## Behavioral/Cognitive

# Spatiotemporal Mapping of Auditory Onsets during Speech Production

# Garret Lynn Kurteff,<sup>1</sup> Alyssa M. Field,<sup>1</sup> Saman Asghar,<sup>1,5</sup> Elizabeth C. Tyler-Kabara,<sup>2,3</sup> Dave Clarke,<sup>2,3,4</sup> Howard L. Weiner,<sup>5</sup> Anne E. Anderson,<sup>6</sup> Andrew J. Watrous,<sup>5</sup> Robert J. Buchanan,<sup>2</sup> Pradeep N. Modur,<sup>4</sup> and <sup>©</sup>Liberty S. Hamilton<sup>1,4\*</sup>

<sup>1</sup>Department of Speech, Language, and Hearing Sciences, Moody College of Communication, The University of Texas at Austin, Austin, Texas 78712, Departments of <sup>2</sup>Neurosurgery <sup>3</sup>Pediatrics, and <sup>4</sup>Neurology, Dell Medical School, The University of Texas at Austin, Austin, Texas 78712, and Departments of <sup>5</sup>Neurosurgery and <sup>6</sup>Pediatrics, Baylor College of Medicine, Houston, Texas 77030

The human auditory cortex is organized according to the timing and spectral characteristics of speech sounds during speech perception. During listening, the posterior superior temporal gyrus is organized according to onset responses, which segment acoustic boundaries in speech, and sustained responses, which further process phonological content. When we speak, the auditory system is actively processing the sound of our own voice to detect and correct speech errors in real time. This manifests in neural recordings as suppression of auditory responses during speech production compared with perception, but whether this differentially affects the onset and sustained temporal profiles is not known. Here, we investigated this question using intracranial EEG recorded from seventeen pediatric, adolescent, and adult patients with medication-resistant epilepsy while they performed a reading/listening task. We identified onset and sustained responses to speech in the bilateral auditory cortex and observed a selective suppression of onset responses during speech production. We conclude that onset responses provide a temporal landmark during speech perception that is redundant with forward prediction during speech production and are therefore suppressed. Phonological feature tuning in these "onset suppression" electrodes remained stable between perception and production. Notably, auditory onset responses and phonological feature tuning were present in the posterior insula during both speech perception and production, suggesting an anatomically and functionally separate auditory processing zone that we believe to be involved in multisensory integration during speech perception and feedback control.

Key words: auditory perception; intracranial electrophysiology; language; speech; speech motor control; speech production

## Significance Statement

Specific neural populations in the auditory cortex preferentially respond to the onset of speech sounds. These "onset responses" aid in perceiving boundaries in continuous speech. We recorded neural responses from patients with intracranial electrodes during a speaking and listening task to investigate the role of onset responses in speech production. Onset responses were present in the auditory cortex during listening, but absent during speaking. On the other hand, onset responses were observed in the insula during both conditions, suggesting a different functional role for the insula in auditory feedback processing. These findings extend our knowledge of how different parts of the brain involved in feedback control operate during speech production by identifying two functionally and anatomically distinct patterns of activity.

Received June 11, 2024; revised July 31, 2024; accepted Oct. 8, 2024.

Author contributions: G.L.K. and L.S.H. designed research; G.L.K., A.M.F., S.A., E.C.T-K., D.C., H.L.W., A.E.A., A.J.W., R.J.B., P.N.M., and L.S.H. performed research; G.L.K., A.M.F., and L.S.H. analyzed data; G.L.K. and L.S.H. wrote the paper.

We thank the patients at Dell Children's Medical Center, Texas Children's Medical Center, and Dell Seton Medical Center for volunteering time during their hospital stay to participate in this research. We thank the members of the clinical team at Dell Children's who assisted in data collection and/or patient referral: Timothy George, MD; Winson Ho, MD; Nancy Nussbaum, PhD; Rosario DeLeon, PhD; William Andy Schraegle, PhD; Fred Perkins, MD; Karen Keough, MD; Aaron Cardon, MD; Karen Skjei, MD; Teresa Ontiveros, RN, MSN; Cassidy Wink, RN; and Bethany Hepokoski, RN. We thank Maansi Desai, PhD, for her assistance with data collection and manuscript edits. Lastly, we thank Stephanie Ries, PhD; Maya Henry, PhD, CCC-SLP; Rosemary A. Lester-Smith PhD, CCC-SLP; and Jun Wang, PhD, for their feedback on early manuscript drafts. This work was supported by the National Institutes of Health National Institute on Deafness and Other Communication Disorders (1R01-DC018579, to L.S.H.) and by a William Orr Dingwall Foundation 100025712 Dissertation Fellowship (to G.L.K.).

\*Lead contact.

The authors declare no competing interests.

Correspondence should be addressed to Liberty S. Hamilton at liberty.hamilton@austin.utexas.edu. https://doi.org/10.1523/JNEUROSCI.1109-24.2024

Copyright © 2024 the authors

## Introduction

Organization of speech cortex during listening and speaking During speech perception, the auditory cortex forms linguistic percepts from incoming acoustic information according to its spectral and temporal characteristics (Appelbaum, 1996). This involves determining the timing of important acoustic events, such as the onset of a sentence or a phrase. Following the detection of these onsets, the content of the sentence must be determined. The posterior superior temporal gyrus (pSTG)-including the classic "Wernicke's area"—is critical to this process, but until recently, little was known about its functional organization. Recent studies have shown that the pSTG can be divided into two regions: a posterior region that is selective to auditory onsets and a more anterior region with sustained activity (Hamilton et al., 2018). These auditory onset responses are critical for segmenting ongoing acoustic boundaries, including sentence, phrase, and syllabic boundaries (Hamilton et al., 2018; Oganian and Chang, 2019). Within these larger regions, the brain encodes phonological information both at the level of local field potentials (LFPs; Mesgarani et al., 2014) and single neurons (Lakretz et al., 2021; Leonard et al., 2023). Specifically, subregions of the STG may respond to phonemes with the same manner of articulation, a linguistic feature that describes the degree of constriction and airflow through the vocal tract (Hayes, 2011; Mesgarani et al., 2014). Phonetically selective regions of the STG represent these phonetic distinctions while invariant to pitch and other acoustic characteristics, which are processed by distinct circuits (Tang et al., 2017; Hamilton et al., 2018). Still, it is unclear whether these organizing principles of the auditory system are affected during speech production, when motor and sensory feedback are also at play.

During speech production, a motor program for the intended utterance is generated prior to articulation from linguistically segmented information (Levelt, 1993; Indefrey and Levelt, 2004; Kawamoto et al., 2014). Also generated is an efference copy, a set of expectations regarding upcoming sensory feedback for use in speech motor control, which contains information about temporal/linguistic landmarks in that feedback (Houde and Nagarajan, 2011; Niziolek et al., 2013; Schneider et al., 2014; Guenther, 2016). During speech perception, onset responses are theorized to have utility in segmenting the auditory stream into discrete linguistic units, such as phrases and sentences. Onset responses may then be differentially processed during speech production, based on the general observation that auditory information is processed differently during speaking compared with listening (Creutzfeldt et al., 1989; Houde et al., 2002; Towle et al., 2008; Cogan et al., 2014; Nourski et al., 2021). Speaker-induced suppression (SIS) is a phenomenon in which self-generated speech generates a lower-amplitude neural response than externally generated speech (Martikainen et al., 2005; Flinker et al., 2010; Behroozmand and Larson, 2011; Ozker et al., 2024). Intracranial recordings have localized SIS to the posterior STG (Chang et al., 2013; Ozker et al., 2024). While this is a similar localization to the onset zone, it is unclear if onset responses are implicated in SIS. If onset responses encode the temporal landmarks of speech, they may then be suppressed as a redundant processing component during speech production.

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. Recent research from our group used scalp EEG recordings to demonstrate that responses to continuous sentences are suppressed during production compared with the 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. SIS was primarily isolated to the N1 and P2 components in this study; early-onset neural responses were observed at acoustic edges with high temporal modulation (Luck, 2014), making these components share characteristics with onset responses observed using invasive recordings. Other top-down cognitive processes can affect speech processing as well, such as expectations about upcoming stimuli evidenced in both speech production (Scheerer and Jones, 2014; Fjaellingsdal et al., 2020; Lester-Smith et al., 2020) and speech perception (Astheimer and 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 (Shadmehr and Krakauer, 2008; Gonzalez Castro et al., 2014), the link between speaker-induced suppression and more general top-down expectation is not well established. The present study aims to investigate how established onset and sustained temporal modulation profiles in the auditory cortex interact with cortical suppression during speech production, linguistic feature representation, and top-down expectancy effects.

## 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). The insula is difficult to record from using several popular neuroimaging techniques due to its placement deep in the Sylvian fissure (Remedios et al., 2009; Mercier et al., 2022). 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). In speech, the insula conventionally plays a role in prearticulatory motor coordination (Dronkers, 1996), but roles in auditory processing have also been documented. Recent work using sEEG reported the insula to be more active for self-generated speech when compared with externally generated speech (Woolnough et al., 2019), an opposite trend to the cortical suppression of self-generated speech observed in the auditory cortex described above. Research in animal models has identified low-latency auditory fields in the posterior insula (Remedios et al., 2009; Sawatari et al., 2011; Takemoto et al., 2014). The human literature also suggests functional distinctions between anterior and posterior insula for emotional vocalizations (Zhang et al., 2018), and lesions to the insula may result in auditory agnosia (Habib et al., 1995; Bamiou et al., 2003). Still, the phonological tuning and temporal specificity of the insular cortex to speech stimuli have not been documented, unlike the temporal cortex. We 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.

# How do the organizational principles of the auditory system change during self-produced speech?

To address how temporal (onset, sustained) and phonological organizational characteristics of the auditory system change during speech production due to suppression of self-generated feedback, we 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 and then passively listened to a playback of their reading to identify potential changes in local field potential recorded by the implanted electrodes. Our first goal was to identify if previously identified onset and sustained response profiles in the auditory cortex (Hamilton et al., 2018) were also present during speech production. We also investigated how linguistic feature tuning changes at individual electrodes during speech production versus perception. Additionally, we 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, we used unsupervised clustering to identify an auditory-responsive region in the posterior insula and conducted similar analyses within to compare the auditory processing of the region to the temporal lobe primary and nonprimary auditory cortex. Our results have implications for understanding important auditory-motor interactions during natural human communication.

## Materials and Methods

#### Participants

Seventeen individuals (sex, 9 F; age,  $16.6 \pm 6.4$ ; range, 8-37 years; race/ ethnicity, 8 Hispanic/Latino, 6 White, 1 Asian, 2 multiracial) undergoing intracranial monitoring of seizure activity via stereoelectroencephalography (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 Extended Data Table 1-1. 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.

#### Neural data acquisition

Intracranial sEEG and electrocorticography (ECoG) data from a total of 2,044 electrodes across subjects were recorded continuously via the epilepsy monitoring teams using a Natus Quantum headbox (Natus Medical Incorporated). At Texas Children's Hospital, sEEG depths (AdTech Spencer Probe Depth Electrodes, 5 mm spacing, 0.86 mm diameter, 4-16 contacts per device), strip electrodes (AdTech), and grids (AdTech custom order, 5 mm spacing, 8×8 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, 0.8 mm diameter, 3.5 mm 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). Speaker (perceived) and microphone (produced) audio were also recorded via RZ2 at 22 kHz to circumvent the downsampling of audio by the clinical recording system. At the other two recording locations, the use of a dedicated research recording system was not possible due to clinical constraints; instead, the auditory stimuli from the iPad were recorded directly on the clinical system using an audio splitter cable. Simultaneous high-resolution audio was recorded for both speaking and playback using an external microphone and a second splitter cable from the iPad both plugged into a MOTU M4 USB audio interface (MOTU) which is 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.

#### Data preprocessing

Data were preprocessed offline using a combination of custom MATLAB scripts and custom Python scripts built off the MNE-python software package (Gramfort et al., 2013). First, data were notch filtered at 60/120/180 Hz to remove line noise, and then bad channels were manually inspected and rejected. 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 multiunit neuronal firing (Ray and Maunsell, 2011), was extracted via Hilbert transform (eight bands, log spaced, Gaussian kernel, 70–150 Hz). Lastly, the eight-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.

#### Electrode localization

Electrodes' locations were registered in the three-dimensional MNI coordinate space (Evans et al., 1993). Electrodes were localized through coregistration of an individual subject's T1 MRI scan with their computed tomography 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 intrasylvian structures such as the insula and Heschl's gyrus (HG) 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 ROI before projection to inflated space. Additionally, any depth electrodes >4 mm 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 >4 mm apart from each other (Muller et al., 2016). Electrodes >4 mm from the cortical surface, while excluded from visualization, were included in analyses if they contained a robust response [p < 0.05 for bootstrap procedure,  $r \ge 0.1$  for temporal receptive field (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 previously (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.

#### Experimental design and statistical analyses

Overt reading and playback 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. Both perception and production trials were preceded by a fixation cross and broadband click tone (Fig. 3A). Production trials consisted of participants overtly reading a sentence, and 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. This task is identical to the one used previously (see Kurteff et al., 2023, for an analysis of this task in noninvasive scalp EEG).

For this study, 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, and 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  $142\pm61$  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 1,000 ms 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 mm 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 the analysis. Lastly, while data were common average referenced in analysis, the data were repreprocessed using a bipolar reference, and any electrodes with a markedly different response when the referencing method was changed were excluded from the analysis. All electrodes rejected through manual inspection of imaging were discussed and agreed upon by three of the authors (G.L.K., A.M.F., and L.S.H.). Electrodes above the significance threshold (p > 0.05) for both perception and production, as determined by the bootstrap procedure described below, were excluded from convex non-negative matrix factorization (cNMF) clustering if the electrode also had a low correlation during the multivariate TRF (mTRF) modeling procedure (r < 0.1). 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.

#### Speech motor control task

A subset of six participants (TC6, DC7, DC10, DC13, DC16, and 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 as follows: "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 and Ruscello, 1981). Each movement was repeated three 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 which was used to align high-resolution task audio with clinical recordings, identified the timing of these clicks and assisted in the generation of the event files.

### Event-related potential (ERP) analysis

We 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 and Liberman, 2008) was used to automatically generate Praat TextGrids (Boersma and Weenink, 2013) using a transcript generated by the iPad log file. Automatically generated TextGrids were checked for accuracy by the first author (G.L.K.). Event files containing start and stop times for each phoneme, word, and sentence, as well as information about the 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 s relative to sentence onset, with the negative window of interest intending to capture any prearticulatory activity related to speech production (Chartier et al., 2018).

Electrode significance was determined by bootstrap *t* test with 1,000 iterations comparing activity during the stimulus to randomly selected interstimulus-interval activity; bootstrapped significance for perception and production activity were calculated separately to identify electrodes that may be selectively responsive to either perceptual or production stimuli. For the bootstrap procedure, we averaged activity 5-550 ms after sentence onset and compared that to average activity during a silent 400-600 ms after the intertrial click as a control. The control time window was selected to not include potential evoked responses from the click sound but still be in the 1,000 ms window between the click sound and stimulus presentation. A similar procedure was used to calculate the significance for the consistent-inconsistent playback contrast (same time windows used). Bootstrap significance for the speech motor control task used activity 500-1,000 ms after the click sound played when text instructions were displayed to avoid including evoked responses to the click sound itself in the procedure. Because there were no intertrial click sounds in the speech motor control task with the click instead marking the display of instructions, activity -500 to 0 ms prior to the click sound was used as the control interval.

In addition to suppression, we were interested to see how onset responses change between speaking and listening. To quantify the presence of an onset response at a particular electrode, we looked at the first 300 ms 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 we required an onset response to begin in the first 300 ms of activity after sentence onset, we 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."

## 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, we employed convex non-negative matrix factorization (cNMF, Fig. 5; Ding et al., 2010). This is an unsupervised clustering technique that reveals the underlying statistical structure of datasets and has previously been used by our 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). We use a similar approach to these papers, summarized by the following equations:

$$X \approx \hat{X} = FG^{\top},$$
$$X_{\text{p,n}} \approx \frac{1}{t} \sum_{n=-1}^{n=1} H \gamma_{p,n} = FG^{\top}$$

where *X* is the 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 from -1 to +2 s relative to sentence onset. Epochs t were averaged within their response type and then concatenated; possible response types were production onset, perception (playback) onset, and intertrial click onset. Our 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, we concatenated additional epoch averages corresponding to the 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. We concatenated ERPs based on the response to production onset, perception (playback) onset, and click onset. We also incorporated information about expected versus unexpected playback as well as presentation of the visual cue in separate supplemental analyses, but these did not significantly alter the clusters observed. Our final concatenation resulted in a matrix X of  $n^*3$  samples (production epochs, perception epochs, and 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, and second, identifying a point at which adding additional clusters resulted in redundant average responses across clusters. We 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 was explained. The average response for each of the k = 9 clusters is provided in Figure 6.

### Suppression index calculation

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 0-1 s 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 5. The suppression index (*SI*), or degree of suppression during speaking as compared with 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 follows:

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

where *SI* of electrode *n* is the difference of high-gamma activity during speaking (*HyS*) subtracted from high-gamma activity during listening (*HyL*) divided by the sum of high-gamma activity during speaking and listening in the first 1 s after the acoustic onset of the sentence. A positive *SI* means that activity was greater during listening as compared with speaking, whereas a negative *SI* means activity was greater during speaking, whereas a negative *SI* means activity was greater during speaking, whereas a negative *SI* means activity was greater during speaking compared with listening. An *SI* of zero would reflect no difference between conditions.

### Linear mixed-effects modeling

Linear mixed-effects (LME) models were fit using the package lmertest (Kuznetsova et al., 2017) in R at several points in the analysis to quantify trends in the data. We chose LME as our statistical testing framework due to its ability to regress across within- and between-subject variability, facilitating generalization across subjects. The general equation takes the following form:

$$y = X\beta + Zu + \epsilon,$$

where  $\beta$  represents fixed-effects parameters, *u* represents random effects, and  $\varepsilon$  represents error. The first LME reported in this paper was used to quantify differences between suppression observed in onset and sustained responses. Suppression index (see above) was used as the response variable with a 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 calculation above: onset SI was calculated between 0 and 750 ms, and sustained SI was calculated between 1,000 and 1,750 ms after sentence onset. We chose these windows based on the average duration of the onset response across all electrodes and chose to make the sustained time window noncontiguous with the onset window to prevent extraneous activity from longer onset responses erroneously being factored as sustained activity in the model. We reported the contrast in the estimated marginal mean (EMM) SI of the two windows. We 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 and Yekutieli, 2001). The second LME reported in this paper was used to quantify response latency within three regions of interest: primary auditory [HG, planum temporale (PT)], nonprimary auditory [STG, superior temporal sulcus (STS)], and posterior and inferior insula. Peak latency values for the onset response (described above) were used as the response variable with ROI (three-way categorical) as a fixed effect and subject as a random effect [in R: peak\_latency ~ roi + (1|subject)]. We reported the EMM peak latencies of the three ROIs as well as their contrasts. The third LME reported in this paper was used to quantify the mTRF 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., 2021). For this LME model, the linear correlation coefficients between  $\widehat{H\gamma}$  and  $H\gamma$ were used as the response variable with model features (i.e., full vs ablated) as a fixed effect and subject and channel as a random effect [in R:  $r \sim \text{model} + (1|\text{subject}) + (1|\text{channel})$ ]. We chose to include channel as a random effect here as we 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 introduced multiple comparisons bias into our analysis. We reported 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. Contrast significance for all LMEs is calculated using F tests with Kenward–Roger approximation with n degrees of freedom specified, where n is the length of matrix X (Kenward and Roger, 1997).

### Multivariate temporal receptive field (mTRF) modeling

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

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

where  $\hat{y}_n(t)$  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  $\tau$  is the time delay between the stimulus and neural activity. For this study, we constructed a stimulus matrix using binary representations of phonological features derived from the phonetic transcript of the task (Fig. 7A). Also included in our stimulus matrix were binary representations of the experimental manipulation present in the current trial (perception vs production trial; followed by consistent playback vs inconsistent playback) in a method similar to our prior work (Kurteff et al., 2023). We fit separate models to predict the high-gamma response in each channel using time delays of -0.3 to 0.5 s. 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 prearticulatory 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 repeating this process 10 times. The ridge parameter was selected at the value that provided the highest average correlation performance across all 10 iterations. Ridge parameters between 10<sup>2</sup> and 10<sup>8</sup> 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. The significance of these correlations was obtained through a permutation t test procedure with 100 iterations in which the training data were shuffled in chunks to remove the relationship between the stimulus and response.

# Results

To examine potential differences in neural processing during speech production and perception, we acquired data from 17 pediatric, adolescent, and adult participants (sex, 9 F; age,  $16.6 \pm 6.4$ ; range, 8-37 years; Table 1) surgically implanted with intracranial 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 (Fig. 3*A*). For all analyses, we extracted the high-gamma analytic amplitude of the local field potentials (Lachaux et al., 2012), which has been shown to correlate with single- and multiunit neuronal firing (Ray and Maunsell, 2011) and tracks both acoustic and phonological characteristics of speech (Mesgarani et al., 2014; Oganian et al., 2023). Based on prior work, we expected to observe strong onset and sustained responses

Table T. Participant demodrat	phics	demograp	pant	Partici	1.	Table
-------------------------------	-------	----------	------	---------	----	-------

Participant	Age	Sex	Seizure focus
TC1	9	М	Left temporoparietal—occipital
DC2	14	М	Right hemisphere
TC3	19	F	Left temporal
DC4	21	М	No access to record
DC5	19	М	Right frontal
TC6*	14	F	Right sensory frontal
DC7*	20	М	Right temporal
DC8	16	F	No strong localization
DS9	37	М	Left temporal
DC10*	14	Х	Left temporal
DC11	13	М	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
DC17*	17	F	Left frontal

Table of age, sex, and seizure localization for each participant in the study. Subjects marked with an asterisk (\*) completed the supplementary speech motor control ask. Subjects in red were excluded from analysis due to a diagnosis of tuberous sclerosis complex. M, male; F, female; X, participant declined to disclose. Letters in identifiers reflect the recording site: TC, Texas Children's Hospital; DC, Dell Children's Medical Center; DS, Dell Ston Medical Center.

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 and Chang, 2014; Chartier et al., 2018). Additionally, our task design allowed us 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. The mean reaction time for reading trials (from which playback trials were generated) across participants was  $279 \pm 161$  ms; neural data were analyzed relative to the acoustic onset of the first phoneme regardless of individual trials' reaction times.

We recorded from a total of 2,044 sEEG depth electrodes implanted in the perisylvian cortex and insula (Fig. 1, Extended Data Fig. 1-1). This included coverage of speech-responsive areas of the lateral superior temporal gyrus, but also within the depths of the superior temporal sulcus, primary auditory cortex, and surrounding regions of the temporal plane. Single-electrode responses can be visualized on a 3D brain in an interactive webviewer at https://hamiltonlabut.github.io/kurteff2024/.

# Onset responses are selectively suppressed during speech production

To examine differences between speech perception and production on individual electrodes, we plotted event-related highgamma responses for speech perception and production trials relative to the beginning of the acoustic onset (the first phoneme) of the sentence. Prior research has demonstrated auditory onset responses are not limited to appearing at sentence onset but can appear at any point in an auditory stimulus following a >200 ms period of silence (Hamilton et al., 2018); we chose to focus on sentence onsets in our analysis as they are the most frequently elicited silence-to-speech transitions in our task. We 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 (Fig. 2B; bootstrap t test, p < 0.05). We quantified individual electrodes' selectivity to speech production or perception by calculating a suppression index (SI; see Materials and Methods). An SI > 0 reflects higher activity during



**Figure 1.** Coverage map. Individual electrodes for all included subjects with imaging (n = 15; excluding TC1 and DC4) plotted on the cvs\_avg35\_inMNI152 atlas brain, color-coded by anatomical region of interest. The cortical surface inflated for better visualization of insular electrodes. Electrode visualization in native subject space is available in Extended Data Figure 1-1.

listening compared with speaking, and an SI < 0 reflects higher activity during speaking compared with listening (Fig. 2*C*).

We observed single electrodes with selective responses to speech perception in bilateral Heschl's gyrus and STG (Fig. 2D). In addition, 51.4% of electrodes in STG (n = 70) and 100% of electrodes in Heschl's gyrus (n = 13) responded significantly to speech perception 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 the primary and nonprimary auditory cortex, onset responses were notably absent during speech production, while sustained responses remained relatively unsuppressed (Estimated marginal mean<sub>onset-sustained</sub> SI = 0.153; p < 0.001;  $t_{(77)} =$ 3.53). Electrodes in the primary sensorimotor cortex were typically more production-selective, in line with conventional localization of sensorimotor control of speech (Penfield and Roberts, 1959; Bouchard et al., 2013; Guenther, 2016). This pattern of responses demonstrates selective suppression of onset responses during speech production in the 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 and Nagarajan, 2011; Tourville and Guenther, 2011).

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

The ability of sEEG to obtain high-resolution recordings of the 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 lesionbased analyses may suffer from vasculature-related confounds in isolating insular responses (Hillis et al., 2004). Here, we present high spatiotemporal resolution recordings from the human insula and identify a functional response profile localized to this region.

While onset responses to speech perception were mostly confined to the 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 a similar fashion to auditory electrodes. Unlike auditory electrodes, however, posterior insular electrodes also showed robust onset responses during speech production (Fig. 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 (Fig. 3). We hypothesized that such responses might reflect a relationship to articulatory motor control or somatosensory processes, which prompted us to trial a nonspeech motor control task in a subset of our participants (n = 6; Table 1). 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 the 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 and 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



**Figure 2.** Auditory onset responses are suppressed during speech production. *A*, Schematic of reading and listening task. Participants read a sentence aloud (purple) and then passively listened to a playback of themselves reading the sentence (green). Pink spikes in the beginning and middle of the audio waveform indicate intertrial 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 (Figs. 3 and 4, respectively). 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; the dark gray indicates sulci while the 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 an anatomical region of interest in primary and nonprimary auditory cortex. Brackets indicate significance (\* = p < 0.05; \*\* = p < 0.01). Additional single-subject electrode profiles are shown in Extended Data Figures 2-1 and 2-2. 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.

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 (Fig. 3*E*). We 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 ms postclick), only 2% of insula electrodes (n = 49) significantly respond 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



**Figure 3.** A functional region of interest in the 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; the dark gray indicates sulci while the light gray indicates gyri. The black outline on the template brain highlights the functional region of interest in the posterior insula with anatomical structures labeled. Electrode color indicates the difference in *Z*-scored high-gamma peaks during the speaking and listening conditions (ΔZ). The right hemisphere is cropped to emphasize insula ROI, while the left hemisphere is shown in its 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 electrode highlighting dual onset responses during speech production and perception. The vertical black line indicates the acoustic onset of a 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, the onset of acoustic activity is visualized relative to the click accompanying the presentation of instructions (pink) and the

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 our 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 (Fig. 3F). In other words, while temporal electrodes did sometimes demonstrate dual onset responses, the amplitudes of these responses were larger for speech perception compared with production. We quantified this restriction of "dual onset" electrodes to the posterior insula by taking the peak amplitude in the first 300 ms of activity prior to sentence onset >1.5 SD above the epoch mean as a measure of the onset response (Fig. 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, our 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 the posterior insula responded with comparable latency to the speech perception onset response electrodes observed in primary (HG and PT) and nonprimary auditory cortex (STG and STS), in some cases responding earlier relative to sentence onset than the auditory cortex electrodes (EMM<sub>A1</sub> peak latency =  $93.7 \pm 16.2$  ms; EMM<sub>Aud. nonprimary</sub> peak latency =  $136.7 \pm 9.4$  ms; EMM<sub>insular</sub> peak latency =  $103.2 \pm 11.7$  ms; A1-Aud. nonprimary p = 0.03,  $t_{(148)} = -2.59$ ; A1-insular p = 0.85,  $t_{(159)} = -0.54$ ; Aud. nonprimary-insular p = 0.03,  $t_{(132)} = 2.61$ ; Fig. 3H). This does not suggest a conventionally proposed serial cascade of information from the primary auditory cortex and is instead indicative of a parallel information flow to the 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 us 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.

# 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 the identification of patterns in the data without access to spatial information or the acoustic content of the stimuli (Ding et al., 2010). This method was used to identify onset and sustained responses in STG (Hamilton et al., 2018). Here, we used cNMF to identify response profiles in our 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 (Fig. 4A). We chose this threshold as increasing the number of clusters in the factorization beyond k = 9 resulted in redundant clusters. Similar response profiles were seen using other numbers of clusters (Materials and Methods). Single-electrode responses to spoken sentences, perceived sentences, and an intertrial 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 a cluster for the k = 9 factorization are shown in Figure 5. 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 (Fig. 4B,C). An additional cluster (c3) was localized to the ventral sensorimotor cortex and showed selectivity to speech production trials, particularly prior to articulation. This cluster is located in the 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, we quantified their functional separation by examining whether individual electrodes contributed strong weighting to both clusters. We observed that despite the spatial proximity of the clusters (which cNMF's clustering technique would not have access to), the majority of electrodes in both onset suppression and dual onset clusters were only strongly weighted within a single cluster (Fig. 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 (Fig. 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 together, the supervised and unsupervised analyses suggest auditory feedback is processed differently by two regions in the temporal and insular cortex. The auditory cortex suppresses responses to self-generated speech

onset of vocalization (red). *F*, Strip plot showing the distribution of channel-by-channel onset response peak amplitudes separated by an anatomical region of interest and whether onset responses occur only during perception (left), only during production (center), or during perception and production (right). Electrodes are colored according to the colormaps of (*A*–*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. *H*, Bar plot showing the estimated marginal mean latency of the onset response in three regions of interest: auditory primary (HG + PT), auditory nonprimary (STG + STS), and posterior and 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/PostCrS, inferior/superior/anterior/posterior circular sulcus of the insula; LGI, long gyrus of the insula; SGI, short gyrus of the insula; PT, planum temporale.



**Figure 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 the total number of clusters in factorization. Threshold of k = 9 factorization plotted as the 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) prearticulatory motor electrodes, characterized by a clusters of stimulus during speech production and localized to ventral sensorimotor cortex. Left, Cluster basis functions for speaking sentences (purple), listening to sentences (green), and intertrial click (pink) for c1, c2, and c3. Center, Right, Two example electrodes for 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 contribute on dual onset and onset suppression cNMF clusters in both hemispheres. The top 50 weighted electrodes for each cluster are plotted on a template brain with an inflated cortical surface; the dark gray indicates sulci while the light gray indicates gyri. The red electrodes contribute more weight to the "onset suppression" cluster while blue electrodes contribute more to the "dual onset" cluster; the purple electrodes contribute equally to both clusters while the 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 nonoverlapping. Abbreviations: STG, s

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.

# 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 (Scheerer and Jones, 2014; Lester-Smith et al., 2020). We were interested if other top-down expectations about the task could affect the responses of electrodes in our data and if these populations overlapped with speaker-induced suppression. To accomplish this, we separated the playback condition into blocks of consistent and inconsistent playback (Fig. 6*A*). In the consistent playback block, participants always played back the sentence they had just produced in the prior speaking trial. In the inconsistent playback block, participants instead 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 5.** Average response of all clusters in reported cNMF analysis. Nine presented clusters explaining 86% of the variance in the data (Fig. 5*A*). "Onset suppression" and "dual onset" clusters presented (Fig. 5*B*) here are labeled as Clusters 2 and 1, respectively, and "prearticulatory motor" cluster presented (Fig. 5*B*) here is labeled as Clusters 3. The responses plotted are the cluster basis functions of individual clusters relative to either sentence onset (production and perception conditions) or the intertrial click tone (click condition).



**Figure 6.** Playback consistency manipulation yields separate, weaker effects than onset suppression. *A*, Task schematic showing playback consistency manipulation. Participants read a sentence aloud (purple) and then passively listened to a 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; the dark gray indicates sulci while the 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, and white indicates a response to neither. Most electrodes are pink, indicating strong responses to both conditions. Example electrodes from *D* are indicated. *C*, Scatter plot of channel-by-channel peak high-gamma activity during consistent playback (*y*-axis) and inconsistent playback (*x*-axis). The vertical black line indicates unity. Color corresponds to a 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 the temporal cortex demonstrating a slight preference for inconsistent playback. Right column (e3 and e4), Electrodes in the frontal/parietal 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 (Fig. 6B, pinkred electrodes; Fig. 6C, electrodes along the unity line). 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 (Fig. 6D). 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 (Fig. 6D, e3 and e4), 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 the superior temporal gyrus and superior temporal sulcus. Curiously, a subset of electrodes localized to the ventral sensorimotor cortex (similar to cluster c3 presented in Fig. 4B) showed an overall preference for speech production trials with prearticulatory activity but, within the playback contrast, demonstrated a preference for consistent playback (Fig. 6D, e5 and e6). We interpret this finding as a speech motor region that indexes predictions of upcoming sensory content for a role in feedback control.

## 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, we observed that onset responses are suppressed during speech production, which motivates the investigation of whether phonological feature tuning is also modulated as part of the auditory system's differential processing of auditory information while speaking. To investigate this, we fit multivariate temporal receptive fields (mTRF) for each electrode to describe the relationship between the neural response at that electrode and selected phonological and task-level features of the stimulus (Fig. 7A). We 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 (Fig. 7B,C).

Onset suppression electrodes in auditory cortex and dual onset electrodes in the posterior insula were both well modeled using this approach ( $r_{onset suppression electrodes = 0.17 \pm 0.08$ ;  $r_{dual onset electrodes = 0.16 \pm 0.11$ ; range, -0.25 to 0.64; Fig. 7D). 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 the auditory cortex (Mesgarani et al., 2014; Berezutskaya et al., 2017; Hamilton et al., 2018, 2021; Oganian and Chang, 2019). Of note,

posterior and inferior insula electrodes were strongly phonologically tuned, with a short temporal response profile as was seen in our prior latency analysis. 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 was explainable by global task-related stimulus features (i.e., whether a sound occurred during a production or a perception trial).

To directly compare phonological feature representations during perception and production, we used variance partitioning techniques to omit or include specific stimulus features in our 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., 2021). In the base model, we included 14 phonological features and 4 task-related features. We first expanded the specificity of phonological feature tuning in our 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 versus produced phonological features separately. However, we saw no significant increase in model performance when expanding the model in this way (Fig. 7F, pink points). 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 (Fig. 7E, far right). The contrast between "base" and "task-specific" model performance, while significant, was a weak effect in favor of the simpler "task-specific" model (EMM<sub>base - task-specific phnfeat</sub>  $\Delta r =$ -0.002, p = 0.04,  $t_{(827)} = -2.01$ ). In contrast, the removal of the playback consistency information from the task-specific portion of the stimulus matrix more substantially affects model performance (EMM<sub>base - omit consistent/inconsistent</sub>  $\Delta r = 0.02$ , p < 0.001,  $t_{(831)} = 10.53$ ). However, the most drastic impairment of model performance emerges when removing information about the contrast of perception and production trials entirely from the (EMM<sub>base - omit perception/production</sub>  $\Delta r = 0.11$ , p < 0.001, model  $t_{(847)} = 27.93$ ). 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 our 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.

## Discussion

We used a sentence reading and playback task that allowed us to compare mechanisms of auditory perception and production while controlling for stimulus acoustics and timing. The primary objective was to determine if the temporal structure and phonological tuning of neural responses change during speaking compared with listening. Using sEEG has the distinct advantage of penetrating deeper structures inside the Sylvian fissure, such as the insula and Heschl's gyrus (Mercier et al., 2022). In the temporal cortex, proximal to where onset responses have been previously identified using surface electrocorticography (Hamilton et al., 2018), we observed a selective suppression of transient responses to sentence onset during speech production, whereas sustained responses remained relatively unchanged between speech perception and production. 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 (Martikainen et al., 2005; Heinks-Maldonado et al., 2006; Hawco et al., 2009; Kurteff et al., 2023). While we do not claim the onset responses observed in our study and others to be equivalent to N/M100, there is a parallel to be drawn between the temporal characteristics of our suppressed cortical activity and the deep literature on suppression of these components during speech production in noninvasive studies. 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 the 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 (Tourville and Guenther, 2011; Houde and Chang, 2015). A notable difference between the original reporting of onset and sustained response profiles in a previous study (Hamilton et al., 2018) and the current study is that many of the electrodes reported in our 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 dorsoposterior STG. Many of the onset electrodes 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 response, 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); our use of sEEG depths may provide greater coverage of these intrasylvian structures. Alternatively, the mixed onset and sustained responses we see in our data may be a mixture of the onset region with the posterior subset of sustained electrodes reported in the original paper. We did observe solely onset-responsive and solely sustained-responsive electrodes (in line with the original paper), but a majority of the onset suppression response profiles described in this study consisted of a mixture of onset and sustained responses at the single-electrode level. Responses to the intertrial click tone observed at some electrodes are another example of pure onset response electrodes in these data.

The suppression of onset responses in the temporal cortex did not impact the structure of phonological feature representations for these electrodes. Phonological feature tuning has been demonstrated previously during speech production, but the analysis focused primarily on the motor cortex and not a direct comparison to the representations present in the temporal cortex during speech perception (Cheung et al., 2016). In the present study, an encoding model capable of differentially encoding phonological features during speech perception and production did not outperform a model only capable of encoding phonological features identically during perception and production, demonstrating that differences in encoding performance during speech production are not due to changes in the phonological feature tuning of individual electrodes. In other words, an electrode that encodes plosive voiced obstruents (like /b/, /g/, and /d/) during speech perception will still encode plosive voiced obstruents during speaker-induced suppression, but the amplitude of the response is reduced during speaking. This is consistent with similar research in scalp EEG conducted by our group (Kurteff et al., 2023) and supports the confinement of cortical suppression during speech production strictly to lower-level sensory components of the auditory system. This is also in line with previous literature showing the degree of suppression observed at an individual utterance is dependent on that utterance's adherence to a sensory goal (Niziolek et al., 2013).

In our analysis, the posterior insula served as a unique functional region in processing auditory feedback during speech production and perception. Unlike the temporal cortex, onset responses were not suppressed during speech production in the posterior insula; the region instead exhibited "dual onset" responses during speech production and perception. Prior work based on lesions and functional imaging suggests that the insula plays a motor role in speech preparation in humans (Dronkers, 1996; Ackermann and Riecker, 2004; Mandelli et al., 2014). However, these studies prescribe this role to the anterior insula, whereas our findings are constrained to the posterior insula, and the insula is far from anatomically or functionally homogenous (Kurth et al., 2010; Zhang et al., 2018; Quabs et al., 2022). A meta-analysis of the functional role of the 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 dorsoposterior granular-dysgranular zone, a ventromedial posterior agranular-dysgranular zone, and a dorsoanterior granular zone (Quabs et al., 2022). Based on the general organizational principles of these articles, the dual onset responses we observed in the posterior insula overlap with functional regions of interest for somatosensory, motor, speech, and interoceptive function and with the dorsoposterior and ventromedial posterior cytoarchitectonic zones. The posterior insula responses we report in this study are purely postarticulatory, indicating a role in auditory feedback monitoring rather than a preparatory motor role. This is corroborated by a recent study that identified an auditory region in the dorsoposterior insula through intraoperative electrocortical stimulation (Zhang et al., 2018), whereby stimulation to the posterior insula resulted in auditory hallucinations. Several studies using animal models, including nonhuman primates, have also identified an auditory field in the posterior insula (Linke and Schwegler, 2000; Rodgers et al., 2008; Remedios et al., 2009). The insular auditory field receives direct input from primary and secondary auditory areas (Sawatari et al., 2011; Takemoto et al., 2014; Jankowski et al., 2023). Additionally, the insular auditory field likely receives parallel direct input from the auditory thalamus given that prior work finds it responds to tonal sounds with similar or shorter latencies than observed in the primary auditory cortex (Sawatari et al., 2011; Takemoto et al., 2014; Jankowski et al., 2023). Our results extend prior findings from animal models, as we observed faster (or equivalently fast) responses to auditory playback stimuli in the posterior insula compared with primary (HG, PT) and higher-order (STG, STS) auditory areas. Thus, this study corroborates parallel auditory pathways between the auditory cortex and posterior insula but in the human brain and with more complex auditory stimuli than pure tones. We also expand upon animal models by showing responses to auditory feedback in the insula are also present during speech production.

While posterior insula and HG are neighboring anatomical structures, we do not believe our 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 corroborated by the functional separation of cluster weights in our cNMF analysis between "onset suppression" and "dual onset" electrodes, alongside the fact that the high-gamma LFP we report on has lower spatial spread than other frequency bands (Muller et al., 2016). Our data are by no means the first to report in vivo recordings of the human insula's responses to speech perception and production: Woolnough et al. (2019) also reported postarticulatory activity in the human insula during speech production and perception. Our insular results are distinct from this study in several ways. First, the authors dichotomize the posterior insula with STG, reporting that the posterior insula is more active for self-generated speech supporting a function that is "opposite of STG." However, our dual onset response electrodes in the posterior insula are equivalently responsive to speech perception and production stimuli, with only a small nonsignificant preference for speech production. Second, Woolnough et al. (2019) found that speech-evoked response magnitudes vary between STG and the posterior insula, with task-evoked activity in STG increasing ~200% in broadband gamma activity from baseline, while the posterior insula shows only ~50% increase in activity from baseline. In our results, temporal and insular evoked activity are similar in magnitude. It is possible that changes in temporal and insular response magnitude between this study and ours are due to differing population

ages: Woolnough et al. (2019) reported data from an older cohort (18-50) than the present study (8-37). Third, the authors used separate tasks with distinct stimuli to compare perception and production, while we generated perceptual stimuli from individual participants' own utterances, allowing us 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. We interpret the posterior insula's role in speech production as a hub for integrating the multiple modalities of sensory feedback (e.g., auditory, tactile, and 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).

Our research motivates further investigation of the role of the posterior insula in feedback control of speech production. Error elicitation paradigms are a potent future direction, as prior feedback perturbation studies have demonstrated that cortical suppression of auditory feedback is diminished during speech errors (Behroozmand and Larson, 2011). Intracranial studies have localized error-dependent suppression to STG (Ozker et al., 2024), but prior work has not included insular coverage, meaning it is unknown whether the auditory onset responses we report in the posterior insula also engage in error-dependent modulatory activity. The lack of speaker-induced suppression during production in the posterior insula leads us to speculate that the posterior insula auditory responses are not functionally engaged in error detection but rather in more general tracking and multisensory processing of self-generated speech. Another potential functional role for our posterior insula auditory responses emerges in emotional saliency processing, supported by results that show high-gamma activity in the insula correlates with subjective aversion to nonspeech stimuli with experimentally manipulated roughness (Arnal et al., 2019). Functional connectivity analyses of the salience network in humans have conventionally implicated the anterior insula rather than the posterior insula, which does not support the idea of such a functional role for the posterior insula (Menon and Uddin, 2010; Uddin, 2015). However, rodent studies of the posterior insula auditory field have identified robust connectivity with the amygdala and have theorized that the insular auditory field may relay auditory information to the amygdala for fear conditioning, warranting future investigation of emotional saliency processing in human posterior insula (Rodgers et al., 2008).

While the primary focus of this study was to describe differences in auditory feedback processing during perception and production, we were motivated to include a consistency manipulation within our 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 (Lester-Smith et al., 2020). The importance of predicting upcoming sensory consequences is visible in the neural data as well: unpredicted auditory stimuli result in suppression of scalp EEG components for selfgenerated speech in pitch perturbation studies (Scheerer and Jones, 2014) as well as the speech of others in a turn-taking sentence production task (Fjaellingsdal et al., 2020). We sought to delineate whether onset responses were an important component of specifically speech perception or involved in a more general predictive processing system. While we did observe that presenting auditory playback in a randomized, inconsistent fashion resulted in a greater response amplitude for some onset suppression electrodes in the auditory cortex, this finding did not hold true for most onset suppression electrodes in our data. Importantly, as perceptual stimuli for both consistent and inconsistent playback conditions were generated from the same set of production trials, all experimental conditions contained the same interstimulus intervals and did not fundamentally differ in their temporal predictability, suggesting that general temporal predictability is not driving responses in the subset of auditory cortex electrodes that demonstrate a contrast between inconsistent and consistent playback. This leads us 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 nonspeech movements are attenuated relative to unpaired tones in mice and in humans (Martikainen et al., 2005; Schneider et al., 2018). With cNMF, we identified a cluster in the ventral sensorimotor cortex that was more active for speech production but, within the consistent/inconsistent playback split, preferred consistent playback. We interpret this response profile as indicative of feedback enhancement for the purposes of speech motor control during speech production. This playback consistency manipulation was also included in a recently published EEG version of this task (Kurteff et al., 2023), but the results of the manipulation were inconclusive. In that EEG study, however, we did see cortical suppression at sentence onset, so perhaps the lack of a result for the consistency manipulation is a mixture of the relatively smaller effect size of the consistency manipulation and the lower signal-to-noise ratio of scalp EEG recordings in comparison with intracranial EEG.

Because our dataset uses sEEG depth electrodes, we were able to record from a wide array of cortical and subcortical areas impractical 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. Visualizations of these single-subject trends are available in Extended Data Figures 2-1 and 2-2. 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 (DC7 PT-MT15  $p_{\text{production}} = 0.01$ ;  $p_{\text{perception}} = 0.9$ ). We identified this area using our unsupervised clustering analysis: cNMF identified a cluster selective to clicks and speech production localized to the occipital lobe (Fig. 5, Cluster 6). We interpret this as a byproduct of our task design, as the text was displayed during speech production trials (the sentence to be read aloud) but not during perception trials. The between-peak duration of the bimodal click response observed in the cNMF cluster is  $\sim$ 1,000 ms, which corresponds with the amount of time a fixation cross was displayed at the beginning of each trial (see Materials and Methods). Based on this information, we conclude these occipital electrodes for DC7 are encoding visual scene changes between fixation cross and text display, but we advise caution in generalizing this to a functional localization as we only observed this trend in a single subject. In a separate single subject (DC5), we 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 intetrial click tones

(DC5 AMF-AI4  $p_{\text{production}} = 0.31$ ;  $p_{\text{perception}} < 0.001$ ). Unlike onset suppression electrodes in the 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 toward pars triangularis, but the final electrode of the depth still had a (barely) nonsignificant response to speech production trials (DC5 AMF-AI8  $p_{\text{production}} =$ 0.06;  $p_{\text{perception}} = 0.45$ ). Unlike the occipital electrodes described above, the inferior frontal perception-selective electrodes of DC5 did not emerge as a functional region in our unsupervised clustering analysis and were interspersed with other perception-selective electrodes from other subjects localized to PT and HG (Fig. 5, Cluster 7). While the convention of the inferior frontal cortex being monolithically a speech production region is increasingly being challenged in contemporary research (Flinker et al., 2015; Tremblay and Dick, 2016; Fedorenko and Blank, 2020; Hickok et al., 2023), the confinement of our perception-selective electrodes in this region to a single subject gives us hesitation to weigh in on this topic. Lastly, we acknowledge that the age distribution of our participant population is rather large, as we recorded from children, adolescents, and adults in this study. The development of the auditory and motor systems that support speech motor control across the lifespan may contribute to the presence of these single-subject trends and the general cross-subject variability of our dataset. We intend to report on the development of these processes in further detail in future work, after an appropriate sample for such analyses has been collected.

Overall, this project 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, such as apraxia of speech (Jacks and Haley, 2015), schizophrenia (Heinks-Maldonado et al., 2007), and stuttering (Max and 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 studies, we 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 our data due to the limitations of naturalistic study design, yet we speculate mechanistically different than onset responses. Our 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.

## Data and code availability

- The neural data reported in this study cannot be deposited in a public repository because they could compromise research participant privacy and consent. To request access, contact the lead contact.
- All original code has been deposited at GitHub and is publicly available as of the date of publication: https://github.com/ HamiltonLabUT/kurteff2024\_code
- Further information and requests for resources and reagents required to reanalyze the data reported in this paper should be directed to and will be fulfilled by the lead contact, Liberty S. Hamilton (liberty.hamilton@austin.utexas.edu).

## References

- 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 AM, Johannesma PI (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. Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96, 3, 1541–1544 vol.3.
- Arnal LH, Kleinschmidt A, Spinelli L, Giraud A-L, Mégevand P (2019) The rough sound of salience enhances aversion through neural synchronisation. Nat Commun 10:3671.
- Astheimer LB, Sanders LD (2011) Predictability affects early perceptual processing of word onsets in continuous speech. Neuropsychologia 49:3512– 3516.
- Bamiou DE, Musiek FE, Luxon LM (2003) The insula (Island of Reil) and its role in auditory processing: literature review. Brain Res Rev 42:143–154.
- Behroozmand R, Larson CR (2011) Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. BMC Neurosci 12:54.
- 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 ZV, Güçlü U, van Gerven MAJ, Ramsey NF (2017) Neural tuning to low-level features of speech throughout the perisylvian cortex. J Neurosci 37:7906–7920.
- Boersma P, Weenink D (2013) Praat: doing phonetics by computer [Computer program]. Version 5.3. 51. Online: Http://Www. Praat. Org/ Retrieved, Last Viewed On, 12.
- Bouchard KE, Chang EF (2014) Control of spoken vowel acoustics and the influence of phonetic context in human speech sensorimotor cortex. J Neurosci 34:12662–12677.
- Bouchard KE, Mesgarani N, Johnson K, Chang EF (2013) Functional organization of human sensorimotor cortex for speech articulation. Nature 495: 327–332.
- Breshears JD, Molinaro AM, Chang EF (2015) A probabilistic map of the human ventral sensorimotor cortex using electrical stimulation. J Neurosurg 123:340–349.
- 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.
- Chang EF (2015) Towards large-scale, human-based, mesoscopic neurotechnologies. Neuron 86:68–78.
- Chang EF, Niziolek CA, Knight RT, Nagarajan SS, Houde JF (2013) Human cortical sensorimotor network underlying feedback control of vocal pitch. Proc Natl Acad Sci U S A 110:2653–2658.
- Chao ZC, 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 GK, Johnson K, Chang EF (2018) Encoding of articulatory kinematic trajectories in human speech sensorimotor cortex. Neuron 98:1042–1054.e4.
- Cheung C, Hamilton LS, Johnson K, Chang EF (2016) The auditory representation of speech sounds in human motor cortex. Elife 5:1–19.
- Cogan GB, Thesen T, Carlson C, Doyle W, Devinsky O, Pesaran B (2014) Sensory-motor transformations for speech occur bilaterally. Nature 507: 94–98.
- 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 MJ, Di Liberto GM, Bednar A, Lalor EC (2016) The multivariate temporal response function (mTRF) toolbox: a MATLAB toolbox for relating neural signals to continuous stimuli. Front Hum Neurosci 10:604.
- Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9: 179–194.
- 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.
- Dichter BK, Breshears JD, Leonard MK, Chang EF (2018) The control of vocal pitch in human laryngeal motor cortex. Cell 174:21–31.e9.
- Di Liberto GM, O'Sullivan JA, Lalor EC (2015) Low-frequency cortical entrainment to speech reflects phoneme-level processing. Curr Biol 25: 2457–2465.
- Ding C, Li T, Jordan MI (2010) Convex and semi-nonnegative matrix factorizations. IEEE Trans Pattern Anal Mach Intell 32:45–55.
- Dronkers NF (1996) A new brain region for coordinating speech articulation. Nature 384:159–161.
- Evans AC, Collins L, Mills SR, Peters TM (1993). 3D statistical neuroanatomical models from 305 MRI volumes. Nuclear Science Symposium and Medical Imaging Conference, 1993., 1993 IEEE Conference Record., 1813–1817, 1813–1817 vol.3.
- Fedorenko E, Blank IA (2020) Broca's area is not a natural kind. Trends Cogn Sci 24:270–284.
- Fjaellingsdal G, Schwenke T, Scherbaum D, Kuhlen S, Bögels AK, Meekes S, & Bleichner J, G M (2020) Expectancy effects in the EEG during joint and spontaneous word-by-word sentence production in German. Sci Rep 10:5460.
- Flinker A, Chang EF, Kirsch HE, Barbaro NM, Crone NE, Knight RT (2010) Single-trial speech suppression of auditory cortex activity in humans. J Neurosci 30:16643–16650.
- Flinker A, Korzeniewska A, Shestyuk AY, Franaszczuk PJ, Dronkers NF, Knight RT, Crone NE (2015) Redefining the role of Broca's area in speech. Proc Natl Acad Sci USA 112:2871–2875.
- Forseth KJ, Hickok G, Rollo PS, Tandon N (2020) Language prediction mechanisms in human auditory cortex. Nat COoMmmun 11:5240.
- Friedman D, Goldman R, Stern Y, Brown TR (2009) The brain's orienting response: an event-related functional magnetic resonance imaging investigation. Hum Brain Mapp 30:1144–1154.
- Gonzalez Castro LN, Hadjiosif AM, Hemphill MA, & Smith MA (2014) Environmental consistency determines the rate of motor adaptation. Curr Biol 24:1050–1061.
- Gramfort A, et al. (2013) MEG and EEG data analysis with MNE-python. Front Neurosci 7:267.
- 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 FH (2016) Neural control of speech. Cambridge, MA: MIT Press.
- Habib M, Daquin G, Milandre L, Royere ML, Rey M, Lanteri A, Salamon G, Khalil R (1995) Mutism and auditory agnosia due to bilateral insular damage–role of the insula in human communication. Neuropsychologia 33: 327–339.
- Hamilton LS, Chang DL, Lee MB, Chang EF (2017) Semi-automated anatomical labeling and inter-subject warping of high-density intracranial recording electrodes in electrocorticography. Front Neuroinform 11:62.
- Hamilton LS, Edwards E, Chang EF (2018) A spatial map of onset and sustained responses to speech in the human superior temporal gyrus. Curr Biol 28:1860–1871.e4.

- Hamilton LS, Oganian Y, Hall J, Chang EF (2021) Parallel and distributed encoding of speech across human auditory cortex. Cell 184:4626–4639.e13.
- Hawco CS, Jones JA, Ferretti TR, Keough D (2009) ERP correlates of online monitoring of auditory feedback during vocalization. Psychophysiology 46:1216–1225.
- Hayes B (2011) Introductory phonology. Hoboken, NJ: John Wiley & Sons.
- Heinks-Maldonado TH, Mathalon DH, Houde JF, Gray M, Faustman WO, Ford JM (2007) Relationship of imprecise corollary discharge in schizophrenia to auditory hallucinations. Arch Gen Psychiatry 64:286–296.
- Heinks-Maldonado TH, Nagarajan SS, Houde JF (2006) Magnetoencephalographic evidence for a precise forward model in speech production. Neuroreport 17:1375–1379.
- 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 AE, Work M, Barker PB, Jacobs MA, Breese EL, Maurer K (2004) Re-examining the brain regions crucial for orchestrating speech articulation. Brain 127:1479–1487.
- Houde JF, Chang EF (2015) The cortical computations underlying feedback control in vocal production. Curr Opin Neurobiol 33:174–181.
- Houde JF, Nagarajan SS (2011) Speech production as state feedback control. Front Hum Neurosci 5:82.
- Houde JF, Nagarajan SS, Sekihara K, Merzenich MM (2002) Modulation of the auditory cortex during speech: an MEG study. J Cogn Neurosci 14: 1125–1138.
- Indefrey P, Levelt WJM (2004) The spatial and temporal signatures of word production components. Cognition 92:101–144.
- Ivanova AA, Hewitt J, Zaslavsky N (2021). Probing artificial neural networks: insights from neuroscience. In arXiv [cs.LG]. arXiv. http://arxiv.org/abs/ 2104.08197
- Jacks A, Haley KL (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.
- Jankowski MM, Karayanni M, Harpaz M, Polterovich A, Nelken I (2023). A rapid anterior auditory processing stream through the insulo-parietal auditory field in the rat. In *bioRxiv* (p. 2023.09.12.557409).
- Kawamoto AH, Liu Q, Lee RJ, Grebe PR (2014) The segment as the minimal planning unit in speech production: evidence based on absolute response latencies. Q J Exp Psychol (Hove) 67:2340–2359.
- Kenward MG, Roger JH (1997) Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 53:983–997.
- 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 GL, et al. (2023) Speaker-induced suppression in EEG during a naturalistic Reading and listening task. J Cogn Neurosci 35:1538–1556.
- Kurth F, Zilles K, Fox PT, Laird AR, Eickhoff SB (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. J Stat Softw 82:1–26.
- Lachaux JP, Axmacher N, Mormann F, Halgren E, Crone NE (2012) High-frequency neural activity and human cognition: past, present and possible future of intracranial EEG research. Prog Neurobiol 98:279–301.
- 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 MK, Cai R, Babiak MC, Ren A, Chang EF (2019) The peri-sylvian cortical network underlying single word repetition revealed by electrocortical stimulation and direct neural recordings. Brain Lang 193:58–72.
- Leonard MK, Gwilliams L, Sellers KK, Chung JE, Xu D, Mischler G, Mesgarani N, Welkenhuysen M, Dutta B, Chang EF (2023) Large-scale single-neuron speech sound encoding across the depth of human cortex. Nature 626:593–602.
- Lester-Smith RA, Daliri A, Enos N, Abur D, Lupiani AA, Letcher S, Stepp CE (2020) The relation of articulatory and vocal auditory-motor control in typical speakers. J Speech Lang Hear Res 63:3628–3642.
- Levelt WJM (1993) *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.
- Linke R, Schwegler H (2000) Convergent and complementary projections of the caudal paralaminar thalamic nuclei to rat temporal and insular cortex. Cerebral Cortex 10:753–771.

- Luck SJ (2014) An Introduction to the event-related potential technique, Ed. 2. Cambridge, MA: MIT Press.
- Mandelli ML, et al. (2014) Frontal white matter tracts sustaining speech production in primary progressive aphasia. J Neurosci 34:9754–9767.
- Martikainen MH, Kaneko K-I, Hari R (2005) Suppressed responses to self-triggered sounds in the human auditory cortex. Cereb Cortex 15: 299–302.
- 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.
- Menon V, Uddin LQ (2010) Saliency, switching, attention and control: a network model of insula function. Brain Struct Funct 214:655–667.
- Mercier MR, et al. (2022) Advances in human intracranial electroencephalography research, guidelines and good practices. Neuroimage 260:119438.
- Mesgarani N, Cheung C, Johnson K, Chang EF (2014) Phonetic feature encoding in human superior temporal gyrus. Science 343:1006–1010.
- Muller L, Hamilton LS, Edwards E, Bouchard KE, Chang EF (2016) Spatial resolution dependence on spectral frequency in human speech cortex electrocorticography. J Neural Eng 13:056013.
- Nguyen D, Isnard J, Kahane P (2022) *Insular epilepsies*. Cambridge, UK: Cambridge University Press.
- Niziolek CA, Nagarajan SS, Houde JF (2013) What does motor efference copy represent? Evidence from speech production. J Neurosci 33:16110–16116.
- Nourski KV, Steinschneider M, Rhone AE, Kovach CK, Banks MI, Krause BM, Kawasaki H, Howard MA (2021) Electrophysiology of the human superior temporal sulcus during speech processing. Cereb Cortex 31: 1131–1148.
- Oganian Y, Bhaya-Grossman I, Johnson K, Chang EF (2023) Vowel and formant representation in the human auditory speech cortex. Neuron 111: 2105–2118.e4.
- Oganian Y, Chang EF (2019) A speech envelope landmark for syllable encoding in human superior temporal gyrus. Sci Adv 5:eaay6279.
- 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* : *The Preprint Server for Biology*.
- Penfield W, Roberts L (1959) Speech and brain mechanisms. Princeton, NJ: Princeton University Press.
- 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.
- Ray S, Maunsell JHR (2011) Different origins of gamma rhythm and highgamma activity in macaque visual cortex. PLoS Biol 19:e1000610.
- Remedios R, Logothetis NK, Kayser C (2009) An auditory region in the primate insular cortex responding preferentially to vocal communication sounds. J Neurosci 29:1034–1045.
- Rodgers KM, Benison AM, Klein A, Barth DS (2008) Auditory, somatosensory, and multisensory insular cortex in the rat. Cereb Cortex 18:2941– 2951.
- 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 NE, Jones JA (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.
- Schneider DM, Nelson A, Mooney R (2014) A synaptic and circuit basis for corollary discharge in the auditory cortex. Nature 513:189–194.
- Schneider DM, Sundararajan J, Mooney R (2018) A cortical filter that learns to suppress the acoustic consequences of movement. Nature 561:391–395.
- Shadmehr R, Krakauer JW (2008) A computational neuroanatomy for motor control. Exp Brain Res 185:359–381.
- St. Louis KO, Ruscello DM (1981) Oral speech mechanism screening examination (OSMSE). Baltimore, MD: University Park Press.
- 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 LS, Chang EF (2017) Intonational speech prosody encoding in the human auditory cortex. Science 357:797–801.
- Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J Neurosci 20: 2315–2331.

- Tourville JA, Guenther FH (2011) The DIVA model: a neural theory of speech acquisition and production. Lang Cogn Process 26:952–981.
- Towle VL, Yoon H-A, Castelle M, Edgar JC, Biassou NM, Frim DM, Spire J-P, Kohrman MH (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 PF (2020) Speech-induced suppression for delayed auditory feedback in adults who do and do not stutter. Front Hum Neurosci 14:150.
- Tremblay P, Dick AS (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 Trans. Inf. Theory 6: 311–329.

- Uddin LQ (2015) Salience processing and insular cortical function and dysfunction. Nat Rev Neurosci 16:55–61.
- Woolnough O, Forseth KJ, Rollo PS, Tandon N (2019) Uncovering the functional anatomy of the human insula during speech. Elife 8:1–14.
- Wrench A (1999) *The MOCHA-TIMIT articulatory database.* Wright S (1921) Correlation and causation. J Agric Res 20:557. Available at:
- https://books.google.com/books?id=lNNdIV\_qpwIC&pg=PA557&hl= pt-PT&source=gbs\_toc\_r&cad=2#v=onepage&q&f=false.
- 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.