortical surface; dark gray indicates sulci while light gray indicates gyri. Red electrodes contribute more weight to the "onset suppression" cluster while blue electrodes contribute more to the "dual onset" cluster; purple electrodes contribute equally to both clusters while white electrodes contribute to neither.
(E) Percent similarity of onset suppression (c1) and dual onset (c2) clusters' top 50 electrodes. The majority of the electrode weighting across these two clusters is non-overlapping.
Abbreviations: STG: superior temporal gyrus; CS: central sulcus. Inf. Ins. = inferior insula, Post. Ins = posterior insula.

3.5.4 Response to playback consistency is a separate mechanism from suppression of onset responses

Speaker-induced suppression of self-generated auditory feedback is one example of how top-down information can influence auditory processing. In rodent studies, animals can learn to associate a particular tone frequency with self-generated movements, and motor-related auditory suppression will occur specifically for that frequency rather than unexpected frequencies that were not paired with movement (Schneider et al. 2018). Expectations about upcoming auditory feedback can also influence the outcomes of feedback perturbation tasks in humans (§1.2.3.3; Lester-Smith et al. (2020); Scheerer & Jones (2014)). I was interested if other top-down expectations about the task could affect the responses of electrodes in these data and if these populations overlapped with speaker-induced suppression. To accomplish this, I separated the playback condition into blocks of consistent and inconsistent playback (Figure 3.5A). In the consistent playback block, participants were always played back the sentence they had just produced in the prior speaking trial. In the inconsistent playback block, participants instead were played back a randomly selected recording of a previous speaking trial. In both cases, the playback stimulus was a recording of their own voice.

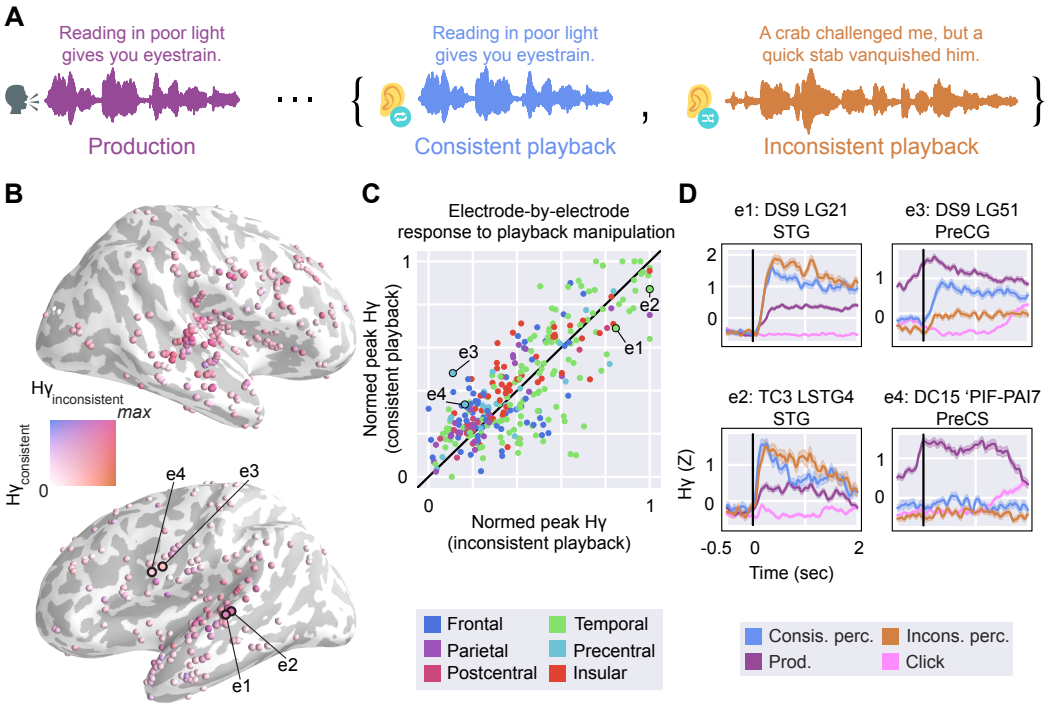

Figure 3.5: **Playback consistency manipulation yields separate, weaker effects than onset suppression**.
(A) Task schematic showing playback consistency manipulation. Participants read a sentence aloud (purple) then passively listened to playback of that sentence (blue) or randomly selected playback of a previous trial (orange).
(B) Whole-brain visualization of responsiveness to playback consistency. Electrodes are plotted on an inflated template brain; dark gray indicates sulci while light gray indicates gyri. Electrodes are colored using a 2D colormap that represents high gamma amplitude during consistent and inconsistent playback; blue indicates a response during consistent playback but not during inconsistent, orange indicates a response during inconsistent playback but not during consistent playback, pink indicates a response to both playback conditions, white indicates a response to neither. Most electrodes are pink, indicating strong responses to both conditions. Example electrodes from (D) are indicated.
Caption continued on next page.

Figure 3.5: (C) Scatter plot of channel-by-channel peak high-gamma activity during consistent playback (Y-axis) and inconsistent playback (X-axis). Vertical black line indicates unity. Color corresponds to gross anatomical region. Example electrodes from (D) are indicated.
(D) Single-electrode plots of high-gamma activity relative to sentence onset (vertical black line). Left column (e1 and e2): Electrodes in temporal cortex demonstrating a slight preference for inconsistent playback. Right column (e3 and e4): Electrodes in frontal cortex demonstrating a slight preference for consistent playback and a larger preference for speech production trials.
Abbreviations: HG: Heschl's gyrus; STG: superior temporal gyrus; PreCS: precentral sulcus; Supramar: supramarginal gyrus.

The majority of electrodes did not differentially respond to consistent or inconsistent playback conditions (pink-red electrodes in Figure 3.5B; electrodes along unity line in Figure 3.5C). While 45.5% of STG electrodes ($n = 55$) were significantly responsive to both consistent and inconsistent playback, only 5.5% were responsive solely during consistent playback and 0% were responsive solely during inconsistent playback. Other auditory areas showed a similar trend, including STS (both = 20.3%; consistent only = 4.3%; inconsistent only = 2.9%; $n = 69$ electrodes), posterior insula (both = 15.4%; Consistent only = 2.6%; Inconsistent only = 0%; $n = 39$ electrodes), and HG (both = 100%; Consistent only = 0%; inconsistent only = 0%; $n = 8$ electrodes). For the subset of electrodes that did differentially respond, most demonstrated a slight amplitude increase during the inconsistent playback condition that started at the time of the onset response and persisted throughout stimulus presentation (Figure 3.5D). Electrodes that selectively responded to inconsistent stimuli did not have an identifiable general response profile. Most electrodes that showed a

preference for inconsistent playback also demonstrated onset suppression during speech production trials (e3 & e4, Figure 3.5D), but this suppression was far stronger than any difference between consistent and inconsistent playback. A contrast between consistent and inconsistent playback was most commonly observed in superior temporal gyrus and superior temporal sulcus. Curiously, a subset of electrodes localized to ventral sensorimotor cortex (similarly to cluster c3 presented in Figure 3.4B) showed an overall preference for speech production trials with pre-articulatory activity, but within the playback contrast demonstrated a preference for consistent playback (e5 & e6, Figure 3.5D). I interpret this finding as a speech motor region that indexes predictions of upcoming sensory content for a role in feedback control.

3.5.5 Despite suppression of onset responses, phonological feature representation is suppressed but stable between perception and production

Prior work shows that circuits within the STG represent phonological feature information that is invariant to other acoustic characteristics such as pitch (Appelbaum 1996; Mesgarani et al. 2014; Tang et al. 2017). Tuning for these phonological features is observed within both posterior onset selective areas of STG and anterior sustained regions (Hamilton et al. 2018). Here, I observed that onset responses are suppressed during speech production, which motivates investigating whether phonological feature tuning is also modulated as part of the auditory system's differential processing of auditory information while speaking. To investigate this, I fit multivariate temporal receptive

fields (mTRF; §3.4.9) for each electrode to describe the relationship between the neural response at that electrode and selected phonological and task-level features of the stimulus (Figure 3.6A). I report the effectiveness of an mTRF model in predicting the neural response as the linear correlation coefficient (r) between a held-out validation response and the predicted response based on the model (Figure 3.6B, C).

Onset suppression electrodes in auditory cortex and dual onset electrodes in the posterior insula were both well modeled using this approach ($\bar{x}_{onset\ suppression\ electrodes} = 0.17 \pm 0.08$; $\bar{x}_{dual\ onset\ electrodes} = 0.16 \pm 0.11$; range -0.25 to 0.64; Figure 3.6D). Within both response profiles, single electrodes exhibited a diversity of preferences to various combinations of phonological features, mirroring previous results showing distributed phonological feature tuning in auditory cortex (Berezutskaya et al. 2017; Hamilton et al. 2018, 2021; Mesgarani et al. 2014; Oganian & Chang 2019). Of note, the posterior and inferior insula electrodes were strongly phonologically tuned, with a short temporal response profile as was seen in my prior latency analysis (§3.5.2). Dual onset and onset suppression electrodes differed from purely production-selective electrodes in this way, as most production-selective electrodes qualitatively did not demonstrate robust phonological feature tuning. Instead, most of the variance in the mTRF instead was explainable by global task-related stimulus features (i.e., whether a sound occurred during a production or a perception trial).

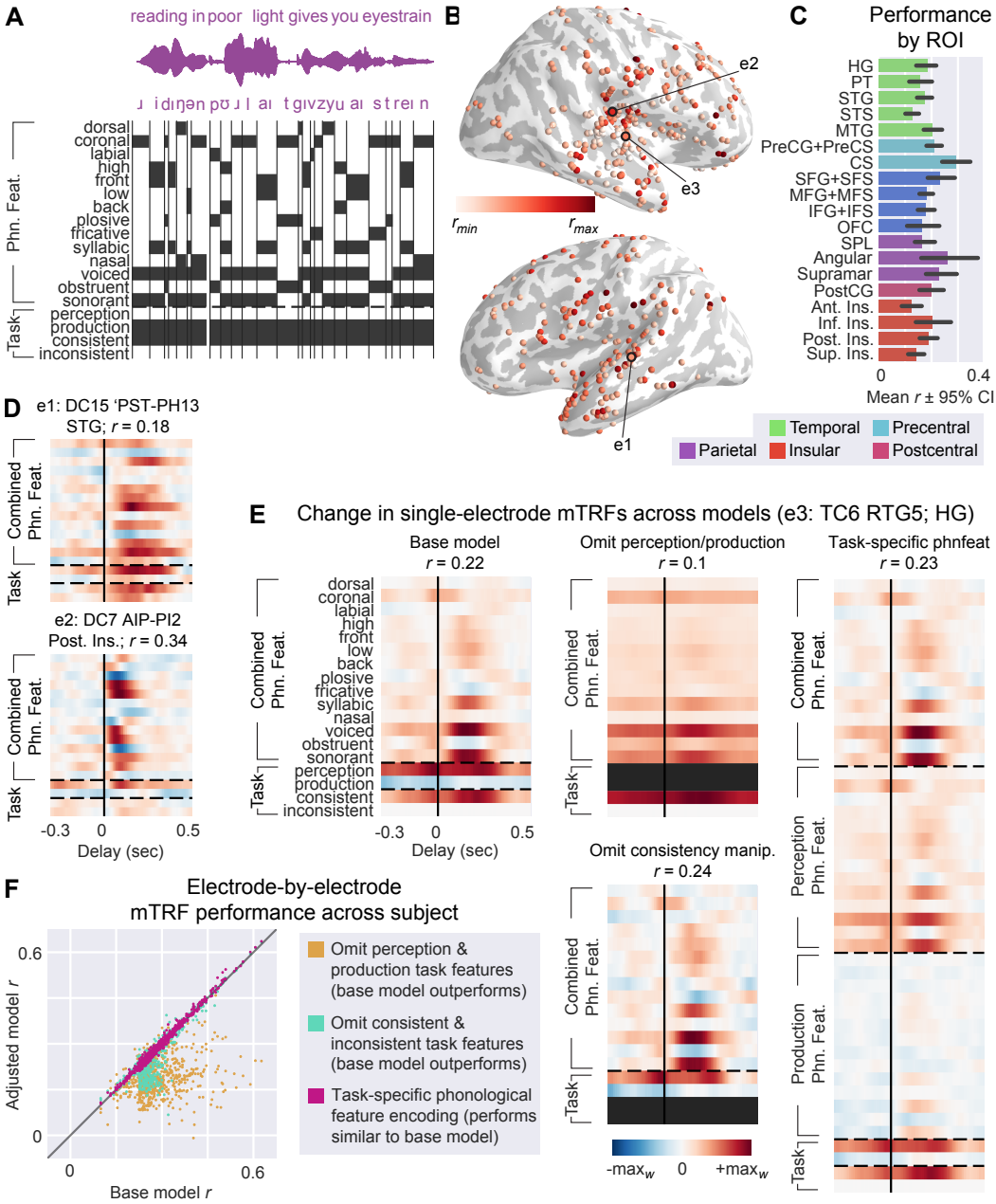
A
reading in poor light gives you eyestrain
ɹ ɪdɪŋən pʊɹ l aɪ tgɪvzyu aɪ streɪ n
Phn. Feat.
Task
dorsal
coronal
labial
high
front
low
back
plosive
fricative
syllabic
nasal
voiced
obstruent
sonorant
perception
production
consistent
inconsistent
B
e2
e3
e1
r_{min}
r_{max}
C
Performance by ROI
HG
PT
STG
STS
MTG
PreCG+PreCS
CS
SFG+SFS
MFG+MFS
IFG+IFS
OFC
SPL
Angular
Supramar
PostCG
Ant. Ins.
Inf. Ins.
Post. Ins.
Sup. Ins.
0
0.4
Mean r ± 95% CI
Parietal
Temporal
Insular
Precentral
Postcentral
D
e1: DC15 'PST-PH13
STG; r = 0.18
Combined Phn. Feat.
Task
e2: DC7 AIP-PI2
Post. Ins.; r = 0.34
-0.3
0
0.5
Delay (sec)
E
Change in single-electrode mTRFs across models (e3: TC6 RTG5; HG)
Base model
r = 0.22
Omit perception/production
r = 0.1
Task-specific phnfeat
r = 0.23
Omit consistency manip.
r = 0.24
Perception Phn. Feat.
Production Phn. Feat.
-max$_w$
0
+max$_w$
F
Electrode-by-electrode
mTRF performance across subject
Adjusted model r
Base model r
0.6
0
Omit perception & production task features (base model outperforms)
Omit consistent & inconsistent task features (base model outperforms)
Task-specific phonological feature encoding (performs similar to base model)

Figure 3.6: **Phonological feature tuning is stable during speaking and listening across brain regions**.
(A) Regression schematic. Fourteen phonological features corresponding to place of articulation, manner of articulation, and presence of voicing alongside four features encoding task-specific information (i.e., whether a phoneme took place during a speaking or listening trial, the playback condition during the phoneme) were binarized sample-by-sample to form a stimulus matrix for use in temporal receptive field modeling.
(B) Model performance as measured by the linear correlation coefficient (r) between the model's prediction of the held-out sEEG and the actual response plotted at an individual electrode level on an inflated template brain; dark gray indicates sulci while light gray indicates gyri. Example electrodes from (D) and (E) are indicated.
(C) Model performance by region of interest. Color corresponds to gross anatomical region.
(D) Temporal receptive fields of two example electrodes in temporal and insular cortex.
(E) Temporal receptive fields of an example electrode for the four models presented in (F).
(F) Scatter plot of channel-by-channel linear correlation coefficients (r) colored by model comparison. The X-axis shows performance for the "base" model whose schematic is presented in (A). The Y-axis for each scatterplot shows performance for a modified version of the base model: task features encoding production and perception were removed from the model (yellow); task features encoding consistent and inconsistent playback conditions were removed from the model (cyan); phonological features were separated into production-specific, perception-specific, and combined spaces (magenta).
Abbreviations: HG: Heschl's gyrus; PT: planum temporale; STG/S: superior temporal gyrus/sulcus; MTG/S: middle temporal gyrus/sulcus; PreCG/S: precentral gyrus/sulcus; CS: central sulcus; SFG/S: superior frontal gyrus/sulcus; MFG/S: middle frontal gyrus/sulcus; IFG/S: inferior frontal gyrus/sulcus; OFC: orbitofrontal cortex; SPL: superior parietal lobule; PostCG: postcentral gyrus; Ant./Post./Sup./Inf. Ins.: anterior/posterior/superior/inferior insula.

To directly compare phonological feature representations during perception and production, I used a variance partitioning technique similar to

the one used in §2.5.2 to omit or include specific stimulus features in the model. In this way, the stimulus matrix serves as a hypothesis about what stimulus characteristics will be important in modeling the neural response. Adding or removing individual stimulus characteristics and observing differences (or lack thereof) in model performance serves as a causal technique for assessing the importance of a stimulus characteristic to the variance of an electrode's response (Ivanova et al. 2021a). In the base model, I included 14 phonological features and 4 task-related features, similar to the base model of Chapter 2 but without an EMG regressor. I first expanded the specificity of phonological feature tuning in my stimulus matrix by separating the phonological feature space into whether the phonemes in question occurred during perception or production (called the "task-specific" model). If phonological feature tuning differed during speech production, model performance should increase when modeling perceived vs. produced phonological features separately. However, I saw no significant increase in model performance when expanding the model in this way (Figure 3.6F, pink points), a result also observed in my EEG data (Figures 2.3, 2.4). Despite no gross difference in model performance, inspection of individual electrodes' receptive fields shows a suppression in the weights for production-specific phonological feature tuning (Figure 5E, far right), again similar to EEG (Figure 2.5). The contrast between "base" and "task-specific" model performance, while significant, was a weak effect in favor of the simpler "task-specific" model ($EMM_{base-task-specific\ phnfeat}$ $\Delta r = -0.002$, $p = 0.04$, $d = -0.1$). In con-

trast, removal of the playback consistency information from the task-specific portion of the stimulus matrix more substantially affects model performance ($EMM_{base-omit\ consistent/inconsistent}$ $\Delta r = 0.02$, $p < .001$, $d = 0.52$). However, the most drastic impairment of model performance emerges when removing information about the contrast of perception and production trials entirely from the model ($EMM_{base-omit\ perception/production}$ $\Delta r = 0.11$, $p < .001$, $d = 1.4$). Upon inspection, the regions exhibiting the largest decline in encoding performance with the omission of the perception-production contrast are frontal production-responsive regions and temporal onset suppression regions, whereas insular electrodes did not see as steep a decline in performance. This suggests that differences in encoding during speech production and perception are the primary explanation of variance in these models. Ultimately, despite onset suppression seen during speech production, higher-order linguistic representations such as phonological features appear to be stable during speech perception and production.

Taken together, these results provide an expanded perspective on how auditory areas of the brain differentially process sensory information during speech production and perception. Transient responses to acoustic onsets in primary and higher order auditory areas are suppressed during speech production, whereas responses of these regions not at acoustic onset remain relatively stable between perception and production. This onset suppression can be seen in the neural time series and is also reflected in the encoding of linguistic information in temporal receptive field models. It is thus possible that the onset

response functions as a stimulus orientation mechanism rather than a higher-order aspect of the perceptual system such as phonological encoding. While expectations about the linguistic content of upcoming auditory playback can influence response profiles, the mechanism appears separate from the suppression of onset responses and is a relatively weak effect by comparison. Lastly, these results provide a unique perspective on the role of the posterior insula during speaking and listening, characterized by its rapid responses to speech production and perception stimuli and phonological tuning without the suppression observed during speech production in nearby temporal areas.

3.6 Conclusion

In this chapter, I used a sentence reading and playback task that allowed me to compare mechanisms of auditory perception and production while controlling for stimulus acoustics. The primary objective was to assess spatiotemporal differences in previously identified onset and sustained response profiles in the auditory cortex (Hamilton et al. 2018) and phonological feature encoding (Mesgarani et al. 2014) during speech production. Using sEEG has the distinct advantage of penetrating into deeper structures inside the Sylvian fissure, such as the insula and Heschl's gyrus (Chang 2015). In temporal cortex, proximal to where onset responses have been previously identified using surface electrocorticography (Hamilton et al. 2018), I observed a selective suppression of transient responses to sentence onset during speech production, whereas sustained responses remained relatively unchanged between speech percep-

tion and production (§3.5.1). The timing of the suppressed onset responses is roughly aligned with scalp-based studies of speaker-induced suppression that posit early components (N1 for EEG, M1 for MEG) as biomarkers of speaker-induced suppression (Chapter 2; Hawco et al. (2009); Heinks-Maldonado et al. (2006); Kurteff et al. (2023); Martikainen et al. (2005)). While I do not claim the onset responses observed in this study and others to be equivalent to N/M1, there is a parallel to be drawn between the temporal characteristics of my suppressed cortical activity and the deep literature on suppression of these components during speech production in noninvasive studies. I will expand upon this more in Chapter 4.

Overall, this study gives clarity to both the differential processing of the auditory system during speech production and the functional role of onset responses as a temporal landmark detection mechanism through high-resolution intracranial recordings of a naturalistic speech production and perception task. To be specific, the suppression of onset responses during speech production lends to the hypothesis that onset responses are an orientational mechanism. Feedforward expectations about upcoming sensory feedback during speech production would nullify the need for temporal landmark detection to the same extent necessary during speech perception, where expectations about incoming sensory content are much less precise. This raises questions about the function of onset responses in populations with disordered feedforward/feedback control systems (§1.3.1), such as apraxia of speech (Jacks & Haley 2015), schizophrenia (Heinks-Maldonado et al. 2007), and stuttering (Max & Daliri 2019; Toyomura

et al. 2020). The presence or absence of onset responses having no effect on the structure of phonological feature representations also supports this hypothesis, as linguistic abstraction is a higher-level perceptual mechanism that need not be implicated in lower-level processing of the auditory system. In future work, I would like to further investigate the role of onset responses in less typical speech production. Just as self-generated speech is less suppressed during errors (Ozker et al. 2022, 2024) and less canonical utterances (Niziolek et al. 2013), the landmark detection services of the onset response may be more necessary in these contexts, leading to a reduced suppression of the onset response. Future research should also aim to better dissociate onset responses from expectancy effects observed in feedback perturbation tasks, which are similar in terms of spatial and temporal profile to onset responses in my data due to the limitations of my naturalistic study design, yet I speculate mechanistically different than onset responses. My findings support a functional network between the lateral temporal lobe, insula, and motor cortex to support natural communication. The differential responses of the speech regions of STG and insula support the role of the posterior insula in auditory feedback control during speaking.

Chapter 4

Discussion

Chapters 2 and 3 of this dissertation include two original research studies on processing of auditory feedback during speech production. There are similarities between the two; for example, they both use the same task: participants read sentences aloud, then passively listened to either consistent (immediate) playback or inconsistent (randomly selected prior trial) playback (Figures 2.1; 3.2A; 3.5A). Notably, this task generates the audio for the playback condition from the production trials, allowing for me to tightly control for acoustic information across my experimental conditions[1].

The EEG results of Chapter 2 demonstrate a suppression of the N1 and P2 components during the speaking trials (Figure 2.2A), a finding in line with previous noninvasive studies of speaker-induced suppression. The novelty of these results lies in the naturalistic, sentence-level stimuli used, as most studies of speech production via noninvasive methods avoid such unconstrained speech stimuli due to issues with motion artifacts (EMG) from the speech articulators.

[1] A caveat concerning the recording technique for both experiments: while perceptual stimuli were generated from production stimuli, minor acoustic differences between the two types of trials may have emerged from bone conduction when perceiving self-generated speech. I did not attempt to mask bone conduction feedback with noise to preserve the naturalistic experimental design of my task.

I was able to correct for those artifacts in my results using a source separation technique during preprocessing (§2.4.4; Appendix A) and by regressing EMG recorded via external electrodes in my encoding models (§2.4.3; Figure 2.4). I showed that the suppression of auditory feedback during speech production observed in the N1/P2 components is unrelated to differential phonetic tuning during speech production, as the linguistic representations in my encoding models are correlated between speech production and speech perception trials (Figure 2.5).

The sEEG results of Chapter 3 further investigated the suppression of auditory feedback using the same task. Because intracranial sEEG has a much higher spatial resolution and is much less susceptible to EMG artifact than scalp EEG, I was able to investigate the results of Chapter 2 with higher precision. In addition to the phonological feature tuning framework I employ in Chapter 2, my analysis in Chapter 3 is also executed through the framework of onset and sustained responses, another organizational principle of the auditory system that has been documented in my advisor's research (Hamilton et al. 2018, 2021). I find that onset responses are suppressed during speech production while sustained responses are not (Figure 3.2). Similar to the SIS observed in the EEG results, onset suppression does not affect phonological feature tuning (Figure 3.6). This onset suppression is localized to primary and non-primary auditory cortex. I did find a separate auditory response region in the posterior insula as well, but this region did not suppress onset responses (Figure 3.3).

This chapter will focus on comparing and contrasting the results of the two studies, as well as discuss the broader implications of both studies.

4.1 Speaker-induced suppression and the auditory system

Suppression of sensory re-afference is believed to be a fundamental component of the motor system and goal-directed movement, which of course includes speech production (Houde & Nagarajan 2011; Parrell et al. 2019). Of specific focus in my dissertation is how this principle of the motor system can affect auditory processing. The literature on this is already fairly deep and I have covered the relevant aspects of it in Chapter 1 (§1.2.3.1). To summarize, research on speaker-induced suppression has focused primarily on the physiology of the phenomenon, and less on the theoretical explanation for it. That being said, one proposed function of SIS is that it is responsible for distinguishing internally and externally generated speech for the purposes of speech motor control (Houde & Nagarajan 2011). SIS emerges from neural synchrony between expectations about utterance content generated before articulation and the sensorimotor/auditory feedback generated during articulation. For example, speech errors are less suppressed than correct speech (Ozker et al. 2022, 2024; Houde et al. 2002; Behroozmand & Larson 2011). Even below the error level, the degree of suppression of auditory feedback during speech production is linked to how well the production of a speech token matches with a canonical sensory goal for that token (Niziolek et al. 2013). For exam-

ple, for vowel production, the first and second formants (f_1, f_2) are essentially proxies for vowel height and backness, respectively (Johnson 2011). Niziolek et al. (2013) had participants produce vowels in words like *heed* /hiːd/ and *head* /hɛd/. If the average formant values for a production of vowel /ɛ/ are f_1 = 550 Hz; f_2 = 1700 Hz, a single production of /ɛ/ with f_1 = 490 Hz; f_2 = 1775 Hz would be *less* suppressed than a single production with f_1 = 555 Hz; f_2 = 1695 Hz because it is farther from the "canonical" production of that vowel, even if it is not consciously perceived as a speech error. So, it is clear that the mechanism of SIS is sensitive to changes in auditory feedback and must interface with the auditory system to some extent. The specifics of this extent are central to the research questions of my dissertation.

4.1.1 Speaker-induced suppression and onset responses

In the original onset and sustained response profile paper (Hamilton et al. 2018), the authors theorized that onset responses may serve a role as an auditory cue detection mechanism based on their utility to detect phrase and sentence boundaries in a decoder framework. Novel stimulus orienting responses have been localized to middle and superior temporal gyrus, which overlaps with the functional region of interest for onset responses (Friedman et al. 2009). These findings are in line with the absence of onset responses during speech production, as auditory orientation mechanisms during speech perception are not necessary to the same extent during speech production due to the presence of a robust forward model of upcoming sensory information

(i.e., efference copy) generated as part of the speech planning process; §1.1.1; Houde & Chang (2015); Tourville & Guenther (2011)).

A notable difference between the original reporting of onset and sustained response profiles in Hamilton et al. (2018) and the ones I present in Chapter 3 is that many of the electrodes reported in my analysis showed a mixture of onset and sustained response profiles, whereas the original paper posits a more stark contrast in the response profiles. This could be due to differences in coverage between the sEEG depth electrodes used here and the pial ECoG grids used in the original study, as the onset response profile was reported to be localized to a relatively small portion of dorsal-posterior STG. Many of onset electrodes in my study were recorded from within STS or other parts of STG; therefore, the activity recorded at those electrodes may represent a mixture of onset and sustained responses, which explains why both would show up in the averaged waveform. Mixed onset-and-sustained responses have been previously reported primarily in HG/PT in a study using ECoG grids covering the temporal plane (Hamilton et al. 2021); the use of sEEG depths in my results may be providing greater coverage of these intra-Sylvian structures. Alternatively, the mixed onset-sustained responses I see in my data may be a mixture of the onset region with the posterior subset of sustained electrodes reported in the original paper. I did observe solely onset-responsive and solely sustained-responsive electrodes (in line with the original paper), but a majority of the onset suppression response profile described in this study consisted of a mixture of onset and sustained responses at the single electrode level.

Responses to the inter-trial click tone observed at some electrodes are another example of pure onset response electrodes in these data.

4.1.2 Speaker-induced suppression and linguistic abstraction

Previous research has shown that electrodes in sensory cortex are preferential to specific classes of phonological features (Mesgarani et al. 2014), which motivated their investigation in this dissertation. Although the degree of SIS remains sensitive to subphonemic changes in auditory feedback (Niziolek et al. 2013), are the invariant representations used in sensory cortex suspect to differences between speech perception and production? Studies such as Niziolek et al. (2013) are important works that help explain the cognitive purpose for SIS, but do not explore in detail the interaction of cortical suppression during speech production with other organizational principles of the auditory system. During speech perception, intermediate, abstract linguistic representations are generated from low-level auditory stimuli in both onset- and sustained-responsive portions of the auditory cortex (Mesgarani et al. 2014; Hamilton et al. 2018). Niziolek et al. (2013) demonstrated that the efference copy, a feedforward expectation about the content of the upcoming auditory stimulus only generated during internally produced speech, contains goal-oriented information. This observation is a critical link in establishing SIS as a neurophysiological biomarker of feedback control, as it shows SIS is sensitive to differences between the efference copy and corollary discharge. A hypothesis put forth by the authors to explain *why* SIS is sensitive

to subphonemic variation is that the efference copy, although itself a precise encoding of motor commands, loses its precision in sensory cortex in favor of invariant encoding of information. Therefore, subphonemic variations would be present in the signals being compared during feedback control. The extent of this theorized goal-based representation in the efference copy (i.e., does it extend to higher levels of linguistic abstraction?) is left as an unanswered question.

During speech production, linguistic representations are used during pre-articulatory planning (e.g., Levelt (1993)'s phonological encoding stage of speech production), meaning the invariant representations are, to some extent, already available to the language network. This could potentially negate the need for additional processing costs associated with segmentation of continuous acoustic information, which is a necessary component of properly perceiving speech. This was the observation that led me to investigate phonological feature representation in Chapters 2 and 3. The results of both studies I present in this dissertation suggest that while SIS is sensitive to subphonemic variation (as demonstrated by Niziolek et al. (2013)), the amplitude differences observed during SIS cannot be explained by differential linguistic feature encoding between speaking and listening (§2.5.2, 3.5.5). In other words, the lack of invariance observed during speech production (Cheung et al. 2016) and speech perception (§1.2.2; Mesgarani et al. (2014)) is not part of the neural signal affected by SIS.

To demonstrate this, I expanded on the N1 suppression observed in

the ERP results of Chapter 2 and the onset suppression observed in Chapter 3 by ablating specific stimulus characteristics from mTRF encoding models and observed how the absence of a specific aspect of the stimulus affects the model's ability to predict the neural response (Figures 2.3, 3.5F). A concurrent reduction in phonological feature response was identified via the mTRF approach in both the EEG and sEEG datasets (Figures 2.5, 3.6E). This suggests that neural activity at multiple levels of representation (sensorimotor activity per SIS, phonological/linguistic representation per mTRF) was suppressed during speaking when compared with listening. Importantly, the structure of the receptive fields themselves was relatively consistent (albeit inverted)—that is, the brain does not shift toward representing different phonological features during perception and production (Figures 2.5, 3.6E). In other words, an electrode that encodes plosive voiced obstruents (like /b/, /g/, /d/) during speech perception will still encode plosive voiced obstruents during speaker-induced suppression, but the amplitude of the response is reduced during speaking.

Although my dissertation results show consistent feature representations between speaking and listening despite magnitude changes within those representations, one previous study did identify changes in feature representation between speaking and listening in the motor cortex (Cheung et al. 2016). In this study, motor electrodes clustered according to place of articulation during speech production, while during passive listening, they clustered according to manner of articulation. However, the authors mention that it is unclear whether these differences in feature representation are the result of a single

intracranial electrode recording from two different populations of neurons (one sensory and one motor) or whether the same population changes its representation depending on task. While I was unable to distinguish between sensory and motor areas with scalp EEG electrodes due to their level of spatial precision, my intracranial recordings are clearly able to distinguish between these areas. I did identify some receptive fields in motor cortex that contain pre-articulatory phonetic feature encoding (Figure C.4), but for the most part my phonologically-tuned electrodes were auditory responses, unlike Cheung et al. (2016).

4.1.3 Biomarkers of speaker-induced suppression

The N1 (and its MEG equivalent M1) has been theorized as a neural indicator of the efference copy, and its suppression has been demonstrated for internally generated speech compared with externally generated speech (Behroozmand & Larson 2011; Martikainen et al. 2005; Heinks-Maldonado et al. 2006), including in the results of Chapter 2 (Figure 2.2A). The P2 is less directly associated with SIS, with limited studies linking it directly to feedback perturbation (Brumberg & Pitt 2019; Behroozmand & Larson 2011), but it is commonly paired with the N1 in speech perception studies to form the N1-P2 complex (Lightfoot 2016). In my results, I do find P2 suppression as well (Figure 2.2B). The N1 and P2 are generally classified as "long-latency" auditory response components in comparison to earlier evoked potentials like the auditory brainstem responses, which have latencies in the

tens of milliseconds (Luck & Kappenman 2013). However, they are still faster than later cognitive components such as the P300 and N400.

The observed suppression of the N1 and P2 components in Chapter 2 without concomitant alteration of phonological feature encoding (§4.1.2) suggests these auditory components do not play a role in linguistic abstraction, even if phonological feature representations emerge on a similar timescale to the N1-P2 complex during speech perception.

4.1.3.1 EEG components and intracranial response profiles

It is tempting to compare the EEG components of Chapter 2 with the intracranial response profiles of Chapter 3, but the comparison is not simple to make. In terms of cognitive function, there are notable similarities between the N1 and the onset response. The N1 was originally conceptualized as an index of "detection of acoustic change" (Hyde 1997), but also plays a more specific role in speech perception: the N1 has been theorized to index speech segmentation as an auditory orientational cueing mechanism in a fashion similar to onset responses (Sanders et al. 2002). In fact, some literature on the N1 explicitly refers to it as an "onset response" (Luck & Kappenman 2013). One notable difference between the EEG auditory response components and the onset responses described in Hamilton et al. (2018) is the size of gap necessary to elicit the response. Onset responses are visible in the N1 after gaps in auditory stimuli as small as 5 milliseconds (Pratt et al. 2005), while the onset responses documented in Chapter 3.5 and originally in Hamilton et al.

(2018) require at least 200 milliseconds of silence to generate.

While the functional links between the N1-P2 and intracranial onset responses are pretty clear, comparisons of the anatomical similarities become more difficult. Source localization in EEG is nontrivial because the scalp and skull act as a filter that spatially smears the data (Luck 2014). Of course, surgically implanted electrodes do not have this issue. The limited number of studies that have directly tried to link intracranial recordings to scalp EEG potentials are generally written with a cautious tone in regards to interpretation of the spatial precision of EEG (Halgren et al. 1995). That being said, the literature on the cortical sources of the auditory N1 agrees that it likely originates from the superior temporal gyrus and Heschl's gyrus, which aligns with the anatomical localization of the intracranial onset responses I report in Chapter 3 (Wolpaw & Penry 1977; Scherg & Picton 1991). However, using this to draw a strong conclusion that the suppression in Chapters 2 and 3 are from the same neural population is still a massive stretch. For example, it is nigh impossible to differentiate between posterior insular and Heschl's responses with EEG, two regions in my data that both exhibit onset responses but with very pronounced differences between them. In the hypothetical scenario that I wanted to conclude that intracranial onset responses were the N1, it would be impossible to localize the N1 activity of Chapter 2 to specifically Heschl's gyrus or the posterior insula.

4.1.4 Expectancy effects during speech perception are a separate mechanism from speaker-induced suppression

A manipulation of whether the perceptual trials immediately followed the production trials from which they were generated (consistent) or not (inconsistent) was included in my dissertation research to assess the hypothesis that SIS is associated with general feedforward auditory processing and not an intrinsic characteristic of corollary discharge during speech production. For the sEEG data specifically, I sought to delineate whether onset responses were an important component of specifically speech perception or involved in a more general predictive processing system.

For the sEEG results, I did observe that presenting auditory playback in a randomized, inconsistent fashion resulted in a greater response amplitude for some onset suppression electrodes in auditory cortex; this finding did not hold true for most onset suppression electrodes in the data. This leads me to believe that the suppression of onset responses is not a byproduct of general expectancy mechanisms modulating the speech perception system, but rather a dedicated component of auditory processing for orienting to novel stimuli. Cortical suppression of self-generated sounds is likely a fundamental component of the sensorimotor system, as neural responses to tones paired with non-speech movements are attenuated relative to unpaired tones in mice and in humans (Martikainen et al. 2005; Schneider et al. 2018). With cNMF, I identified a cluster in ventral sensorimotor cortex that was more active for speech production, but within the consistent/inconsistent playback split, pre-

ferred consistent playback. I interpret this response profile as indicative of feedback enhancement for the purposes of speech motor control during speech production.

In my EEG results, differences between consistent and inconsistent perceptual trials were small, with only three individual participants demonstrating a significant difference. Suppression can emerge from many cortical sources (§4.1.3); therefore, linking any suppression observed during this perception-only manipulation to the cross-modal suppression observed between speaking and listening is not trivial using a scalp recording technique. This lack of a result could be a mixture of the smaller effect size for this manipulation (as observed in the sEEG data) and the lower signal-to-noise ratio of scalp EEG recordings in comparison to intracranial sEEG.

I was motivated to include this manipulation within the speech perception condition by several findings. Behaviorally, participants' habituation to the task can affect results: inconsistent perturbations of feedback during a feedback perturbation task elicit larger corrective responses than consistent, expected perturbations; however, there is no corroborated link between these results and SIS (Lester-Smith et al. 2020; Mollaei et al. 2016; Gonzalez Castro et al. 2014; Jones & Munhall 2000). The importance of predicting upcoming sensory consequences is visible in neural data as well: unpredicted auditory stimuli result in suppression of scalp EEG components for self-generated speech in pitch perturbation studies (Scheerer & Jones 2014) as well as the speech of others in a turn-taking sentence production task (Goregliad Fjaellingsdal

et al. 2020). Other top–down influences on auditory processing and a non-uniformity of suppression across cognitive processes makes the initial research question regarding a connection between forward modeling of auditory stimuli in perception and production difficult to investigate with an unconstrained experimental paradigm.

4.2 The insular auditory field

The intracranial data presented in Chapter 3 offer several distinct advantages[2] over EEG; perhaps the largest advantage is the ability to image individual brain structures separately. sEEG depths in particular have another advantage over similar intracranial techniques in that it can image the insula, which would require arduous dissection of the Sylvian fissure to reach with a standard ECoG grid (Chang 2015; Remedios et al. 2009; Nguyen et al. 2022). In my analysis, the posterior insula served as a unique functional region in processing auditory feedback during speech production and perception (§3.5.2). Unlike temporal cortex, onset responses were not suppressed during speech production in posterior insula; the region instead exhibited "dual onset" responses during speech production and perception. The large amount of non-overlap in weighting of "dual onset" and "onset suppression" clusters' top electrodes suggests that the posterior insula auditory responses I report are not simply spatial runoff from neighboring Heschl's gyrus. Furthermore, my results are focused on the high gamma frequency band, which has less spatial

[2]Of course, intracranial electrophysiology also has its limitations; see §4.4.3.

spread than lower-frequency bands (Muller et al. 2016).

A meta-analysis of the functional role of human insula parcellated the lobe into four primary zones: social-emotional, cognitive, sensorimotor, and olfactory-gustatory (Kurth et al. 2010). As speech production involves sensorimotor and cognitive processes, even speech cannot be constrained to one functional region of the insula. Cytoarchitectonically, the human insula consists of eleven distinct regions which can be grossly clustered into three zones: a dorsal-posterior granular-dysgranular zone, a ventral-middle-posterior agranular-dysgranular zone, and a dorsal-anterior granular zone (Quabs et al. 2022). Based on the general organizational principles of these articles, the dual onset responses I observed in the posterior insula overlap with functional regions of interest for somatosensory, motor, speech, and interoceptive function, and with the dorsal-posterior and ventral-middle-posterior cytoarchitectonic zones. The posterior insula responses I report in this dissertation are purely post-articulatory, indicating a role in auditory feedback monitoring rather than a preparatory motor role. To corroborate, the most robust responses I observed during the speech motor control task were not to the task itself, but rather the click sound that played before each trial (Figure 3.3E). Another recent study identified an auditory region in dorsal-posterior insula through intraoperative electrocortical stimulation, whereby stimulation to posterior insula resulted in auditory hallucinations, confirming the role of this region in auditory processing (Zhang et al. 2018).

A particularly fascinating component of my results in the posterior

insula is the response latency of the zone: I observed faster (or equivalently fast) responses to auditory playback stimuli in the posterior insula compared to primary (HG, PT) and higher order (STG, STS) auditory areas. While posterior insula and HG are neighboring anatomical structures, I do not believe my posterior insula responses to be simply miscategorized HG activity due to the distinction between how HG and posterior insula respectively suppress or do not suppress auditory feedback during speech production. This is supported by animal research that has demonstrated a direct cellular pathway between the auditory thalamus and the posterior insula (§4.2.2), as well as data-driven insights from the Human Connectome Project that have identified functional connectivity between posterior insula and the medial geniculate nucleus of the thalamus (Rolls et al. 2023). While the insular auditory field does receive input from primary and non-primary auditory areas, it also receives direct parallel input from the auditory thalamus, evidenced in part by pure-tone responses in the insular auditory field sometimes having a lower response latency than the primary auditory cortex in a fashion consistent with my results (Jankowski et al. 2023; Sawatari et al. 2011; Takemoto et al. 2014). Thus, the results of Chapter 3 corroborate parallel auditory pathways between auditory cortex and posterior insula but in the human brain and with more complex auditory stimuli than pure tones. I also expand upon prior work by showing responses to auditory feedback in the insula are also present during speech production.

These data are by no means the first documentation of *in vivo* recordings of the human insula's responses to speech perception and production:

Woolnough et al. (2019) also reported post-articulatory activity in the human insula during speech production and perception. My results are distinct from this study in several ways. First, the authors dichotomize the posterior insula with STG, reporting that posterior insula is more active for self-generated speech "opposite of STG." However, my dual onset response electrodes in the posterior insula are equivalently responsive to speech perception and production stimuli, with only a small non-significant preference for speech production. Second, the responses reported in Woolnough et al. (2019) differ in magnitude between STG and the posterior insula, with task-evoked activity in STG increasing $\sim 200\%$ in broadband gamma activity from baseline, while posterior insula showed only $\sim 50\%$ increase in activity from baseline. In my results, temporal and insular evoked activity are similar in magnitude. Third, the authors did not use the same stimuli for speech production and speech perception trials, instead comparing self-generated speech during production to a listening task where speech was generated by another speaker. In my study, I generated perceptual stimuli from individual participants' own utterances, allowing me to control for temporal and spectral characteristics of the stimuli and more directly compare speech perception with production within the posterior insula for the same stimulus.

4.2.1 Multisensory integration in posterior insula

Overall, I interpret the posterior insula's role in speech production as a hub for integrating the multiple modalities of sensory feedback (e.g., auditory,

tactile, proprioceptive) available during speech production for the purposes of speech monitoring, based in part on previous work establishing the insula's role in multisensory integration (Kurth et al. 2010). Diffusion tensor imaging reveals that the posterior insula in particular is characterized by strong connectivity to auditory, sensorimotor, and visual cortices, supporting such a role (Zhang et al. 2018). My research motivates further investigation of the role of the posterior insula in auditory perception and, more specifically, feedback control of speech production.

While most lobes of the brain have clear-cut macro-functionality (e.g., the occipital lobe processes vision), this is not true for the insula. However, Kurth et al. (2010) provide a meta-analysis of insula neuroimaging studies and produce a compelling hypothesis for a potential macro-function in multisensory processing. They categorize the insula according to four functional subdivisions (olfactory/gustatory, somatosensory, cognitive, and social/emotional). Their interpretation for a myriad of functionality is that all of these subfunctions are necessary for generating a "coherent experience of the world," or some sort of perception about the internal/external states of an individual's environment. While this sounds pretty abstract, perception of internal state is a fundamental objective of the speech motor control system, as state estimation is what allows for error detection during feedback monitoring (Houde & Nagarajan 2011).

Speech motor control is fundamentally a multisensory process, as multiple sensory domains provide feedback during the process: tactile/proprioceptive

feedback from the movement of the articulators combined with auditory feedback from speech itself. The posterior insula region of interest I describe in Chapter 3 appears to be mostly auditory, as it was active during speech perception and click responses (pure auditory) as well as speech production (auditory-motor), but not during the nonspeech motor control task (pure motor). If the region was responsive during the speech motor control task, that could have served as an alternative explanation for the rapid response latencies, as tactile feedback is available during speech production earlier than auditory feedback as evidenced by phonetic phenomena such as voice onset time (Johnson 2011); the lack of pre-articulatory activity during speech production trials further suggests this is not the case. But, the multisensory integration documented extensively in the insula could serve as a motivator for why this auditory field I describe is active during speech production—although my posterior insula ROI is not responsive to motor feedback, it may interact with nearby integration circuits in the insula. Further sEEG experimentation explicitly addressing this hypothesis (perhaps also in a more constrained experimental context) could help conclude this speculation.

4.2.2 Insular auditory fields in animal models

Several animal models have been used to identify an auditory field in the posterior insula (Linke & Schwegler 2000; Remedios et al. 2009; Rodgers et al. 2008). This is partially due to the difficulties in recording from human insula that I have already described. Most of the animal research I will

discuss is histological, meaning pathways between neurons are traced using a mixture of anterograde and retrograde staining techniques, which clearly define the connectivity between different brain regions. In nonhuman primates, invasive electrophysiology similar in the abstract to the sEEG data I present is common, while finer spatial resolution microelectrode arrays are utilized in rodent models. Of course, there are caveats of using animal models to study (especially) speech research, as many aspects of vocal communication found in humans are absent from animals both behaviorally and neurobiologically. But, animal research is in many ways the best source of information on auditory processing in the insula currently.

4.2.2.1 In nonhuman primates

The earliest documentation of an insular auditory field in primates showed that neurons in the caudal insular cortex are responsive to click tones in mangabey, rhesus, and squirrel monkeys (Sudakov et al. 1971; Pribram et al. 1954). These responses were low-latency, leading the authors to propose direct projections from the auditory thalamus to the insular auditory field, but this hypothesis was not tested. These early investigations were expanded on by Remedios et al. (2009) to show that insular auditory responses in rhesus monkeys differ from nearby auditory cortex in several ways. First, the insular neurons were less tonotopically organized than auditory cortex. Second, the caudal insula responded preferentially to conspecific vocal communication over other sounds. The insular auditory field documented in this paper bears

further resemblance to my results in their finding that some insular electrodes actually responded faster than primary auditory cortex, further suggesting a direct thalamic projection.

Nonhuman primate research has also demonstrated the multisensory integrative aspect of the insula I described above (§4.2.1). In macaques, the posterior granular insula is posited to integrate sensation from many modalities, including auditory, proprioceptive, and visual information (Evrard 2019). The proposed function for such an integrative region is to assist in the monitoring of self-motion, a fundamental component of feedback control of speech production in humans.

4.2.2.2 In rodents

Histological studies in rodents corroborate the proposed direct thalamic projection to posterior insula auditory areas. Linke & Schwegler (2000) used a staining technique to show that the medial geniculate body of the auditory thalamus projects to the primary auditory cortex, amygdala, and insular cortex. Rodgers et al. (2008) recorded evoked cortical potentials to auditory stimuli in the posterior insula of rat and made several interesting observations. First, they were able to record auditory responses from the insula even when primary auditory cortex was lesioned, confirming a direct projection from the auditory thalamus, as auditory responses would not be possible if they were simply "downstream" of primary auditory cortex. They also suggest the insular auditory field has a specialization for both multisensory integration (based

on overlap with a somatotopic representation) and auditory fear responses (e.g., perception of a tone that was previously conditioned to appear with an adverse stimulus such as a shock). My finding that posterior insular auditory responses have comparable or shorter latencies to auditory responses in primary auditory cortex is also reflected in the animal literature: Sawatari et al. (2011) found that insular auditory field response latencies were consistently faster than core auditory areas in mice. The authors also argue against the insular auditory field being considered an auditory belt area (which would put it directly downstream of primary auditory cortex) based on the response latencies, which is similar to why I believe my posterior insula responses are a parallel-processed auditory area.

While multisensory integration supports higher-order auditory feedback control, I was unable to test for a fear-conditioned auditory processing preference in the confines of my task. However, an abstract parallel can be drawn between the conspecific vocal preferences of macaque insular auditory field (Remedios et al. 2009) and the fear response in rats (Rodgers et al. 2008) in that both of these stimulus types involve some sort of emotional or social component, which could be further supported by the insula's connectivity to the amygdala (Rolls et al. 2023). Returning to human subjects research, Zhang et al. (2018) propose that auditory responses in the posterior insula are an early stage of a posterior-to-anterior sensory-to-affective gradient within the insula. It is possible the conspecific preferences of macaque insula and the fear responses in rat insula are homologues of this gradient in humans. This belief

is stated in Ackermann & Riecker (2004), where they theorize that emotional association regions in the insula eventually evolved to serve a more domain-specific fine motor control role in speech production.

4.2.3 A separate speech planning mechanism in anterior insula

A large portion of the research on the human insula's involvement in speech and language comes from lesion and functional imaging studies that posit a preparatory motor role for the insula in speech (Ackermann & Riecker 2004; Dronkers 1996; Mandelli et al. 2014). However, these studies prescribe this role to the *anterior* insula, whereas my findings are constrained to *posterior* insula, and the insula is far from anatomically or functionally homogenous (Kurth et al. 2010; Quabs et al. 2022; Zhang et al. 2018). Circling back to theoretical models of speech motor control, apraxia of speech being a deficit in motor speech programming would localize AOS to premotor cortex, where a phonological code is converted into a motor program (Tourville & Guenther 2011). However, localization accounts of apraxia of speech vary greatly. Dronkers (1996) localized AOS to the anterior insula, while modern case studies from neurosurgical impairment have instead localized AOS to the posterior middle frontal gyrus (Chang et al. 2020) and middle precentral gyrus (Levy et al. 2023). The latter case is of particular interest, as the middle precentral gyrus has been recently proposed as a phonological-motor coordination region (Silva et al. 2022), which is in line with theoretical conceptualization of AOS as a disorder of motor speech programming. If we assume this classifica-

tion of AOS to be true, the insular localization of AOS put forth in Dronkers (1996) could be either a byproduct of confounding vasculature (as suggested n Hillis et al. (2004)) or indicative of functional connectivity between the insula and middle precentral gyrus, something which has been demonstrated in diffusion tensor imaging research (Mandelli et al. 2014). However, recent theories have posited a separation of AOS into phonemic and prosodic subtypes based on whether the underlying motor impairment is to laryngeal motor control (dorsal-middle precentral gyrus) or articulatory control (ventral precentral gyrus); further research is needed to see if the anterior insula is functionally linked to one of these subdivisions of the speech production stream over the other (Hickok et al. 2023). Regardless, the absence of pre-articulatory motor activity in my insular "dual onset" electrodes clearly separates this anterior motor region from the posterior auditory one I describe.

4.3 Pre-articulatory activity

Before articulation, a communicative desire must be morphologically, syntactically, and lexically encoded before it is transformed into a motor program for the speech articulators (Flinker et al. 2015; Tourville & Guenther 2011; Levelt 1993). In my EEG analysis, I observed a positive deflection in the grand average ERP (Figure 2.2) that began ∼200 milliseconds before articulation and peaked ∼100 milliseconds before articulation present in the speech production trials. I believe this activity to be related to the feedforward linguistic and motoric preparation that must take place before articulation.

I also observed pre-articulatory motor activity in some prefrontal/premotor electrodes in my sEEG dataset. However, the exact pre-articulatory stages of speech production are difficult to dissociate with this task, as there is no epoched timing information available as to when these processes occur in a naturalistic context. Even in sEEG, the supplementary speech motor control task does not segment out these stages, simply providing a "go" signal. Regardless, the presence of this pre-articulatory activity exclusively during the speech production trials motivates these stages as an explanation.

In the EEG results, prestimulus activity was also observed in the grand average during *perception* trials in the form of positive activity starting at -600 milliseconds and peaking at stimulus onset. This activity may be related to predictive components of speech perception, as feedforward processing is an important aspect of successful speech perception (Hamilton et al. 2021; Heald & Nusbaum 2014; Poeppel & Monahan 2011). This speculation is supported by the structure of the task allowing participants to anticipate when they would hear a sentence; however, this task was not operationalized in a way that allows a more granular analysis of this phenomenon. The lack of a strong contrast between consistent and inconsistent playback in the EEG results is also contrary to this hypothesis, as the consistent playback condition is fundamentally more predictable than the inconsistent playback condition and thus would show an enhancement of activity if prediction is what is driving the pre-stimulus activity in perception trials. Notably, for both perception and production, the polarity of the prestimulus activity was inconsistent from subject to subject.

This internal inconsistency suggests the activity is not related to previously described ERP components (e.g., readiness potential/Bereitschaftspotential) as these components have a canonical negative polarity (Jahanshahi & Hallett 2003; Yoshida et al. 1999; Wohlert 1993). An alternative explanation for pre-articulatory activity in this task is this activity is reflective of residual uncorrected EMG; however, the integrity of task-related neural components suggests any EMG activity capable of producing such a large deflection would not be present in the corrected data (see Appendix A).

When the pre-stimulus perceptual activity in the EEG results are re-contextualized with the results of the consistency manipulation in sEEG, a new hypothesis emerges. A subset of precentral electrodes that clustered with other pre-articulatory electrodes in the cNMF analysis also showed an enhanced response during consistent playback relative to inconsistent playback (although altogether more responsive to production trials; Figure 3.5D, e3 & e4). An interesting speculation that also explains why this result may not generalize across all participants is that this consistency enhancement reflects sub-vocalic rehearsal during consistent trials. Because consistent playback trials are temporally much closer to their matched production trial (cf. inconsistent playback trials, whose corresponding production trial is taken from a previous block and may be several minutes in the past), information about the auditory and articulatory feedback present in the efference copy may be accessible still to participants, allowing tandem silent articulation with the perceptual stimulus. Unfortunately, it is impossible to know which participants

were engaging in this activity, but future inclusion of a dedicated silent production component of the task may help better define these curious precentral responses.

4.4 Limitations

There is a fundamental tradeoff between the ecological validity and size of a dataset with ease of interpretability and experimental control (Ivanova et al. 2021b). Because many aspects of my experiments presented in this dissertation are relatively unconstrained, there are several limitations I would like to discuss that may be alleviated with future task design.

4.4.1 Electromyographic artifact

In any noninvasive neuroimaging study of speech production, movement artifacts caused by articulation are a concern to the integrity of the data. I extend recent results studying speech production at a four-word phrase level (Ries et al. 2021) by scaling up to the sentence level with evoked responses to speech appearing relatively cleaned of EMG artifact as evidenced by the integrity of the N1 and P2 components. Because of the success of prior studies in analyzing event-related EEG data during overt speech production (Ries et al. 2021; Riès et al. 2013; Vos et al. 2010), I did not present further corroboration of the artifact correction techniques as a primary result of my study; however, because this study used a more continuous speech stimulus than the prior studies described above, I provide an investigation into the efficacy of my

artifact correction techniques in Appendix A. This appendix is in many ways a "recycling" of components of my Masters thesis and a subsequent manuscript dedicated to this method that was rejected from publication after peer review and ultimately reworked into the content of Chapter 2 and its corresponding journal article (Kurteff et al. 2023).

One reason to assume that residual EMG is affecting the results is the differing performance of encoding models that do or do not regress EMG (Figure 2.4). Models that accounted for EMG as a stimulus characteristic on the whole outperformed models that did not, which means there is variance remaining in the postprocessed data that is well explained by EMG activity. The inclusion of an EMG regressor was only made possible by recording facial muscle activity using auxiliary electrodes in conjunction with the EEG, akin to how EEG researchers will record auxiliary VEOG and HEOG to assist with artifact correction. Although previous research has demonstrated blind source separation-based artifact correction techniques are sufficient in correcting EMG artifact for ERP analysis, the substantial difference in model performance when this normalized EMG activity was ablated from the stimulus matrix leads me to strongly recommend the use of auxiliary EMG recordings to any researchers who wish to fit similar linear encoding models to speech production data, especially as portable/multi-person EEG studies (i.e., increased susceptibility to movement artifact) become more popular. Furthermore, I only recorded single-channel EMG, whereas there are a plethora of facial muscles that contribute to EMG artifact in the electroencephalogram. It is possi-

ble that including activity from multiple auxiliary channels as a regressor in mTRF models would further improve their performance, but future research is needed to substantiate this claim. For more translational applications (such as brain-computer interfaces), there is a trade-off between ease of use and amount of sensors which should inform EMG sensor location and quantity.

There are several reasons I do not believe the residual EMG in my response nullifies the interpretation of these results. First, the integrity of purely auditory responses is preserved after post-processing as evidenced by evoked responses to intertrial click tones, which suggests the evoked responses seen at the sentence level are not false positives caused by EMG artifact (see Appendix A). Second, despite the contribution of EMG to linear encoding models, I observe strong phonological feature tuning consistent with previous research (Desai et al. 2021; Hamilton et al. 2021). Third, including EMG as a regressor in linear encoding models ensures that phonological feature tuning (or a similar feature space of interest) is not obscured or affected by muscle artifact. Models that include an EMG regressor show similar trends to the models I report on in my sEEG analysis, which corroborates their integrity. Lastly, evoked responses to sentence onset contained robust N1 and P2 components that would not be visible in the presence of substantial noise from EMG. The general profile of suppression of early activity during speech production is also visible in sEEG.

It is necessary to include a disclaimer here about sEEG. sEEG is not special when compared to other recording techniques in that there is still no

way to obtain fundamental ground truth in the activity. Meaning, artifact can never be fully ruled out. In EEG recordings, EMG artifact power is strongest in the 20-30 Hz (β) range (Goncharova et al. 2003). In theory, performing a Hilbert transform to extract high gamma ($H\gamma$; 70-150 Hz) analytic amplitude for analysis as I did in my sEEG results should sidestep much of the movement artifact, as this frequency band is outside the typical range for EMG artifact. The 20-30 Hz range for EMG is not fundamentally linked to the recording technique (EEG vs. sEEG) but is rather a function of the firing rate of motor neurons which are consistent regardless of recording technique. However, as EMG artifact originates from a myriad of sources, it has an astonishingly wide frequency range, with some studies reporting EMG artifact as high as 300 Hz (Chen et al. 2019). A recent study used external EMG electrodes similar to my approach in Chapter 3 but in conjunction with sEEG activity and found a narrowband (at approximately the participant's fundamental frequency (f_0)) gamma component that correlated strongly with the timecourse of the external EMG recording, suggesting that mechanical vibration caused by speech production may be a source of artifact in addition to motor neuron activity (Bush et al. 2022). Similarly to how I advocate for external EMG recordings in scalp EEG experiments based on the results of Chapter 3, the authors of this manuscript advocate for external EMG recordings in conjunction with sEEG to minimize spurious interpretation of speech production data. While I agree with this precaution, it was not possible for me to implement in my dissertation study due to my experimental design predating the publication of

this manuscript. A supplemental analysis of my speech production trials and their correlation with the spectrotemporal characteristics of the participant's f_0 has the potential to serve as a post-hoc validation of the absence of EMG artifact, but I have not conducted such an analysis at this time.

4.4.2 The playback consistency manipulation

The absence of results in this manipulation in Chapter 2 and a relatively weak effect in Chapter 3 may be related to the task's block design, which was used to explicitly avoid eliciting an oddball response. My study presented the consistent and inconsistent perceptual trials in blocks of 50 (for EEG) or 20-25 (for sEEG) trials each, which means participants could identify when perceptual stimuli would be inconsistent with the preceding trial, a fundamental difference from the oddball tasks where deviant stimuli are presented randomly. I additionally chose not to present inconsistent stimuli in an oddball fashion because the perceptual stimuli were generated from the recorded productions of the participant. Thus, to generate the full range of inconsistent perceptual stimuli in the task, a full block of production trials is needed, and collecting this as a baseline before introducing oddball inconsistent stimuli would greatly extend the time of the recording sessions, and I judged that more repetitions of each condition would be more important to my research questions[3]. The block design of the task may also cause listeners

[3]Time is a limited resource for intracranial data collection. Intraoperative sessions are limited to twenty minutes at most, and even bedside research studies (such as mine) have the lowest priority for patient interaction compared to the wide variety of clinician visits.

to adapt to the inconsistent playback stimuli over the course of the block. A similar experiment in which all inconsistent playback trials were interspersed randomly among consistent playback trials would facilitate a comparison to conventional "oddball" studies of predictability in the EEG literature. Designing a study in this way may also minimize the influence of potential extraneous top–down manipulations on auditory processing (§1.2.3.3). Top–down manipulations of expectations about perceptual content were a variable of interest for this dissertation, which is why I avoided an oddball design.

The naturalistic design of the stimuli introduces many potential top–down processes, not just the forward modeling of perceptual trial content that I sought to investigate. For example, participant engagement with the stimuli can affect the degree of SIS observed (§1.2.3.3). My participants were not instructed to actively listen and were not required to make any responses concerning the playback condition, meaning within-subject differences in attentive listening during speech perception were left up to the independent engagement of the participant with the task. On the other hand, speech production necessitates active listening as part of the feedback control system (§1.1). Thus, without prompting active listening during perception, participant engagement may have comprised some of the fundamental differences between speaking and listening task conditions. More attention to the precise manipulation of an attentional contrast in future studies may yield more informative results, including potentially exploring differences in the N1/onset response between active and passive speech perception trials.

4.4.3 Recording from people with intractable epilepsy

Intracranial recordings have many advantages over noninvasive recording techniques which I have discussed extensively at this point, but I have not discussed the downsides of intracranial recordings. First, the data is incredibly rare to acquire compared to noninvasive recordings and is only possible within a handful of clinical populations (Chang 2015). This is the first limitation that makes generalization of results from intracranial studies difficult. Furthermore, the fact that every single clinical population in which sEEG and ECoG are recorded has a clinical necessity for brain surgery of some sort means that most (if not all) have abnormal neuroanatomy/physiology, which imposes another limit on generalization to the healthy brain. It also limits generalization within the population, as common pathologies for sEEG studies like epilepsy can affect individuals' brains drastically different. This may help to explain some of the single-subject response profiles in my sEEG results, such as the frontal production-responsive electrodes of DC5 (Figure B.2). One of the most common techniques for generalizing across subjects, which I employ, is to "warp" electrodes to a common template space. But, in severe cases, abnormal pathologies such as tumors or resection cavities may create difficulties in the generation of 3D reconstructions from the subject's MRI, which can in turn affect the reliability of the anatomical parcellation atlas (Hamilton et al. 2017). Nevertheless, all 3D reconstructions are manually checked and corrected in instances where abnormal pathology affects surface reconstruction and electrode localization which ideally mitigates these issues. This potential

scenario is relevant to my dissertation dataset as many of my participants are pediatric or adolescent, and severe epilepsy during the early stages of neurodevelopment can drastically alter the typical structure-function relationships of the brain (Karami et al. 2020). Even in adults epilepsy can alter language lateralization, with many people with epilepsy exhibiting more bilateral language than people without epilepsy (Möddel et al. 2009). However, it is important to mention that this is the *only* way to acquire intracranial recordings from the human brain—there are no alternatives for researchers who want these data. So, any caveats related to the nature of the clinical population serve as fundamental limitations. The only way to avoid this limitation is to corroborate intracranial research with noninvasive recordings in healthy controls, which I provide in Chapter 2.

4.5 Future directions

Because the datasets I present in my dissertation are both large naturalistic corpora, there are many additional analyses that could be conducted within them before new data are collected. For example, a collaborator of my lab (Yao Chen) is currently conducting a sub-analysis on the data presented in Chapter 3 focusing on neural responses to speech errors during the production component of the task. In general, speech errors are an intriguing next step, as the speaker-induced suppression literature shows that speech errors result in less suppression of neural activity during speech production (Behroozmand & Larson 2011; Ozker et al. 2022). While natural error analyses (such as the one

described above) are certainly possible, an alternative possibility is to employ an experimental paradigm that elicits speech errors, such as tongue twisters (Ries et al. 2021) or feedback perturbation paradigms (Jones & Munhall 2000; Toyomura et al. 2020). I have collected pilot data on a delayed auditory feedback task in which auditory feedback is temporally delayed ∼200 milliseconds in a closed-loop system (e.g., headphones) to induce speech errors. I chose this method over pitch perturbation as delayed auditory feedback is impossible to compensate for, making the task more difficult/more likely to elicit errors (Stuart et al. 2002). The rest of the task is designed to resemble the task of Chapters 2 and 3, in that it involves sentence-level overt speech production and a passive playback condition, which hopefully will facilitate comparison to the results of my dissertation. I plan to use this paradigm to investigate how the suppression of onset responses differs during speech errors in both the auditory cortex and posterior insula; a difference in suppression modulation between these two regions could help differentiate the functions of these parallel auditory areas.

4.5.1 Speech motor control across the lifespan

One advantage of recording sEEG from children's hospitals is that my dataset contains neural activity from a wide range of age groups, including late childhood, early adolescence, late adolescence, and adulthood. While the data presented in Chapter 3 are only from seventeen participants, a larger sample size will afford comparisons across age groups, something my group has begun

to investigate in other tasks. The speech motor control system does not reach maturity until adolescence (Walsh & Smith 2002), which leads me to hypothesize that suppression of onset responses (and speaker-induced suppression more generally) varies by developmental stage, with the amount of suppression increasing with age until adolescence. Simplifying the task may help to recruit younger age groups, as my task as-is necessitates literacy, something which is not guaranteed in children with epilepsy who often have concomitant difficulties with literacy (Lah et al. 2017). The delayed auditory feedback pilot I discuss above attempts to address this, as the task is a picture description task based on a pre-recorded auditory stimulus (e.g., the participant hears What are the "dinosaurs doing?" when presented with an illustration of dinosaurs playing soccer), which removes the literacy requirement but also adds an additional passive listening condition (externally generated playback, versus the internally generated playback stimuli used in Chapters 2 and 3). A pediatric scalp EEG study similar to what I presented in Chapter 2 may also help to provide developmental insights with a larger sample size.

4.5.2 Speaker-induced suppression in apraxia of speech

Several clinical populations have demonstrated abnormal SIS when compared to healthy controls, including people who stutter (Toyomura et al. 2020; Max & Daliri 2019), people with schizophrenia (Heinks-Maldonado et al. 2007; McGuire et al. 1995), people with Parkinson's disease (Railo et al. 2020), but I want to focus on apraxia of speech for this discussion, as I believe the

potential links between the disorder and SIS are less well-defined. AOS is conventionally defined as a deficit in motor speech programming with a relatively spared feedback monitoring system: Ballard et al. (2018) used an auditory feedback perturbation paradigm to show that individuals with AOS exhibit typical compensatory responses to feedback perturbation of f_0 and f_1 . Jacks & Haley (2015) also altered auditory feedback through pitch perturbation and noise masking. Their results showed that noise masking, but not pitch perturbation, increased speech fluency in people with AOS but not in healthy control participants. The authors conclude that an increase in speech fluency when auditory feedback is unavailable/unreliable indicates an over-reliance on feedback control in people with AOS to compensate for an impaired feedforward control system.

So, Jacks & Haley (2015) theorized that an over-reliance on auditory feedback may be a fundamental compensatory mechanism for people with apraxia of speech. What does that mean for the neural processes that support the feedback monitoring system? It is unknown whether neural responses to auditory feedback in people with AOS would appear similar to healthy controls, partially due to difficulties in gathering a large cohort of people with AOS to investigate, as AOS is often confounded with expressive aphasia (§1.3.1; Patidar et al. (2013); Kobayashi & Ugawa (2013)). My hypothesis is that people with AOS may exhibit less speaker-induced suppression during speech production than healthy controls. Recall that (1) SIS is an index of adherence to the efference copy (§1.2.3.1; Niziolek et al. (2013); Behroozmand & Larson

(2011)); (2) the degree of cortical suppression is diminished during speech errors (Ozker et al. 2022, 2024); and (3) SIS decreases during non-error acoustic deviances from a typical production (Niziolek et al. 2013). An over-reliance on auditory feedback could potentially lead to less suppression in people with AOS, as feedback monitoring systems are more active.

4.5.3 Onset responses in brain-computer interfaces

Brain-computer interfaces (BCI), specifically speech-generating devices, are an emerging neurotechnology that offer a unique mode of communication to people with severe motor disorders such as amyotrophic lateral sclerosis and brainstem stroke (Herff et al. 2015; Moses et al. 2021; Rabbani et al. 2019). While the bulk of research at this point is not conducted in real time and/or focuses on healthy controls, the ultimate goal of the field is naturalistic, embodied communication for people who are completely "locked-in," meaning real-time decoding from clinical populations is the future of speech BCI (Metzger et al. 2023). As people with locked-in-syndrome have a complete loss of voluntary motor movement, use of any acoustic or motor activity in these speech decoders is not possible: neural activity is the only source of input for the BCI. This poses a unique challenge that resembles imagined speech production research in the abstract: it is impossible to tell precisely when a speech event begins without a corresponding audible, visible, or palpable event. Thus, "event detection" algorithms which decode the beginning of a communication attempt are a necessary piece of the puzzle for developing speech BCI. Fur-

thermore, such a mechanism must be incredibly precise, as both failing to detect (false negative) or erroneously decoding neural activity not linked to an explicit communication attempt (false positive) are both frustrating fail cases for an individual dependent on such a device to communicate. Preliminary case studies of speech BCI in paralyzed patients have begun to employ speech event detection in their decoding pipeline, but the examples in the literature are still quite limited (Moses et al. 2021). Onset responses, which index the beginning of an acoustic event, could potentially be utilized as a neural landmark for a speech event decoder. Their suppression during speech production could help to minimize false positives in an event decoder, as any decoded speech event that contains an onset response would be either speech perception or an erroneous production, both of which should not be decoded. A decoder which recorded activity from temporal cortex in a sliding window and attempted to decode a speech event whenever an onset response is detected as suppressed could be an elegant solution to the problem of ambiguous timing information in imagined speech. A future study that explicitly demonstrates the suppression of onset responses in imagined speech would further motivate such a use case for onset suppression.

4.6 Conclusion

Speech perception and production have been siloed and studied independently for decades; only recently is this beginning to change. The two studies in this dissertation approached this dichotomy more from the speech

perception side due to my mentor's training. As a result, I was interested in how organizational principles of the auditory system might change during speech production, given the context of more coarse differences in neural processing between speaking and listening (e.g., speaker-induced suppression). Phonological feature tuning, a theoretical proxy for linguistic abstraction during speech perception, was an attractive phenomenon to investigate because recent studies had shown that phonological feature tuning was present in non-auditory regions of the brain (Cheung et al. 2016) and was encoded separately in auditory cortex within two spatially separated "onset" and "sustained" regions (Hamilton et al. 2018). So, phonological feature tuning appeared to me as somewhat malleable: perhaps local tuning might change during speech production, or, more generally, when the predictability of an auditory stimulus changes. While I conclude that phonological feature tuning in local populations actually does not shift in these experimental manipulations, I was able to demonstrate a suppression in the tuning of phonological features during speech production that more resembles a global "gain change" than some fundamental representational shift. Differences in neural activity were visible in the weighting (or waveform, depending on the analytic technique), but not in the aspects of the stimulus that my models suggest the brain were encoding.

When contextualized with the goal-oriented and acoustic sensitivity of speaker-induced suppression, I believe this to be an interesting result. SIS is a biomarker of the error detection process, as self-generated auditory feedback becomes un-suppressed when one makes an error; however, it can also become

un-suppressed without being detected as an error based on very subtle changes in the corollary discharge relative to the efference copy. Meaning, while our brain is able to interpret slight deviations from expectations as acceptable (a form of invariance), non-errored speech, some other aspects of the auditory system are still quite sensitive to these deviations. While the parts of the brain that suppress corollary discharge (via SIS) are responding differently to low-level, non-invariant acoustics, it is reassuring to see in my results that phonological feature tuning is not also being shifted. If I *did* report a change in phonological feature tuning during SIS, this would imply that phonological features are directly "downstream" of acoustics and not a dedicated intermediate form of representing sensory information.

My intracranial results provided two other findings worth concluding on: the suppression of onset responses during speech production and the existence of a low-latency insular auditory field. The suppression of onset responses (but not sustained responses) during speech production lends to the theory that onset responses are involved in the speech segmentation process as a temporal landmark detection mechanism. When speaking, we have knowledge of where these temporal landmarks are in the efference copy, so the auditory system does not need onset responses to help suss them out. The auditory responses in the insula, on the other hand, did not suppress onset responses during speech production, so whatever role the insula has in auditory processing is likely unrelated to speech segmentation. My working hypothesis is that the insula is helping to concatenate sensory processing from the auditory,

tactile, and other sensory domains during speech, but as of now that's a future direction.

We still have much to understand about speech production, speech perception, language, and motor control as individual complex cognitive processes. Auditory feedback processing is a marriage of all these domains, meaning we have even more to learn. The studies I present in this dissertation are a small step towards understanding this process in the brain, with hopes of someday benefitting translational research relating to the assessment and treatment of disordered auditory feedback processing, something present in conditions such as apraxia of speech, stuttering, Parkinson's disease, aphasia, and schizophrenia, to name a few.

Appendices

Appendix A

Validation of EMG artifact correction during EEG task

In Chapter 2, I used a blind source separation technique based on CCA to identify and correct for EMG artifact (De Clercq et al. 2006). This approach has been successful in removing EMG artifact associated with articulation in previous EEG studies of overt speech production (Ries et al. 2021; Riès et al. 2013; Vos et al. 2010); my approach is described in the Methods section of that chapter (§2.4). These prior studies I reference have all focused on speech production at the word or phrase level whereas my study focuses on sentence-level speech production, a more naturalistic form of speech that could potentially elevate the risk for EMG contamination of the data. As a safeguard, I recorded EMG via auxiliary facial electrodes during the task (Figure 2.1). Including these recordings as a regressor in linear encoding models reduces the influence of residual EMG artifact on the response (Figure 2.4); however, for the ERP analyses, external validation of the data set's integrity was necessary to deem it suitable for analysis.

To accomplish this, I compared EEG responses to the task before and after CCA artifact correction. Responses were epoched to the acous-

tic onset of the first phoneme of the sentence (as in Figure 2.2), as well as two non-task-related events: the acoustic onset of the intertrial click tone and peaks in the auxiliary EMG electrode activity as detected by the function `mne.preprocessing.create_eog_epochs()` (Figure 2.1). Significance between pre-and post-CCA-corrected epochs was assessed via LME modeling using a technique similar to the one described in Chapter 2 (§2.4): a fixed effect of trial type and a random effect of subject (`RMS_difference` $\sim$ `Condition + (1|Subject)`). However, instead of raw voltage values, difference waves between uncorrected and CCA-corrected activity were calculated at each epoch by subtracting the root-mean-square of the two responses averaged across channels, then averaging across the first 300 msec of activity relative to the epoch of interest. Although polarity is important for interpreting EEG components such as N1 and P2, I opted to make no assumptions about the polarity of potential EMG artifact by using the root-mean-square of the response. If EMG were successfully removed from the data, epochs associated with EMG activity (speech production and peak auxiliary activity) should show a larger difference wave between uncorrected and CCA-corrected activity than epochs unassociated with EMG activity (speech perception and intertrial clicks).

Linear-mixed effects modeling (Equation 2.1) provided a confirmation that EMG was successfully removed from the data set using CCA while preserving the integrity of the neural response (Figure A.1). I report the *EMM* and standard error of the difference waves here. Epochs associated with articulatory activity showed a large difference before and after CCA artifact

correction ($EMM_{\text{Production}} = 58.40 \pm 78.6\ \mu V$; $EMM_{\text{Aux Peak}} = 87.7 \pm 79.4\ \mu V$), whereas epochs associated with passive listening showed a small difference before and after CCA artifact correction ($EMM_{\text{Perception}} = 12.1 \pm 79.2\ \mu V$; $EMM_{\text{Click}} = 19.8 \pm 77.5\ \mu V$). The difference in how these epoch types were affected by CCA was further corroborated by the effect sizes of the LME model's contrasts, calculated as Cohen's d (Cohen 2013). Contrast between epochs that both involve articulation or both involve passive listening were small ($\Delta_{\text{Aux Peak}-\text{Production}}\ d = 0.038,\ p = .39$; $\Delta_{\text{Click}-\text{Perception}}\ d = 0.01,\ p = .96$), whereas contrasts between epochs that differed in their expected contamination with EMG activity were large ($\Delta_{\text{Click}-\text{Aux Peak}}\ d = -0.089,\ p < .001$; $\Delta_{\text{Click}-\text{Production}}\ d = -0.05,\ p = .07$; $\Delta_{\text{Aux Peak}-\text{Perception}}\ d = -0.1,\ p < .001$; $\Delta_{\text{Perception}-\text{Production}}\ d = -0.06,\ p = .06$).

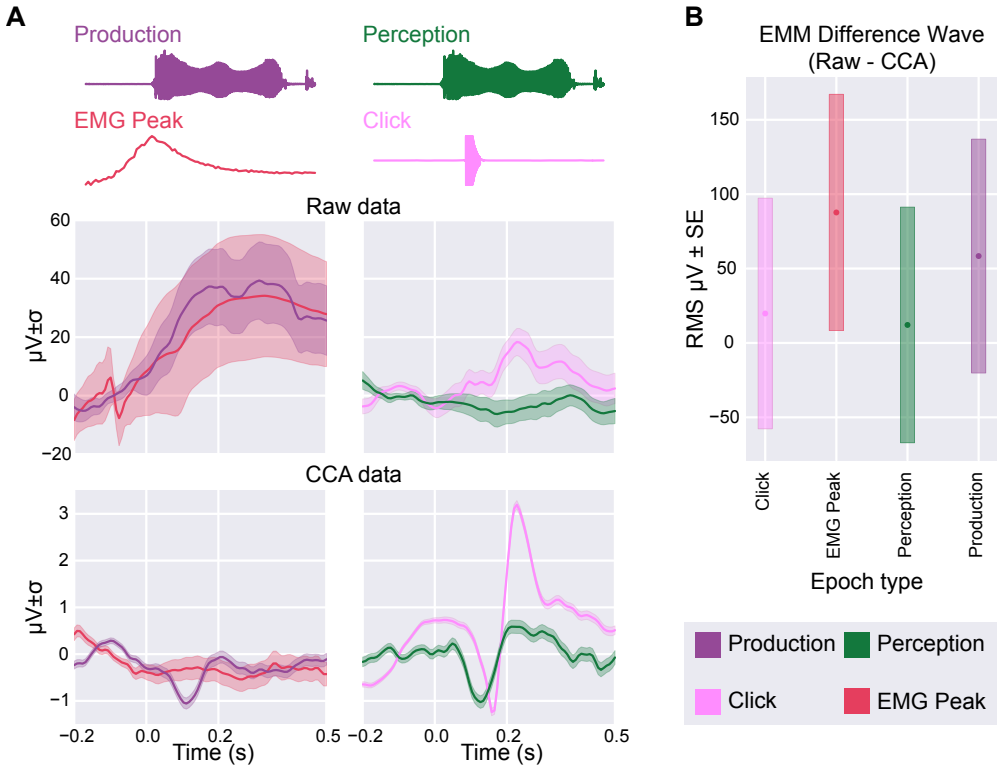

Figure A.1: **Comparison of EEG activity before and after EMG artifact correction**.
(A) Stimuli (top) and grand average ERP of raw data (middle) and CCA-corrected data (bottom) relative to displayed stimuli. Grand average plots are separated by the epochs' anticipated level of contamination with EMG artifact. Left panels (red, purple) show epochs that are anticipated to be contaminated because of their association with articulation. Right shows (green, pink) epochs that are anticipated to contain relatively less EMG artifact because of their association with passive listening; however, jaw clenching during passive listening means these data cannot be assumed to be EMG free.
(B) LME model EMMs for the RMS amplitudes of 0–300 msec raw-CCA difference waves for each of the four epochs of interest. Shaded area represents standard error. A value closer to zero indicates less activity was subtracted from the EEG response during CCA artifact correction.

These validation techniques suggest that CCA is a sensitive and specific method for correcting EMG activity in my data set. Although these results are promising, an important caveat is that there is no guaranteed method of confirming an artifact technique is both successful (no Type I error) and accurate (no Type II error); EEG has no "ground truth" for source localization (Bradley et al. 2016). This caveat motivates the use of external validation techniques described here, but also imposes a fundamental limitation on all EEG studies which employ artifact correction techniques. I argue for the integrity of my results despite this limitation, and I encourage those interested in using artifact correction techniques to study naturalistic speech production via EEG to do so.

Appendix B

Unique single-subject response profiles in the sEEG results

Because the dataset from Chapter 3 uses sEEG depth electrodes, I was able to record from a wide array of cortical and subcortical areas impractical or impossible to cover with ECoG grids. As a result, there were several interesting trends observed within single subjects that were not robust enough to report upon earlier but do warrant a more speculative discussion.

Occipital coverage was generally limited for this study, but one subject (DC7) had three electrodes in the right lateral occipital cortex that strongly preferentially responded to speech production trials and to click responses (Figure B.1 e2 PT-MT15 $p_{production} = 0.01$; $p_{perception} = 0.9$). I identified this area using my unsupervised clustering analysis: cNMF identified a cluster selective to clicks and speech production localized to the occipital lobe (Figure C.1, cluster 6). I interpret this as a byproduct of my task design, as text was displayed during speech production trials (the sentence to be read aloud) but not during perception trials (Figure B.1D). The between-peak duration of the bimodal click response observed in the cNMF cluster is ∼1000 milliseconds, which corresponds with the amount of time a fixation cross was displayed at

the beginning of each trial (§3.4.5). Based on this information, I conclude these occipital electrodes for DC7 are encoding visual scene changes between fixation cross and text display, but I advise caution in generalizing this to a functional localization as I only observed this trend in a single subject.

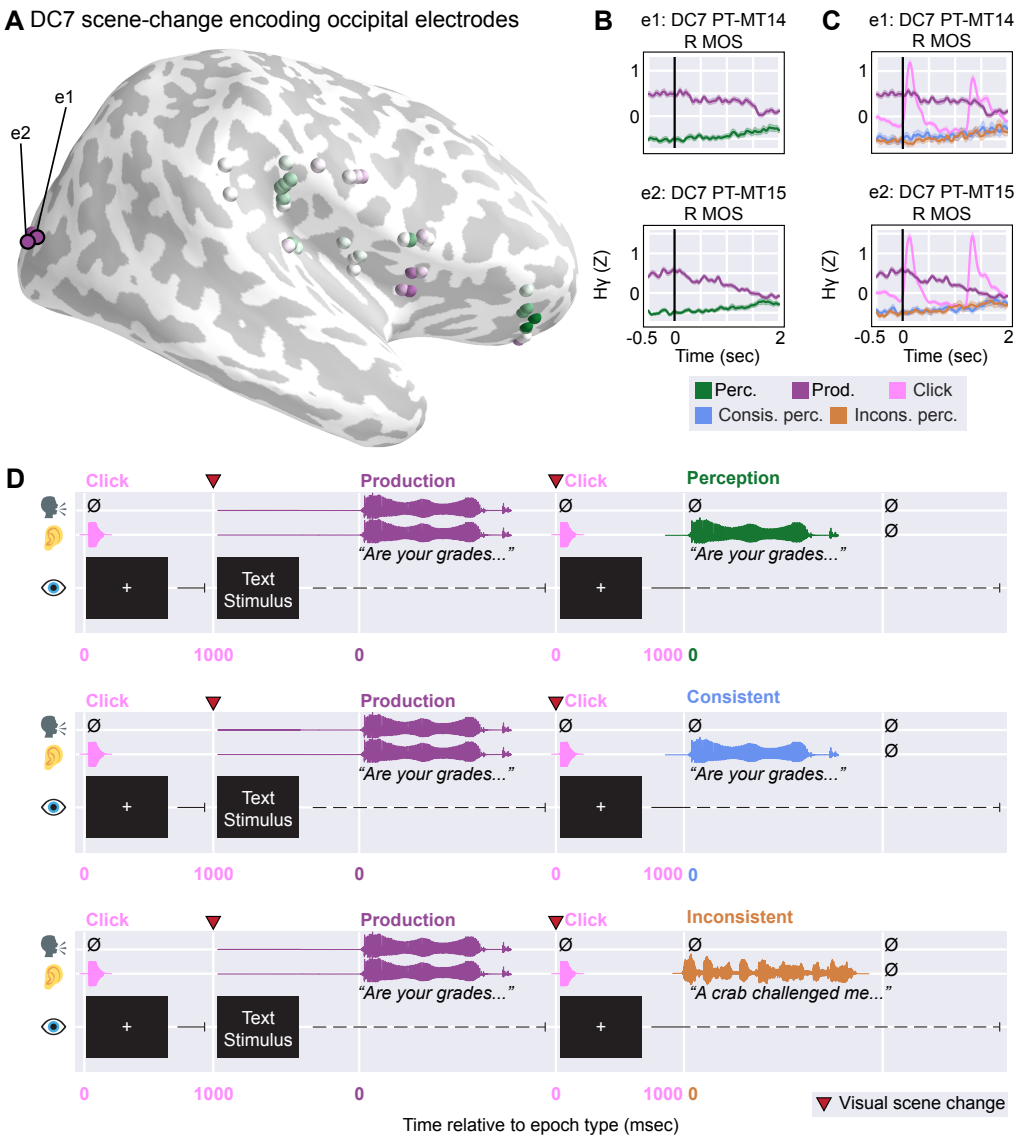

Figure B.1: **Single-subject visual scene change responses in occipital cortex**.
(A) Inflated cortical reconstruction of single-subject (DC7) right hemisphere with significant electrodes (*SI* bootstrap t-test; see §3.4.8) visualized. Light gray represents gyri while dark gray represents sulci. Electrodes are colored according to their *SI* values. Example electrodes in (B) and (C) are indicated.
(B) Single-electrode plots showing visual scene change responses in middle occipital sulcus during speech production (purple) and perception (green). Shaded area represents margin of error. Subplot titles reflect the participant ID and electrode name from the clinical montage.
(C) Single-electrode plots showing responses to speech production (purple), consistent (blue) and inconsistent (orange) playback conditions, and the inter-trial click (pink). Shaded area represents margin of error. Subplot titles reflect the participant ID and electrode name from the clinical montage. The electrodes in this panel appear to be most responsive during speech production and the click sound, both of which temporally correlate with visual scene changes.
(D) Expanded task schematic to illustrate where visual scene changes occur in the task. Rows represent information seen, heard, and spoken by the participant over the course of a trial. The time on the X-axis is not to scale due to trial-to-trial variability in reaction time duration in participant responses and is instead relative to the different types of events visualized at t=0 in (B) and (C). Multiple panels are provided to emphasize that the timing of events does not fundamentally change for consistent versus inconsistent playback. Visual scene changes are indicated on the timeline with a red triangle.
Abbreviations: MOS: middle occipital sulcus.

In a separate single subject (DC5), I observed electrodes in the right inferior frontal sulcus (just dorsal of pars triangularis of the inferior frontal gyrus) that responded selectively to speech perception and inter-trial click tones (Figure B.2 e1 $p_{production} = 0.31$; $p_{perception} < .001$). Unlike onset suppression electrodes in auditory cortex, these electrodes were silent during speech production for onset and sustained responses. The amplitude of production responses increased as the depth progressed laterally towards pars triangularis, but the

final electrode of the depth still had a (barely) non-significant response to speech production trials (DC5 AMF-AI8 $p_{production} = 0.06$; $p_{perception} = 0.45$). Unlike the occipital electrodes described above, the inferior frontal perception-selective electrodes of DC5 did not emerge as a functional region in the unsupervised clustering analysis and were interspersed with other perception-selective electrodes from other subjects localized to PT and HG (Figure C.1, cluster 7). While the convention of inferior frontal cortex being monolithically a speech production region is increasingly being challenged in contemporary research (Fedorenko & Blank 2020; Flinker et al. 2015; Hickok et al. 2023; Tremblay & Dick 2016), the confinement of the perception-selective electrodes in this region to a single subject gives me hesitation to bolster those claims with these data.

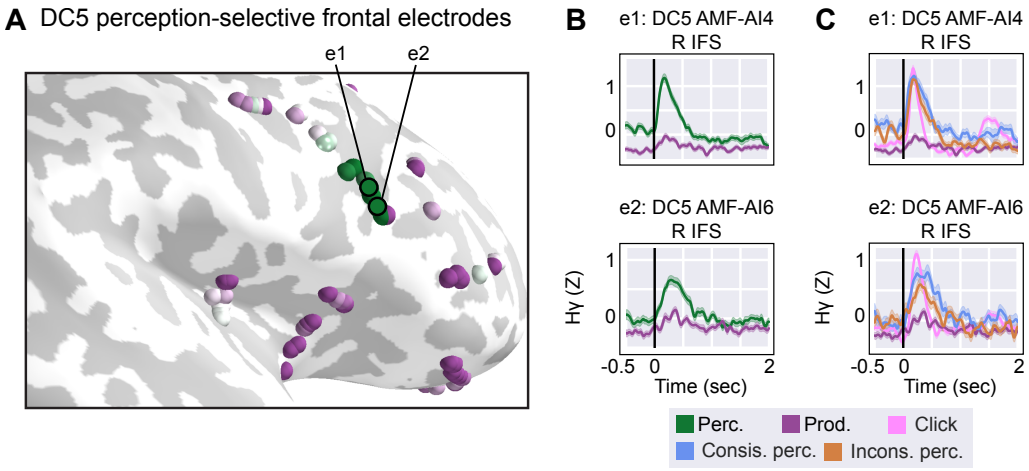

Figure B.2: **Single-subject perceptual responses in inferior frontal cortex**.
(A) Inflated cortical reconstruction of single-subject (DC5) right hemisphere with significant electrodes (*SI* bootstrap t-test; see §3.4.8) visualized. Light gray represents gyri while dark gray represents sulci. Electrodes are colored according to their *SI* values. Example electrodes in (B) and (C) are indicated.
(B) Single-electrode plots showing perceptual responses in inferior frontal cortex during speech production (purple) and perception (green). Shaded area represents margin of error. Subplot titles reflect the participant ID and electrode name from the clinical montage.
(C) Single-electrode plots showing responses to speech production (purple), consistent (blue) and inconsistent (orange) playback conditions, and the inter-trial click (pink). Shaded area represents margin of error. Subplot titles reflect the participant ID and electrode name from the clinical montage.
Abbreviations: IFS: inferior frontal sulcus.

Appendix C

Supplemental figures

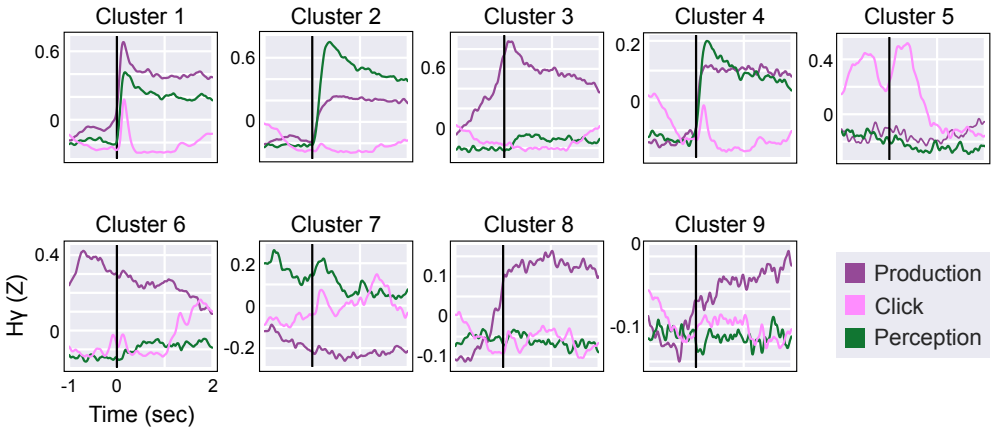

Figure C.1: **9 presented cNMF clusters explain 86% of the variance in the data** (§3.4.7; Figure 3.4A).
"Onset Suppression" and "Dual Onset" clusters presented in Results (Figure 3.4B) here are labeled as Clusters 2 and 1, respectively. "Pre-articulatory Motor" cluster presented in Results (Figure 3.4B) here is labeled as Cluster 3. The responses plotted are the cluster basis functions of individual clusters relative to either sentence onset (production and perception conditions) or the inter-trial click tone (click condition).

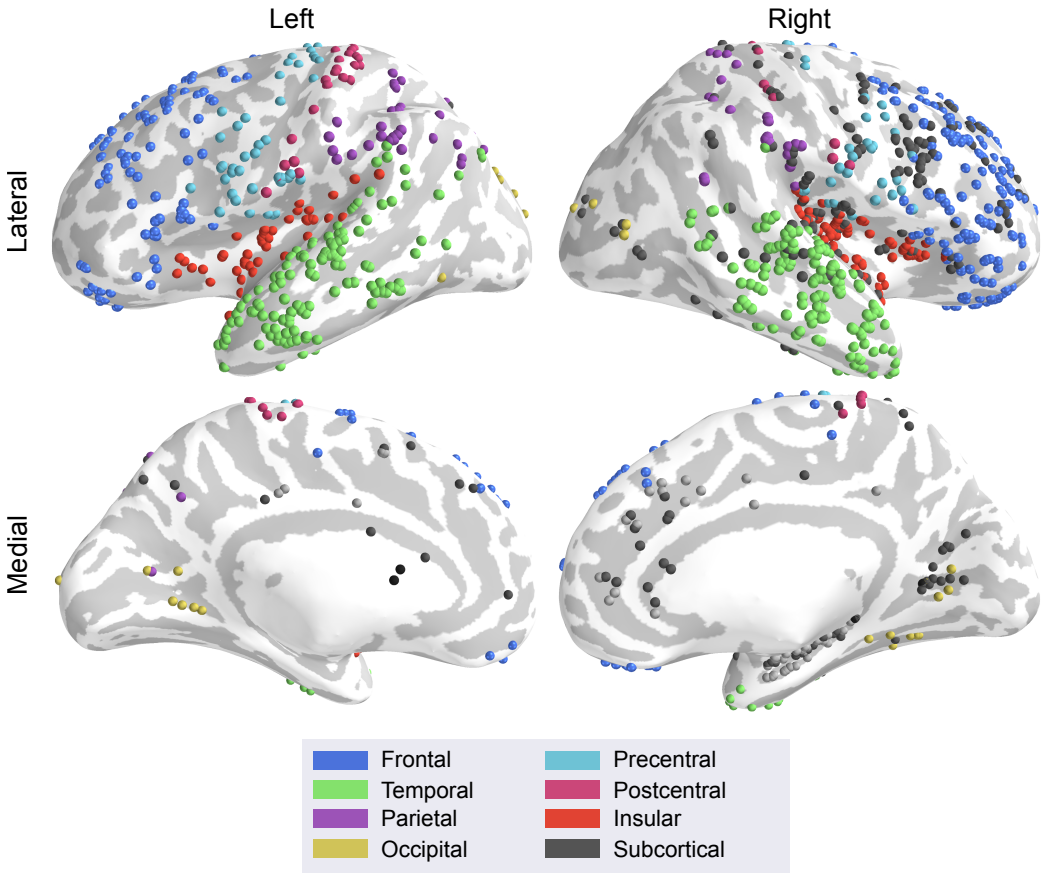

Figure C.2: **Individual electrodes for all subjects with available imaging (n=15) plotted on the `cvs_avg35_inMNI152` atlas brain, color-coded by anatomical region of interest**.
Cortical surface inflated for better visualization of insular electrodes. Electrode visualization in native subject space is shown in Figure C.3.

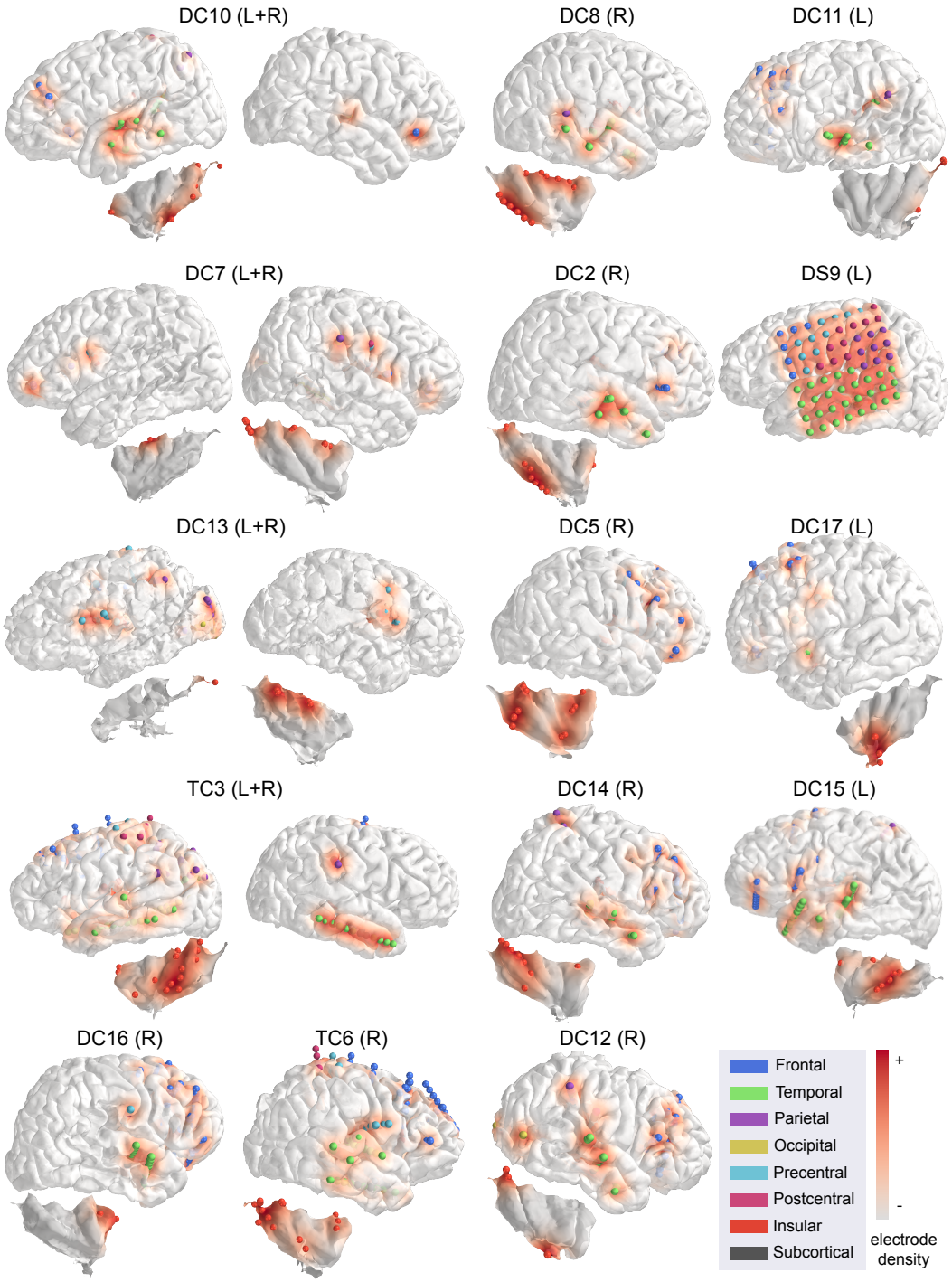
DC10 (L+R)
DC8 (R)
DC11 (L)
DC7 (L+R)
DC2 (R)
DS9 (L)
DC13 (L+R)
DC5 (R)
DC17 (L)
TC3 (L+R)
DC14 (R)
DC15 (L)
DC16 (R)
TC6 (R)
DC12 (R)
Frontal
Temporal
Parietal
Occipital
Precentral
Postcentral
Insular
Subcortical
+
-
electrode density

Figure C.3: **Electrodes visualized on 3D reconstructions of individual subjects' MRIs, color-coded by anatomy**.
Color gradient represents density of electrode coverage. A separate reconstruction of individual subjects' insulas is provided for visualization of insular electrodes not visible from lateral cortical surface. Each subject displayed here is visualized on an averaged brain in Figure C.2.

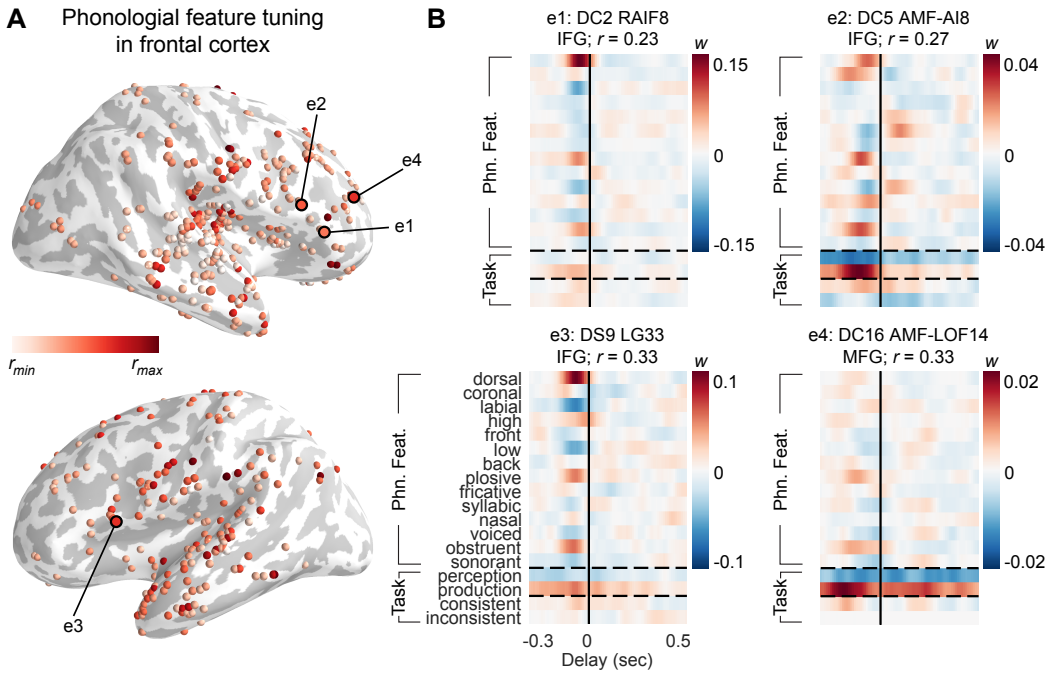

Figure C.4: **Phonological feature representation in negative delays in inferior frontal cortex**.
(A) Inflated template brain reconstruction identical to Figure 3.6B but with example electrodes from (B) indicated instead. Dark gray indicates sulci while light gray indicates gyri. Color corresponds to linear correlation coefficient (r) values of mTRF models at a single-electrode level.
(B) Single-electrode temporal receptive fields demonstrating phonological feature tuning in inferior frontal cortex across participants. Notably, the strongest weighting for phonological features is consistency at negative delays (pre-articulatory). Phonological feature tuning is strongest in IFG across participants (e1, 2, 3) and receptive fields in other areas of frontal cortex are better modeled by task-level features (e4), but show the same temporal selectivity as phonologically tuned electrodes in IFG.
Abbreviations: IFG: inferior frontal gyrus; MFG: middle frontal gyrus.

Appendix D

Supplemental tables

ID	Age	Sex	Languages Spoken	EMG Placement
OP1	24	X	English	N/A
OP2	25	F	English, Gujarati	N/A
OP3	21	F	English, Spanish	orbicularis oris & mandible
OP4	23	F	English	orbicularis oris & mandible
OP5	18	F	English, Mandarin	orbicularis oris & mandible
OP6	22	F	English	orbicularis oris & mandible
OP7	27	F	English, Spanish	orbicularis oris & mandible
OP8	23	F	English	orbicularis oris & mandible
OP9	24	F	English, Spanish, Polish	orbicularis oris & mandible
OP10	30	M	English	orbicularis oris & mandible
OP11	21	F	English	submentalis
OP12	26	M	English	masseter
OP13	35	M	English	orbicularis oris & mandible
OP14	23	M	English	masseter
OP15	23	F	English	mylohyoid
OP16	25	M	English	orbicularis oris & mandible
OP17	30	M	English	masseter
OP18	28	M	English	masseter
OP19	23	M	English	masseter
OP20†	21	M	English	orbicularis oris & mandible
OP21	20	F	English	masseter

Table D.1: **Participant demographics for EEG participants discussed in Chapter 2**.
Participant IDs marked with (†) are excluded from analysis due to a recording error.

ID	Age	Sex	Seizure Focus
TC1†	9	M	Left temporo-parieto occipital
DC2	14	M	Right hemisphere
TC3	19	F	Left temporal
DC4	21	M	No access to record
DC5	19	M	Right frontal
TC6*	14	F	Right sensory frontal
DC7*	20	M	Right temporal
DC8	16	F	No strong localization
DS9	37	M	Left temporal
DC10*	14	X	Left temporal
DC11	13	M	Bilateral frontal
DC12	14	F	Right hemisphere and midline
DC13*	8	F	Left hemisphere
DC14	18	F	Right frontal white matter
DC15	20	F	Left frontotemporal
DC16*	9	F	Right frontal
DC7*	17	F	Left frontal

Table D.2: **Table of age, sex, and seizure localization for each participant discussed in Chapter 3**.
Participant IDs marked with (*) participated in the supplementary speech motor control task described in §3.4.5.1. Participant IDs marked with (†) were excluded from analysis due to the presence of tuberous sclerosis complex. M = male, F = female, X = patient declined to disclose.

Appendix E

Glossary of acronyms

A

A1 *primary auditory cortex*; first used in §1.2.1 ◇ The final output of the ascending auditory pathway and the first part of the cerebral cortex that processes auditory sensation. Also referred to as Heschl's gyrus.

AOS *apraxia of speech*; first used in §1.3.1 ◇ A motor speech disorder characterized by an impairment of the speech articulators without concomitant muscle weakness.

B

BCI *brain-computer interface*; first used in §4.5.3 ◇ A term for any piece of technology (software or hardware) that allows a human to control a device using neural activity alone. These are an attractive future direction for speech neurosciences as a whole, as people who cannot communicate by other means (e.g., people with locked-in syndrome) could benefit from such a device.

C

CCA *canonical correlation analysis*; first used in §2.4.4 ◇ A source separation technique used for EMG artifact correction in EEG (see **EMG**, **EEG**).

cNMF *convex non-negative matrix factorization*; first used in §3.4.7 ◇ An unsupervised clustering technique that weights electrodes according to the similarity of their responses without access to anatomical information about the electrodes.

D

DIVA *directions of velocities into articulators*; first used in §1.1.1 ◇ A neuroanatomically and computationally precise model of speech production and speech motor control (Tourville & Guenther 2011).

E

ECoG *electrocorticography*; first used in §1.1.2.1 ◇ An intracranial neuroimaging technique which records the local field potential of neurons through a grid of electrodes that are placed on the pial surface of the brain during surgery.

EEG *electroencephalography*; first used in §1 ◇ A neuroimaging technique that records excitatory postsyaptic potentials of neurons (Buzsáki

et al. 2012). Usually refers to scalp-based recording techniques but intracranial variants exist as well (see **ECoG**, **sEEG**).

EMG *electromyography*; first used in §2.2 ◇ A recorded electrical potential from a motor neuron and a common source of artifact in EEG recordings (see **EEG**).

EMM *estimated marginal means*; first used in §2.5.1 ◇ A common statistic reported in linear-mixed effects model analyses that takes the mean over the fitted results of the model.

EOG *electrooculography*; first used in §2.4.3 ◇ Electrical potentials generated by eye movement and a common source of artifact in EEG recordings (see **EEG**). Commonly divided into horizontal (hEOG) and vertical (vEOG) components, which correspond to eyeblinks and saccades, respectively.

ERP *event-related potential*; first used in §2.2 ◇ An analysis technique commonly used in EEG research where neural response is averaged relative to the presentation of a stimulus to identify a canonical response to that stimulus (see **EEG**; Luck (2014)).

F

F_1**;** F_2 *first and second formants*; first used in §4.1 ◇ The first and second formants are common measures utilized in acoustic phonetics and serve as a proxy for vowel height and laterality (i.e., front vs. back).

FACTS *feedback-aware control of tasks in speech*; first used in §1.1.1 ◇ A theoretical model of speech motor control which emphasizes the goal-based nature of speech production (see Parrell et al. (2019)).

fMRI *functional magnetic resonance imaging*; first used in §1 ◇ A neuroimaging technique which measures hemodynamic changes in blood oxygenation via a superconducting magnet as a metric of brain activity. The "f" in fMRI refers to the comparison of activity during a task to a baseline control to measure task-related neural activity.

H

hEOG *horizontal electrooculography*; first used in §2.4.3 ◇ See **EOG**.

HG *heschl's gyrus*; first used in §1.2.1 ◇ A gyrus on the top of the temporal lobe that receives early auditory information from the auditory thalamus. Part of the primary auditory cortex (A1) with the planum temporale (PT). Not to be confused with high gamma, which I abbreviate as $H\gamma$.

HSFC *hierarchical state-feedback control*; first used in §1.1.1 ◇ A theoretical model of speech control which posits the brain monitors sensory feedback during speech production using an internal state estimation of the vocal tract (Houde & Nagarajan 2011).

I

ICA *independent components analysis*; first used in §2.4.4 ◇ A source separation technique used to correct artifact in EEG activity (see **EEG**).

IFG *inferior frontal gyrus*; first used in §1.1.2 ◇ The ventral-most gyrus of the frontal lobe, bordered posteriorly by precentral gyrus and inferiorly by the Sylvian fissure. Consists of three parts: pars triangularis, opercularis, and orbitalis. Broca's area is thought to be somewhere in IFG but there is disagreement in the field about exactly where. The definition I use in this dissertation is pars triangularis and opercularis, a.k.a. the posterior 2/3 of IFG.

J

JoCN *journal of cognitive neuroscience*; first used in §2.1 ◇ A peer-reviewed journal in which the results of Chapter 2 are published (Kurteff et al. 2023).

L

LME *linear mixed-effects modeling*; first used in §2.4 ◇ A statistical model capable of separately regressing fixed and random effects; used in both results chapters to account for across-subject variation in neural response.

M

MEG *magnetoencephalography*; first used in §1.2.3.1 ◇ A noninvasive neuroimaging technique which measures the magnetic fields generated by neuron potentials via a superconducting magnet.

MNI *montreal neurological institute*; first used in §3.4.4 ◇ A common three-dimensional coordinate system for the brain, named after the titular institute affiliated with McGill University.

MOCHA *multichannel articulatory database*; first used in §2.4 ◇ A corpus of sentences to be spoken that accounts for the natural phonetic variation of English, originally curated by Texas Instruments and MIT (Wrench 1999).

MRI *magnetic resonance imaging*; first used in §1 ◇ See **fMRI**.

mTRF *multivariate temporal receptive field*; first used in §2.4.7 ◇ see **TRF**.

P

PET *positron emission tomography*; first used in §1 ◇ An invasive neuroimaging technique which measures change in metabolic activity using a radioactive tracer as an index of neural activity.

PLT *perceptual loop theory*; first used in §1.1.1 ◇ A theoretical model of speech motor control which posits the perceptual systems of the brain

are responsible for feedforward and feedback control (Indefrey & Levelt 2004).

PT *planum temporale*; first used in §1.2.1 ◇ A portion of cortex just posterior to Heschl's gyrus (HG), on top of the temporal lobe and tucked into the Sylvian fissure. Part of primary auditory cortex (A1).

R

ROI *region of interest*; first used in §3.4.4 ◇ A neuroanatomical unit specifying an area of the brain that is either functionally (e.g., frontal eye fields) or anatomically (e.g., superior temporal gyrus) constrained.

S

sEEG *stereo-electroencephalography*; first used in §1 ◇ An invasive neuroimaging technique which records the local field potential of neurons through strips of surgically implanted electrode contacts that penetrate the cortex.

SI *suppression index*; first used in §3.4.8 ◇ A quantification of how much an electrode responds during speech production versus perception trials that has been utilized in prior research of auditory feedback processing (Flinker et al. 2010).

SIS *speaker-induced suppression*; first used in §1.2.3.1 ◇ A phenomenon in which the neural response interally generated speech is suppressed in

relation to externally generated speech.

Spt *temporo-parietal junction*; first used in §1.1.2.2 ◇ Spt is a commonly referenced auditory-motor integration area situated somewhere in posterior superior temporal gyrus and/or supramarginal gyrus. The term was coined by Hickok and colleagues in the early 2000s and is technically not an acronym, but instead an ironic shortening of "the spot" (Hickok, p.c.). A common backronym for Spt is "Sylvian parietal temporal" (Hickok 2007).

STG *superior temporal gyrus*; first used in §1.1.1 ◇ An important structure in the temporal lobe for speech perception; sometimes referred to as non-primary auditory cortex or Wernicke's area.

STS *superior temporal sulcus*; first used in §3.4.8 ◇ The sulcus that runs inferior of superior temporal gyrus (see **STG**).

STRF *spectrotemporal temporal receptive field*; first used in §2.4.7 ◇ see **TRF**.

T

TRF *temporal receptive field*; first used in §2.4.7 ◇ A linear modeling technique based on ridge regression that I utilize in both chapters. It has many names in the literature, but mTRF (multivariate TRF), STRF (spectrotemporal TRF), and linear encoding model are the most common.

V

vEOG *vertical electrooculography*; first used in §2.4.3 ◇ See **EOG**.

Bibliography

Ackermann, H. & Riecker, A. (2004). The contribution of the insula to motor aspects of speech production: a review and a hypothesis. Brain Lang., 89, 320–328.

Aertsen, A. M. & Johannesma, P. I. (1981). The spectro-temporal receptive field. a functional characteristic of auditory neurons. Biol. Cybern., 42, 133–143.

Appelbaum, I. (1996). The lack of invariance problem and the goal of speech perception. In Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96, vol. 3, pp. 1541–1544 vol.3.

Ardila, A. (2010). A proposed reinterpretation and reclassification of aphasic syndromes. Aphasiology, 24, 363–394.

Arsenault, J. S. & Buchsbaum, B. R. (2016). No evidence of somatotopic place of articulation feature mapping in motor cortex during passive speech perception. Psychon. Bull. Rev., 23, 1231–1240.

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

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.

Barry, R. J., Kirkaikul, S., & Hodder, D. (2000). EEG alpha activity and the ERP to target stimuli in an auditory oddball paradigm. Int. J. Psychophysiol., 39, 39–50.

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

Benabid, A. L. (2003). Deep brain stimulation for parkinson's disease. Curr. Opin. Neurobiol., 13, 696–706.

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.

Berezutskaya, J., Freudenburg, Z. V., Güçlü, U., van Gerven, M. A. J., & Ramsey, N. F. (2017). Neural tuning to Low-Level features of speech throughout the perisylvian cortex. J. Neurosci., 37, 7906–7920.

Boatman, D., Gordon, B., Hart, J., Selnes, O., Miglioretti, D., & Lenz, F. (2000). Transcortical sensory aphasia: revisited and revised. Brain, 123 (Pt 8), 1634–1642.

Boersma, P. (2002). Praat, a system for doing phonetics by computer. Glot. Int., 5, 341–345.

Bouchard, K. E. & Chang, E. F. (2014). Control of spoken vowel acoustics and the influence of phonetic context in human speech sensorimotor cortex. J. Neurosci., 34, 12662–12677.

Bouchard, K. E., Mesgarani, N., Johnson, K., & Chang, E. F. (2013). Functional organization of human sensorimotor cortex for speech articulation. Nature, 495, 327–332.

Bradley, A., Yao, J., Dewald, J., & Richter, C.-P. (2016). Evaluation of electroencephalography source localization algorithms with multiple cortical sources. PLoS One, 11, e0147266.

Breshears, J. D., Molinaro, A. M., & Chang, E. F. (2015). A probabilistic map of the human ventral sensorimotor cortex using electrical stimulation. J. Neurosurg., 123, 340–349.

Broca, P. (1865). Sur le siège de la faculté du langage articulé. Bull. Mem. Soc. Anthropol. Paris, 6, 377–393.

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.

Buchsbaum, B. R., Baldo, J., Okada, K., Berman, K. F., Dronkers, N., D'Esposito, M., & Hickok, G. (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory - an aggregate analysis of lesion and fMRI data. Brain Lang., 119, 119–128.

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

Bush, A., Chrabaszcz, A., Peterson, V., Saravanan, V., Dastolfo-Hromack, C., Lipski, W. J., & Richardson, R. M. (2022). Differentiation of speech-induced artifacts from physiological high gamma activity in intracranial recordings. Neuroimage, 250, 118962.

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

Casserly, E. D. & Pisoni, D. B. (2010). Speech perception and production. Wiley Interdiscip. Rev. Cogn. Sci., 1, 629–647.

Caucheteux, C., Gramfort, A., & King, J.-R. (2023). Evidence of a predictive coding hierarchy in the human brain listening to speech. Nat Hum Behav, 7, 430–441.

Caviness, V. S., Makris, N., Montinaro, E., Sahin, N. T., Bates, J. F., Schwamm, L., Caplan, D., & Kennedy, D. N. (2002). Anatomy of stroke, part i: an MRI-based topographic and volumetric system of analysis. Stroke, 33, 2549–2556.

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

Chang, E. F., Kurteff, G., Andrews, J. P., Briggs, R. G., Conner, A. K., Battiste, J. D., & Sughrue, M. E. (2020). Pure apraxia of speech after resection based in the posterior middle frontal gyrus. Neurosurgery, 87, E383–E389.

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.

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

Chartier, J., Anumanchipalli, G. K., Johnson, K., & Chang, E. F. (2018). Encoding of articulatory kinematic trajectories in human speech sensorimotor cortex. Neuron, 98, 1042–1054.e4.

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.

Cheung, C., Hamilton, L. S., Johnson, K., & Chang, E. F. (2016). The auditory representation of speech sounds in human motor cortex. Elife, 5.

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

Cohen, J. (2013). Statistical power analysis for the behavioral sciences. (London, England: Routledge), 2nd edn.

Creutzfeldt, O., Ojemann, G., & Lettich, E. (1989). Neuronal activity in the human lateral temporal lobe. II. responses to the subjects own voice. Exp. Brain Res., 77, 476–489.

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.

Cui, H. & Andersen, R. A. (2007). Posterior parietal cortex encodes autonomously selected motor plans. Neuron, 56, 552–559.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. i. segmentation and surface reconstruction. Neuroimage, 9, 179–194.

Darley, F., Aronson, A., & Brown, J. (1975). Motor speech disorders. (Saunders).

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.

de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L., & Theunissen, F. E. (2017). The hierarchical cortical organization of human speech processing. J. Neurosci., 37, 6539–6557.

Desai, M., Holder, J., Villarreal, C., Clark, N., Hoang, B., & Hamilton, L. S. (2021). Generalizable EEG encoding models with naturalistic audiovisual stimuli. J. Neurosci., 41, 8946–8962.

Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage, 53, 1–15.

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.

Dichter, B. K., Breshears, J. D., Leonard, M. K., & Chang, E. F. (2018). The control of vocal pitch in human laryngeal motor cortex. Cell, 174, 21–31.e9.

Dick, F., Tierney, A. T., Lutti, A., Josephs, O., Sereno, M. I., & Weiskopf, N. (2012). In vivo functional and myeloarchitectonic mapping of human primary auditory areas. J. Neurosci., 32, 16095–16105.

Ding, C., Li, T., & Jordan, M. I. (2010). Convex and semi-nonnegative matrix factorizations. IEEE Trans. Pattern Anal. Mach. Intell., 32, 45–55.

Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. Nature, 384, 159–161.

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. Proc. Natl. Acad. Sci. U. S. A., 111, 7126–7131.

Duffy, J. R. (2019). Motor Speech Disorders E-Book: Substrates, Differential Diagnosis, and Management. (Elsevier Health Sciences).

Englot, D. J. & Chang, E. F. (2014). Rates and predictors of seizure freedom in resective epilepsy surgery: an update. Neurosurg. Rev., 37, 389–404; discussion 404–5.

Evans, A. C., Collins, L., Mills, S. R., & Peters, T. M. (1993). 3D statistical neuroanatomical models from 305 MRI volumes. In Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record., vol. 1813–1817, pp. 1813–1817 vol.3.

Evrard, H. C. (2019). The organization of the primate insular cortex. Front. Neuroanat., 13, 43.

Fedorenko, E. & Blank, I. A. (2020). Broca's area is not a natural kind. Trends Cogn. Sci., 24, 270–284.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. Proc. Natl. Acad. Sci. U. S. A., 110, 16616–16621.

Flinker, A., Chang, E. F., Kirsch, H. E., Barbaro, N. M., Crone, N. E., & Knight, R. T. (2010). Single-trial speech suppression of auditory cortex activity in humans. J. Neurosci., 30, 16643–16650.

Flinker, A., Korzeniewska, A., Shestyuk, A. Y., Franaszczuk, P. J., Dronkers, N. F., Knight, R. T., & Crone, N. E. (2015). Redefining the role of broca's area in speech. Proc. Natl. Acad. Sci. U. S. A., 112, 2871–2875.

Forseth, K. J., Hickok, G., Rollo, P. S., & Tandon, N. (2020). Language prediction mechanisms in human auditory cortex. Nat. Commun., 11, 5240.

Fox, P. T., Ingham, R. J., Ingham, J. C., Hirsch, T. B., Downs, J. H., Martin, C., Jerabek, P., Glass, T., & Lancaster, J. L. (1996). A PET study of the neural systems of stuttering. Nature, 382, 158–161.

Friedman, D., Goldman, R., Stern, Y., & Brown, T. R. (2009). The brain's orienting response: An event-related functional magnetic resonance imaging investigation. Hum. Brain Mapp., 30, 1144–1154.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. Psychon. Bull. Rev., 13, 361–377.

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.

Gauvin, H. S. & Hartsuiker, R. J. (2020). Towards a new model of verbal monitoring. J Cogn, 3, 17.

Germann, J. & Petrides, M. (2020). The ventral part of dorsolateral frontal area 8A regulates visual attentional selection and the dorsal part auditory attentional selection. Neuroscience, 441, 209–216.

Geschwind, N. (1970). The organization of language and the brain. Science, 170, 940–944.

Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., Nastase, S. A., Feder, A., Emanuel, D., Cohen, A., Jansen, A., Gazula, H., Choe, G., Rao, A., Kim, C., Casto, C., Fanda, L., Doyle, W., Friedman, D., Dugan, P., Melloni, L., Reichart, R., Devore, S., Flinker, A., Hasenfratz, L., Levy, O., Hassidim, A., Brenner, M., Matias, Y., Norman, K. A., Devinsky, O., & Hasson, U. (2022). Shared computational principles for language processing in humans and deep language models. Nat. Neurosci., 25, 369–380.

Gómez-Herrero, G. (2007). Automatic artifact removal (AAR) toolbox v1. 3 (release 09.12. 2007) for MATLAB. Tampere University of Technology.

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.

Gonzalez Castro, L. N., Hadjiosif, A. M., Hemphill, M. A., & Smith, M. A. (2014). Environmental consistency determines the rate of motor adaptation. Curr. Biol., 24, 1050–1061.

Goodglass, H. & Kaplan, E. (1972). The Assessment of Aphasia and Related Disorders. (Lea & Febiger).

Goregliad Fjaellingsdal, T., Schwenke, D., Scherbaum, S., Kuhlen, A. K., Bögels, S., Meekes, J., & Bleichner, M. G. (2020). Expectancy effects in the EEG during joint and spontaneous word-by-word sentence production in german. Sci. Rep., 10, 5460.

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.

Guenot, M., Isnard, J., Ryvlin, P., Fischer, C., Ostrowsky, K., Mauguiere, F., & Sindou, M. (2001). Neurophysiological monitoring for epilepsy surgery: the talairach SEEG method. StereoElectroEncephaloGraphy. indications, results, complications and therapeutic applications in a series of 100 consecutive cases. Stereotact. Funct. Neurosurg., 77, 29–32.

Guenther, F. H. (2016). Neural Control of Speech. (MIT Press).

Halgren, E., Baudena, P., Clarke, J. M., Heit, G., Liégeois, C., Chauvel, P., & Musolino, A. (1995). Intracerebral potentials to rare target and distractor auditory and visual stimuli. i. superior temporal plane and parietal lobe. Electroencephalogr. Clin. Neurophysiol., 94, 191–220.

Hamberger, M. J. (2007). Cortical language mapping in epilepsy: a critical review. Neuropsychol. Rev., 17, 477–489.

Hamilton, L. S. (2024). Neural processing of speech using intracranial electroencephalography: Sound representations in the auditory cortex. In Oxford Research Encyclopedia of Neuroscience. (Oxford University Press).

Hamilton, L. S., Chang, D. L., Lee, M. B., & Chang, E. F. (2017). Semi-automated anatomical labeling and inter-subject warping of High-Density

intracranial recording electrodes in electrocorticography. Front. Neuroinform., 11, 62.

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. Language, Cognition and Neuroscience, 35, 573–582.

Hamilton, L. S., Oganian, Y., Hall, J., & Chang, E. F. (2021). Parallel and distributed encoding of speech across human auditory cortex. Cell, 184, 4626–4639.e13.

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.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? Science, 298, 1569–1579.

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.

Heald, S. L. M. & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. Front. Syst. Neurosci., 8, 35.

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.

Herff, C., Heger, D., de Pesters, A., Telaar, D., Brunner, P., Schalk, G., & Schultz, T. (2015). Brain-to-text: decoding spoken phrases from phone representations in the brain. Front. Neurosci., 9, 217.

Hickok, G. (2007). Where is area spt? Accessed on May 26, 2024.

Hickok, G. (2014). The architecture of speech production and the role of the phoneme in speech processing. Lang. Cogn. Process., 29, 2–20.

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area spt. J. Cogn. Neurosci., 15, 673–682.

Hickok, G., Okada, K., & Serences, J. T. (2009). Area spt in the human planum temporale supports sensory-motor integration for speech processing. J. Neurophysiol., 101, 2725–2732.

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

Hickok, G., Venezia, J., & Teghipco, A. (2023). Beyond broca: neural architecture and evolution of a dual motor speech coordination system. Brain, 146, 1775–1790.

Hillis, A. E., Work, M., Barker, P. B., Jacobs, M. A., Breese, E. L., & Maurer, K. (2004). Re-examining the brain regions crucial for orchestrating speech articulation. Brain, 127, 1479–1487.

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.

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. Nature, 532, 453–458.

Hyde, M. (1997). The N1 response and its applications. Audiol. Neurootol., 2, 281–307.

Indefrey, P. & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. Cognition, 92, 101–144.

Ivanova, A. A., Hewitt, J., & Zaslavsky, N. (2021a). Probing artificial neural networks: insights from neuroscience.

Ivanova, A. A., Schrimpf, M., Anzellotti, S., Zaslavsky, N., Fedorenko, E., & Isik, L. (2021b). Beyond linear regression: mapping models in cognitive neuroscience should align with research goals.

Jacks, A. & Haley, K. L. (2015). Auditory masking effects on speech fluency in apraxia of speech and aphasia: Comparison to altered auditory feedback. J. Speech Lang. Hear. Res., 58, 1670–1686.

Jackson, R. L., Hoffman, P., Pobric, G., & Lambon Ralph, M. A. (2016). The semantic network at work and rest: Differential connectivity of anterior temporal lobe subregions. J. Neurosci., 36, 1490–1501.

Jahanshahi, M. & Hallett, M. (2003). The Bereitschaftspotential: Movement-Related Cortical Potentials. (Springer Science & Business Media).

Jankowski, M. M., Karayanni, M., Harpaz, M., Polterovich, A., & Nelken, I. (2023). A rapid anterior auditory processing stream through the Insulo-Parietal auditory field in the rat.

Johns, L. C., Rossell, S., Frith, C., Ahmad, F., Hemsley, D., Kuipers, E., & McGuire, P. K. (2001). Verbal self-monitoring and auditory verbal hallucinations in patients with schizophrenia. Psychol. Med., 31, 705–715.

Johnson, K. (2011). Acoustic and Auditory Phonetics. (John Wiley & Sons).

Jones, J. A. & Munhall, K. G. (2000). Perceptual calibration of F0 production: evidence from feedback perturbation. J. Acoust. Soc. Am., 108, 1246–1251.

Kalinowski, J. & Stuart, A. (1996). Stuttering amelioration at various auditory feedback delays and speech rates. Eur. J. Disord. Commun., 31, 259–269.

Karami, M., Nilipour, R., Barekatain, M., & Gaillard, W. D. (2020). Language representation and presurgical language mapping in pediatric epilepsy: A narrative review. Iran J Child Neurol, 14, 7–18.

Kearney, E. & Guenther, F. H. (2019). Articulating: The neural mechanisms of speech production. Lang Cogn Neurosci, 34, 1214–1229.

Kenward, M. G. & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. Biometrics, 53, 983–997.

Khalighinejad, B., Cruzatto da Silva, G., & Mesgarani, N. (2017). Dynamic encoding of acoustic features in neural responses to continuous speech. J. Neurosci., 37, 2176–2185.

Khalilian-Gourtani, A., Wang, R., Chen, X., Yu, L., Dugan, P., Friedman, D., Doyle, W., Devinsky, O., Wang, Y., & Flinker, A. (2022). A corollary discharge circuit in human speech.

Khanna, A. R., Muñoz, W., Kim, Y. J., Kfir, Y., Paulk, A. C., Jamali, M., Cai, J., Mustroph, M. L., Caprara, I., Hardstone, R., Mejdell, M., Meszéna, D., Zuckerman, A., Schweitzer, J., Cash, S., & Williams, Z. M. (2024). Single-neuronal elements of speech production in humans. Nature.

Kobayashi, S. & Ugawa, Y. (2013). Relationships between aphasia and apraxia. J. Neurol. Transl. Neurosci.

Kunii, N., Kamada, K., Ota, T., Kawai, K., & Saito, N. (2013). Characteristic profiles of high gamma activity and blood oxygenation level-dependent responses in various language areas. Neuroimage, 65, 242–249.

Kurteff, G. (2020). Modulation of neural responses to naturalistic speech production and perception. Master's thesis, The University of Texas at Austin.

Kurteff, G. L., Field, A. M., Asghar, S., Tyler-Kabara, E. C., Clarke, D., Weiner, H. L., Anderson, A. E., Watrous, A. J., Buchanan, R. J., Modur, P. N., & Hamilton, L. S. (2024). Processing of auditory feedback in perisylvian and insular cortex.

Kurteff, G. L., Lester-Smith, R. A., Martinez, A., Currens, N., Holder, J., Villarreal, C., Mercado, V. R., Truong, C., Huber, C., Pokharel, P., & Hamilton, L. S. (2023). Speaker-induced suppression in EEG during a naturalistic reading and listening task. J. Cogn. Neurosci., 35, 1538–1556.

Kurth, F., Zilles, K., Fox, P. T., Laird, A. R., & Eickhoff, S. B. (2010). A link between the systems: functional differentiation and integration within the human insula revealed by meta-analysis. Brain Struct. Funct., 214, 519–534.

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

Lachaux, J.-P., Axmacher, N., Mormann, F., Halgren, E., & Crone, N. E. (2012). High-frequency neural activity and human cognition: past, present and possible future of intracranial EEG research. Prog. Neurobiol., 98, 279–301.

Lah, S., Castles, A., & Smith, M. L. (2017). Reading in children with temporal lobe epilepsy: A systematic review. Epilepsy Behav., 68, 84–94.

Laiho, A., Elovaara, H., Kaisamatti, K., Luhtalampi, K., Talaskivi, L., Pohja, S., Routamo-Jaatela, K., & Vuorio, E. (2022). Stuttering interventions for children, adolescents, and adults: a systematic review as a part of clinical guidelines. J. Commun. Disord., 99, 106242.

Lakretz, Y., Ossmy, O., Friedmann, N., Mukamel, R., & Fried, I. (2021). Single-cell activity in human STG during perception of phonemes is organized according to manner of articulation. Neuroimage, 226, 117499.

Leonard, M. K., Cai, R., Babiak, M. C., Ren, A., & Chang, E. F. (2019). The peri-sylvian cortical network underlying single word repetition revealed by

electrocortical stimulation and direct neural recordings. Brain Lang., 193, 58–72.

Leonard, M. K., Gwilliams, L., Sellers, K. K., Chung, J. E., Xu, D., Mischler, G., Mesgarani, N., Welkenhuysen, M., Dutta, B., & Chang, E. F. (2023). Large-scale single-neuron speech sound encoding across the depth of human cortex. Nature.

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. J. Speech Lang. Hear. Res., 63, 3628–3642.

Levelt, W. J. M. (1993). Speaking: From Intention to Articulation. (MIT Press).

Levy, D. F., Silva, A. B., Scott, T. L., Liu, J. R., Harper, S., Zhao, L., Hullett, P. W., Kurteff, G., Wilson, S. M., Leonard, M. K., & Chang, E. F. (2023). Apraxia of speech with phonological alexia and agraphia following resection of the left middle precentral gyrus: illustrative case. J Neurosurg Case Lessons, 5.

Lewandowsky, M. (1912). Praktische Neurologie für Ärzte. (Springer-Verlag).

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychol. Rev., 74, 431–461.

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

Linke, R. & Schwegler, H. (2000). Convergent and complementary projections of the caudal paralaminar thalamic nuclei to rat temporal and insular cortex. Cereb. Cortex, 10, 753–771.

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

Luck, S. J. & Kappenman, E. S. (2013). The oxford handbook of event-related potential components. Oxford Library of Psychology. (New York, NY: Oxford University Press).

Manasco, H. (2013). The aphasias. In Introduction to Neurogenic Communication Disorders. pp. 71–103.

Mandelli, M. L., Caverzasi, E., Binney, R. J., Henry, M. L., Lobach, I., Block, N., Amirbekian, B., Dronkers, N., Miller, B. L., Henry, R. G., & Gorno-Tempini, M. L. (2014). Frontal white matter tracts sustaining speech production in primary progressive aphasia. J. Neurosci., 34, 9754–9767.

Manto, M., Bower, J. M., Conforto, A. B., Delgado-García, J. M., da Guarda, S. N. F., Gerwig, M., Habas, C., Hagura, N., Ivry, R. B., Mariën, P., Molinari, M., Naito, E., Nowak, D. A., Oulad Ben Taib, N., Pelisson, D., Tesche, C. D., Tilikete, C., & Timmann, D. (2012). Consensus paper: roles of the cerebellum in motor control–the diversity of ideas on cerebellar involvement in movement. Cerebellum, 11, 457–487.

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.

Massaro, D. W. & Chen, T. H. (2008). The motor theory of speech perception revisited. Psychon. Bull. Rev., 15, 453–7; discussion 458–62.

Matusz, P. J., Dikker, S., Huth, A. G., & Perrodin, C. (2019). Are we ready for real-world neuroscience? J. Cogn. Neurosci., 31, 327–338.

Max, L. & Daliri, A. (2019). Limited Pre-Speech auditory modulation in individuals who stutter: Data and hypotheses. J. Speech Lang. Hear. Res., 62, 3071–3084.

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.

Mercier, M. R., Dubarry, A.-S., Tadel, F., Avanzini, P., Axmacher, N., Cellier, D., Vecchio, M. D., Hamilton, L. S., Hermes, D., Kahana, M. J., Knight, R. T., Llorens, A., Megevand, P., Melloni, L., Miller, K. J., Piai, V., Puce, A., Ramsey, N. F., Schwiedrzik, C. M., Smith, S. E., Stolk, A., Swann, N. C., Vansteensel, M. J., Voytek, B., Wang, L., Lachaux, J.-P., & Oostenveld, R. (2022). Advances in human intracranial electroencephalography research, guidelines and good practices. Neuroimage, 260, 119438.

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

Metzger, S. L., Littlejohn, K. T., Silva, A. B., Moses, D. A., Seaton, M. P., Wang, R., Dougherty, M. E., Liu, J. R., Wu, P., Berger, M. A., Zhuravleva, I., Tu-Chan, A., Ganguly, K., Anumanchipalli, G. K., & Chang, E. F. (2023). A high-performance neuroprosthesis for speech decoding and avatar control. Nature.

Meyyappan, S., Rajan, A., Mangun, G. R., & Ding, M. (2021). Role of inferior frontal junction (IFJ) in the control of feature versus spatial attention. J. Neurosci., 41, 8065–8074.

Miozzo, M., Williams, A. C., McKhann, 2nd, G. M., & Hamberger, M. J. (2017). Topographical gradients of semantics and phonology revealed by temporal lobe stimulation. Hum. Brain Mapp., 38, 688–703.

Möddel, G., Lineweaver, T., Schuele, S. U., Reinholz, J., & Loddenkemper, T. (2009). Atypical language lateralization in epilepsy patients. Epilepsia, 50, 1505–1516.

Mollaei, F., Shiller, D. M., Baum, S. R., & Gracco, V. L. (2016). Sensorimotor control of vocal pitch and formant frequencies in parkinson's disease. Brain Res., 1646, 269–277.

Moore, D. R., Fuchs, P. A., Rees, A., Palmer, A., & Plack, C. J. (2010). The Oxford Handbook of Auditory Science: The Auditory Brain. (OUP Oxford).

Moses, D. A., Metzger, S. L., Liu, J. R., Anumanchipalli, G. K., Makin, J. G., Sun, P. F., Chartier, J., Dougherty, M. E., Liu, P. M., Abrams, G. M., Tu-Chan, A., Ganguly, K., & Chang, E. F. (2021). Neuroprosthesis for decoding speech in a paralyzed person with anarthria. N. Engl. J. Med., 385, 217–227.

Muller, L., Hamilton, L. S., Edwards, E., Bouchard, K. E., & Chang, E. F. (2016). Spatial resolution dependence on spectral frequency in human speech cortex electrocorticography. J. Neural Eng., 13, 056013.

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.

Nguyen, D., Isnard, J., & Kahane, P. (2022). Insular Epilepsies. (Cambridge University Press).

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.

Nourski, K. V., Steinschneider, M., Rhone, A. E., Kovach, C. K., Banks, M. I., Krause, B. M., Kawasaki, H., & Howard, M. A. (2021). Electrophysiology

of the human superior temporal sulcus during speech processing. Cereb. Cortex, 31, 1131–1148.

Oganian, Y., Bhaya-Grossman, I., Johnson, K., & Chang, E. F. (2023). Vowel and formant representation in the human auditory speech cortex. Neuron, 111, 2105–2118.e4.

Oganian, Y. & Chang, E. F. (2019). A speech envelope landmark for syllable encoding in human superior temporal gyrus. Sci. Adv., 5, eaay6279.

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

Ozker, M., Doyle, W., Devinsky, O., & Flinker, A. (2022). A cortical network processes auditory error signals during human speech production to maintain fluency. PLoS Biol., 20, e3001493.

Ozker, M., Yu, L., Dugan, P., Doyle, W., Friedman, D., Devinsky, O., & Flinker, A. (2024). Speech-induced suppression and vocal feedback sensitivity in human cortex. bioRxiv.

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., Kim, H. E., Breska, A., Saxena, A., & Ivry, R. (2021). Differential effects of cerebellar degeneration on feedforward versus feedback control across speech and reaching movements. J. Neurosci., 41, 8779–8789.

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.

Patidar, Y., Gupta, M., Khwaja, G. A., Chowdhury, D., Batra, A., & Dasgupta, A. (2013). A case of crossed aphasia with apraxia of speech. Ann. Indian Acad. Neurol., 16, 428–431.

Penfield, W. & Roberts, L. (1959). Speech and Brain Mechanisms. (Princeton University Press).

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.

Poeppel, D. & Monahan, P. J. (2011). Feedforward and feedback in speech perception: Revisiting analysis by synthesis. Language and Cognitive Processes, 26, 935–951.

Pratt, H., Bleich, N., & Mittelman, N. (2005). The composite N1 component to gaps in noise. Clin. Neurophysiol., 116, 2648–2663.

Pribram, K. H., Rosner, B. S., & Rosenblith, W. A. (1954). Electrical responses to acoustic clicks in monkey: extent of neocortex activated. J. Neurophysiol., 17, 336–344.

Quabs, J., Caspers, S., Schöne, C., Mohlberg, H., Bludau, S., Dickscheid, T., & Amunts, K. (2022). Cytoarchitecture, probability maps and segregation of the human insula. Neuroimage, 260, 119453.

Rabbani, Q., Milsap, G., & Crone, N. E. (2019). The potential for a speech Brain-Computer interface using chronic electrocorticography. Neurotherapeutics, 16, 144–165.

Railo, H., Nokelainen, N., Savolainen, S., & Kaasinen, V. (2020). Deficits in monitoring self-produced speech in parkinson's disease. Clin. Neurophysiol., 131, 2140–2147.

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.

Ray, S., Crone, N. E., Niebur, E., Franaszczuk, P. J., & Hsiao, S. S. (2008). Neural correlates of high-gamma oscillations (60-200 hz) in macaque local field potentials and their potential implications in electrocorticography. J. Neurosci., 28, 11526–11536.

Ray, S. & Maunsell, J. H. R. (2011). Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. PLoS Biol., 9, e1000610.

Remedios, R., Logothetis, N. K., & Kayser, C. (2009). An auditory region in the primate insular cortex responding preferentially to vocal communication sounds. J. Neurosci., 29, 1034–1045.

Riès, S., Janssen, N., Burle, B., & Alario, F.-X. (2013). Response-locked brain dynamics of word production. PLoS One, 8, e58197.

Ries, S. K., Pinet, S., Nozari, N. B., & Knight, R. T. (2021). Characterizing multi-word speech production using event-related potentials. Psychophysiology, 58, e13788.

Rodgers, K. M., Benison, A. M., Klein, A., & Barth, D. S. (2008). Auditory, somatosensory, and multisensory insular cortex in the rat. Cereb. Cortex, 18, 2941–2951.

Rolls, E. T., Rauschecker, J. P., Deco, G., Huang, C.-C., & Feng, J. (2023). Auditory cortical connectivity in humans. Cereb. Cortex, 33, 6207–6227.

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: an event-related potential index of perceived onsets in continuous speech. Nat. Neurosci., 5, 700–703.

Sawatari, H., Tanaka, Y., Takemoto, M., Nishimura, M., Hasegawa, K., Saitoh, K., & Song, W.-J. (2011). Identification and characterization of an insular auditory field in mice. Eur. J. Neurosci., 34, 1944–1952.

Scheerer, N. E. & Jones, J. A. (2014). The predictability of frequency-altered auditory feedback changes the weighting of feedback and feedforward input for speech motor control. Eur. J. Neurosci., 40, 3793–3806.

Scherg, M. & Picton, T. W. (1991). Separation and identification of event-related potential components by brain electric source analysis. Electroencephalogr. Clin. Neurophysiol. Suppl., 42, 24–37.

Schneider, D. M., Nelson, A., & Mooney, R. (2014). A synaptic and circuit basis for corollary discharge in the auditory cortex. Nature, 513, 189–194.

Schneider, D. M., Sundararajan, J., & Mooney, R. (2018). A cortical filter that learns to suppress the acoustic consequences of movement. Nature, 561, 391–395.

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.

Shadmehr, R. & Krakauer, J. W. (2008). A computational neuroanatomy for motor control. Exp. Brain Res., 185, 359–381.

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

Silva, A. B., Liu, J. R., Zhao, L., Levy, D. F., Scott, T. L., & Chang, E. F. (2022). A neurosurgical functional dissection of the middle precentral gyrus during speech production. J. Neurosci., 42, 8416–8426.

Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. Brain Lang., 164, 77–105.

St. Louis, K. O. & Ruscello, D. M. (1981). Oral Speech Mechanism Screening Examination (OSMSE). (University Park Press).

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

Stuart, A., Kalinowski, J., Rastatter, M. P., & Lynch, K. (2002). Effect of delayed auditory feedback on normal speakers at two speech rates. J. Acoust. Soc. Am., 111, 2237–2241.

Sudakov, K., MacLean, P. D., Reeves, A., & Marino, R. (1971). Unit study of exteroceptive inputs to claustrocortex in awake, sitting, squirrel monkey. Brain Res., 28, 19–34.

Sun, Y. & Poeppel, D. (2023). Syllables and their beginnings have a special role in the mental lexicon. Proc. Natl. Acad. Sci. U. S. A., 120, e2215710120.

Takemoto, M., Hasegawa, K., Nishimura, M., & Song, W.-J. (2014). The insular auditory field receives input from the lemniscal subdivision of the auditory thalamus in mice. J. Comp. Neurol., 522, 1373–1389.

Tang, C., Hamilton, L. S., & Chang, E. F. (2017). Intonational speech prosody encoding in the human auditory cortex. Science, 357, 797–801.

Tate, M. C., Herbet, G., Moritz-Gasser, S., Tate, J. E., & Duffau, H. (2014). Probabilistic map of critical functional regions of the human cerebral cortex: Broca's area revisited. Brain, 137, 2773–2782.

Theunissen, F. E., Sen, K., & Doupe, A. J. (2000). Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J. Neurosci., 20, 2315–2331.

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.

Towle, V. L., Yoon, H.-A., Castelle, M., Edgar, J. C., Biassou, N. M., Frim, D. M., Spire, J.-P., & Kohrman, M. H. (2008). ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. Brain, 131, 2013–2027.

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.

Turin, G. (1960). An introduction to matched filters. IRE Transactions on Information Theory, 6, 311–329.

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.

Walsh, B. & Smith, A. (2002). Articulatory movements in adolescents: evidence for protracted development of speech motor control processes. J. Speech Lang. Hear. Res., 45, 1119–1133.

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.

Wohlert, A. B. (1993). Event-related brain potentials preceding speech and nonspeech oral movements of varying complexity. J. Speech Hear. Res., 36, 897–905.

Wolpaw, J. R. & Penry, J. K. (1977). Hemispheric differences in the auditory evoked response. Electroencephalogr. Clin. Neurophysiol., 43, 99–102.

Woolnough, O., Forseth, K. J., Rollo, P. S., & Tandon, N. (2019). Uncovering the functional anatomy of the human insula during speech. Elife, 8.

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

Wright, S. (1921). Correlation and causation. J. Agric. Res., 20, 557.

Yoshida, K., Kaji, R., Hamano, T., Kohara, N., Kimura, J., & Iizuka, T. (1999). Cortical distribution of bereitschaftspotential and negative slope potential preceding mouth-opening movements in humans. Arch. Oral Biol., 44, 183–190.

Youngerman, B. E., Khan, F. A., & McKhann, G. M. (2019). Stereoelectroencephalography in epilepsy, cognitive neurophysiology, and psychiatric disease: safety, efficacy, and place in therapy. Neuropsychiatr. Dis. Treat., 15, 1701–1716.

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

Zhang, Y., Zhou, W., Wang, S., Zhou, Q., Wang, H., Zhang, B., Huang, J., Hong, B., & Wang, X. (2018). The roles of subdivisions of human insula in emotion perception and auditory processing. Cereb. Cortex, 29, 517–528.

Zhao, L., Silva, A. B., Kurteff, G. L., & Chang, E. F. (2023). Inhibitory control of speech production in the human premotor frontal cortex.

Zhao, S. & Rudzicz, F. (2015). Classifying phonological categories in imagined and articulated speech. In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 992–996.

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.

Index

Vita

G. Lynn Kurteff was born in Monterey, California. They received a BA in Linguistics & Psychology from UC Berkeley in 2015. At Berkeley they were mentored by Drs. Keith Johnson, Lev D. Michael, and Joseph J. Campos. They next conducted research in the laboratory of Dr. Edward F. Chang at UCSF in 2016, where they studied recovery from aphasia after neurosurgery and the neural basis of speech and syntax via cortical stimulation mapping. They joined a joint MS-PhD program in Speech, Language, and Hearing Sciences at UT Austin under the supervision of Liberty S. Hamilton in 2018. They received their MSSLHS from UT Austin in 2020. As a speech-language pathologist, they are interested in aphasia, apraxia of speech, gender-affirming voice therapy, and brain-computer interfaces. Lynn was awarded the William Orr Dingwall Foundation's "Foundations of Language" fellowship for their dissertation research. After their doctorate, they plan to work as a postdoctoral researcher and pursue clinical certification as a speech-language pathologist. Lynn identifies as transfeminine and uses they/them pronouns.

Permanent address: 26126 Camino Real, Carmel-by-the-Sea, CA

This dissertation 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.