Encoding of phonological features in intracranial recordings of human speech cortex during a dual perception-production task

Garret Kurteff & Liberty Hamilton

The University of Texas at Austin
Moody College of Communication

Introduction and Methods

Background

Speaker-induced suppression (SIS): responses to self-generated speech are suppressed relative to externally generated speech.

- Something encoded during perception is not encoded during production?
    - Phonological features (Mesgarani et al. 2014)
    - Onset and sustained responses (Hamilton et al. 2018)
- Speech production involves feedforward representations while speech perception does not
    - Stimulus predictability related to speaker-induced suppression?
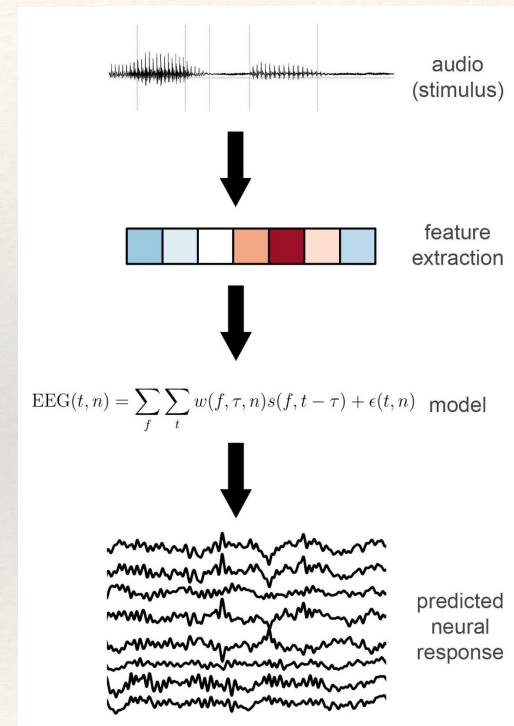

Ozker et al. 2020

The University of Texas at Austin
Moody College of Communication
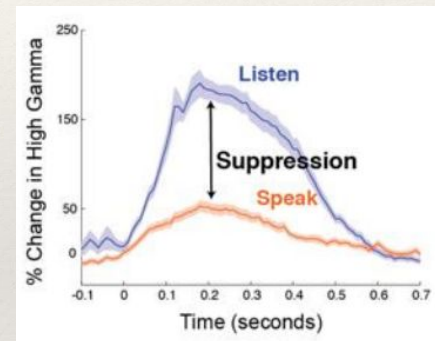
Temporal receptive field modeling

Multivariate temporal receptive fields (mTRF) provide a measure of how much of a neural signal can be explained by stimulus characteristics (DiLiberto et al. 2015)

- Evaluation of model: correlate predicted response with actual response
- Evaluation of stimulus features: Individual features assigned "weight" in mTRF depending on how much they contribute to predicting neural activity

audio (stimulus)

feature extraction

$$\text{EEG}(t, n) = \sum_f \sum_t w(f, \tau, n)s(f, t - \tau) + \epsilon(t, n)$$ model

predicted neural response

The University of Texas at Austin
Moody College of Communication

Research questions

1. Is SIS the loss of phonological feature encoding?

2. Is SIS the loss of encoding onset/sustained responses?

3. Is SIS a feature of predictable speech instead of produced speech?
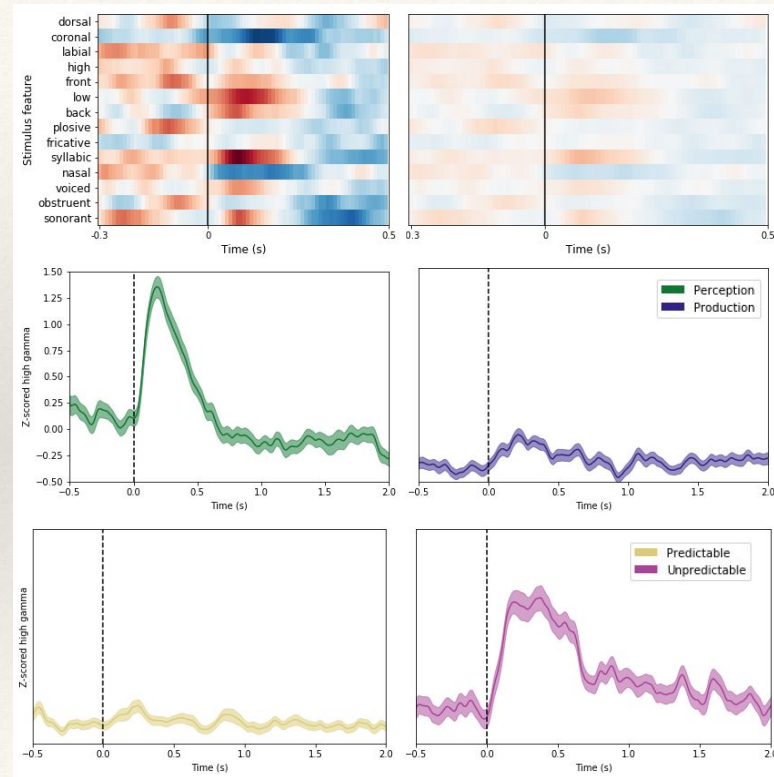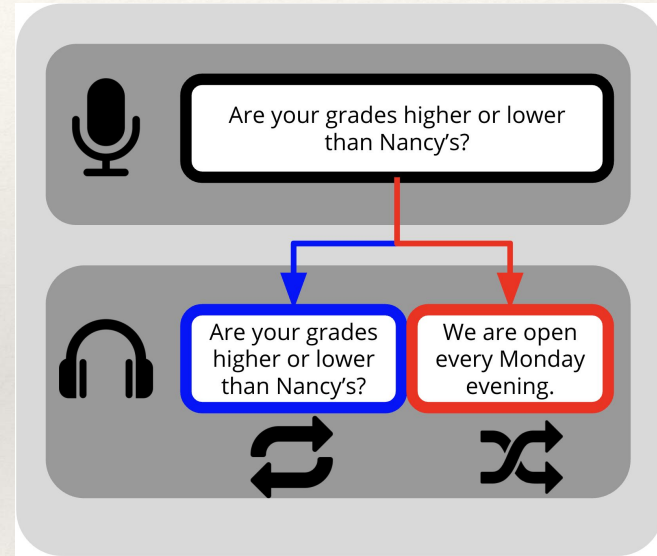
Methods

n = 3 patients undergoing seizure monitoring via sEEG (2M, 1F, 16.3±2)

Dual production-perception task. Each trial consists of:

1. Overt production of MOCHA-TIMIT sentences
2. Listening to playback of their production, split into two conditions:
   - Predictably: immediate playback
   - Unpredictably: random playback of previous trial
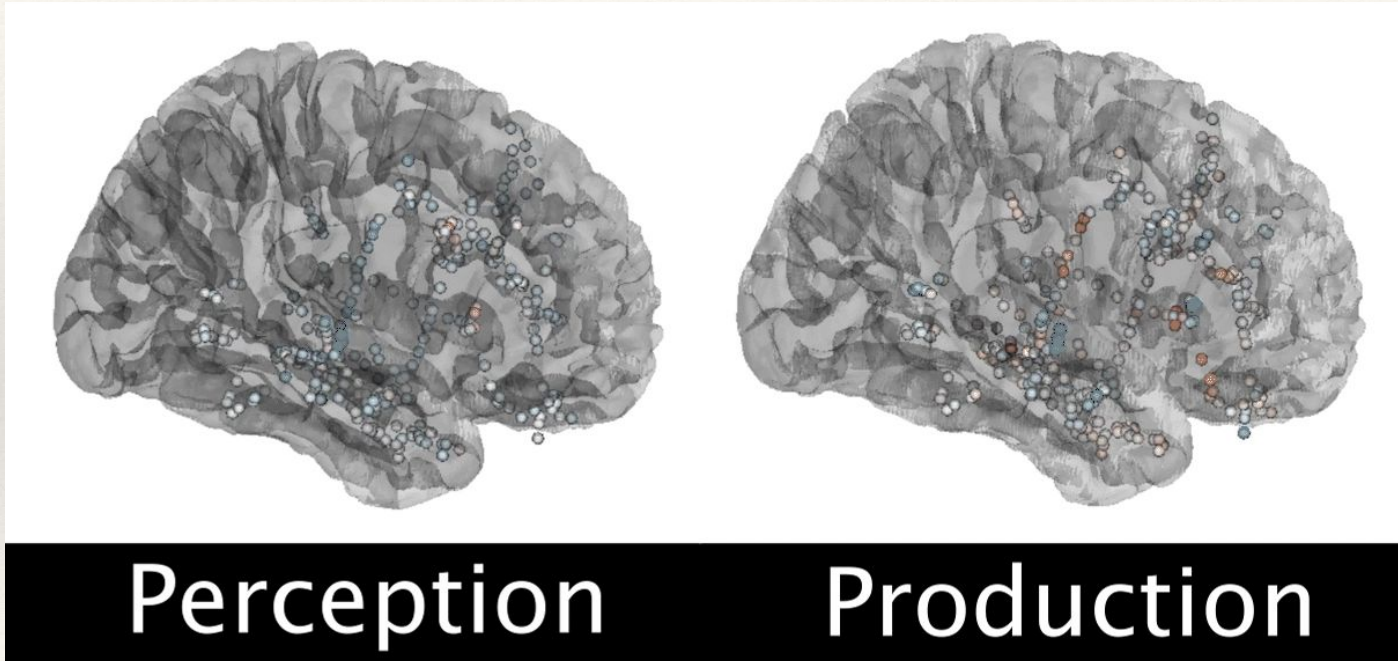

Kurteff et al. (in prep)

The University of Texas at Austin
Moody College of Communication
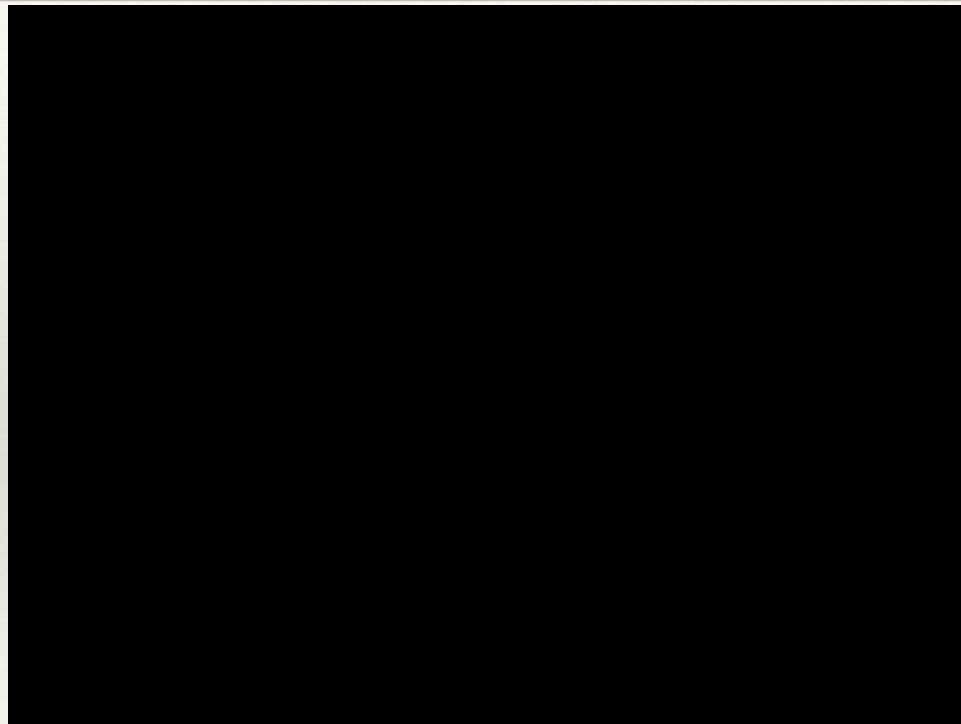
Task video

Results

High-gamma responses to perception and production in all electrodes across subjects

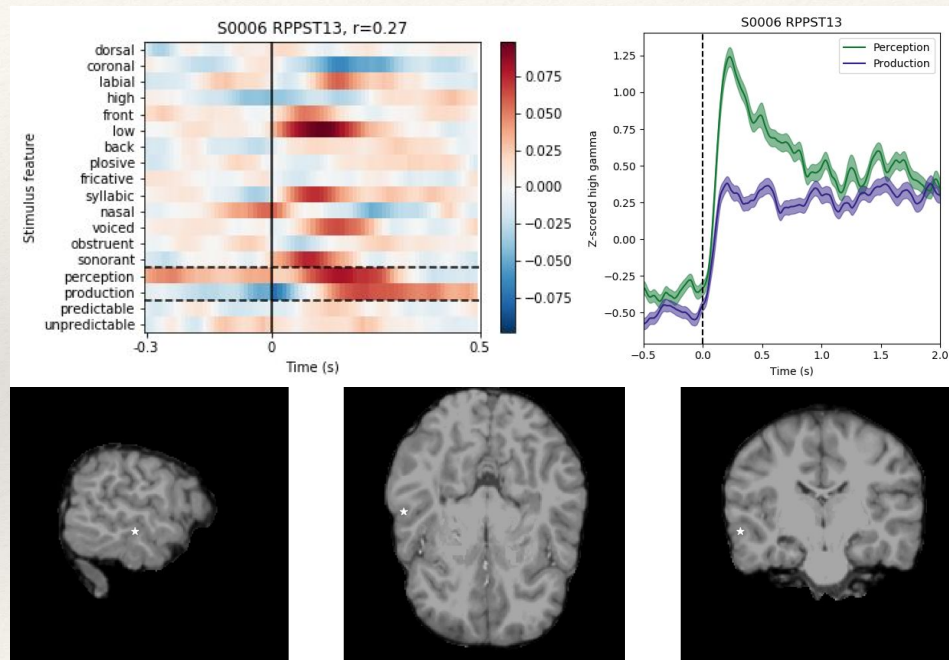Perception   Production

Feature space comparison

- Omitting phonological features affects model performance
- Omitting task information affects performance *much more*
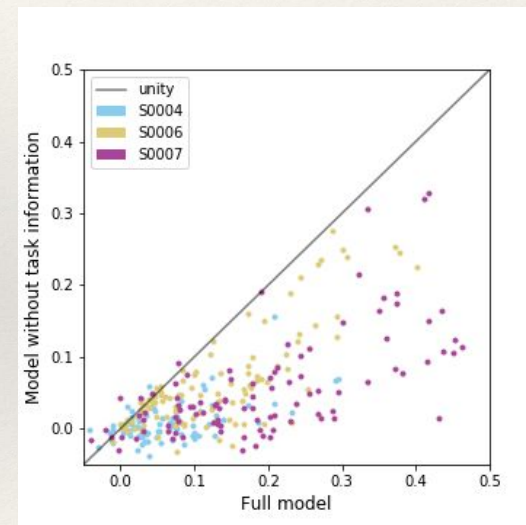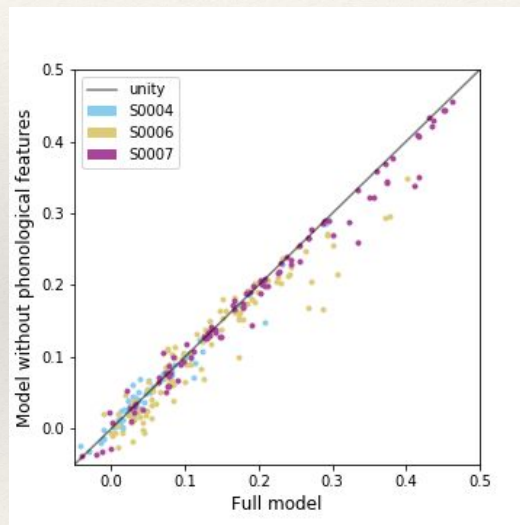
Onset responses are selectively suppressed during speech production in temporal lobe

- Onset responses present during speech perception, but not production
- Phonological feature encoding did not differ between perception and production
- Predictability not differentially encoded
- Percent of electrodes that selectively suppress onsets during speech production:
  - STG: 100%
 - STS: 86%
  - MTG: 100%

Onset responses are *also* selectively suppressed during speech production in frontal lobe

- Onset responses present during speech perception, but not production
- Phonological feature encoding did not differ between perception and production
- Predictability not differentially encoded
- Percent of electrodes showing onset suppression during speech production:
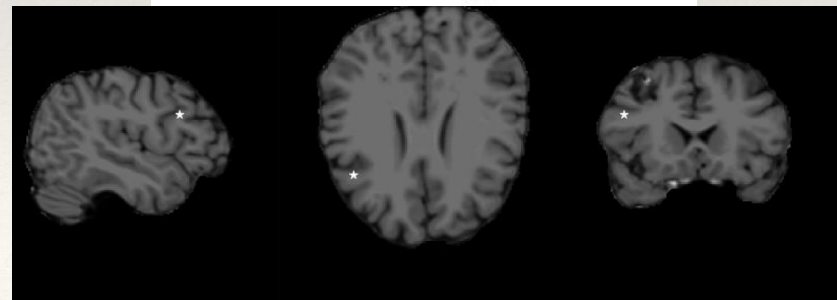 - IFS: 100%
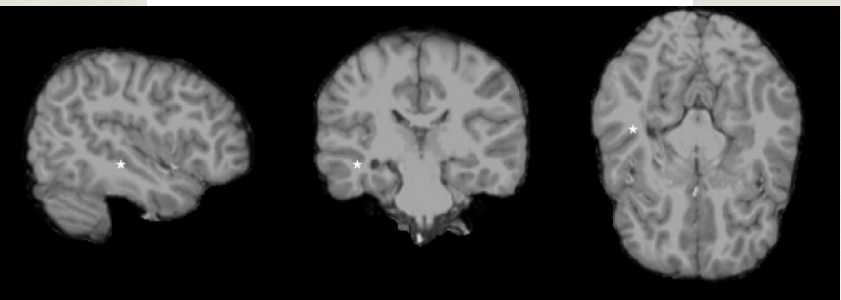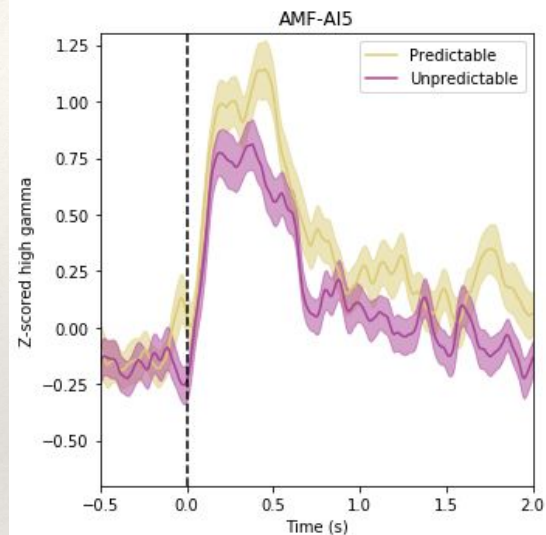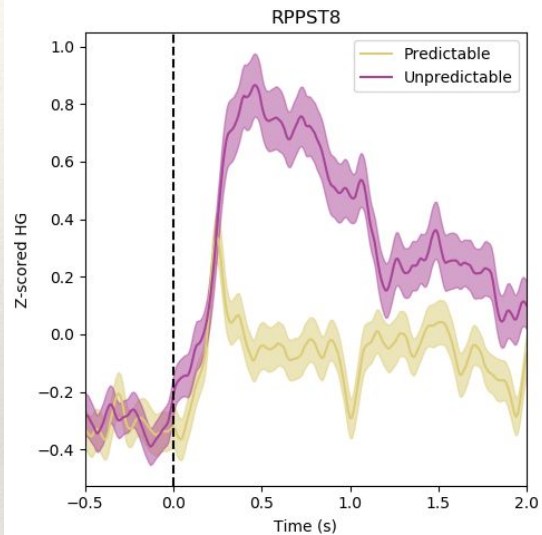 - Pars triangularis: 100%
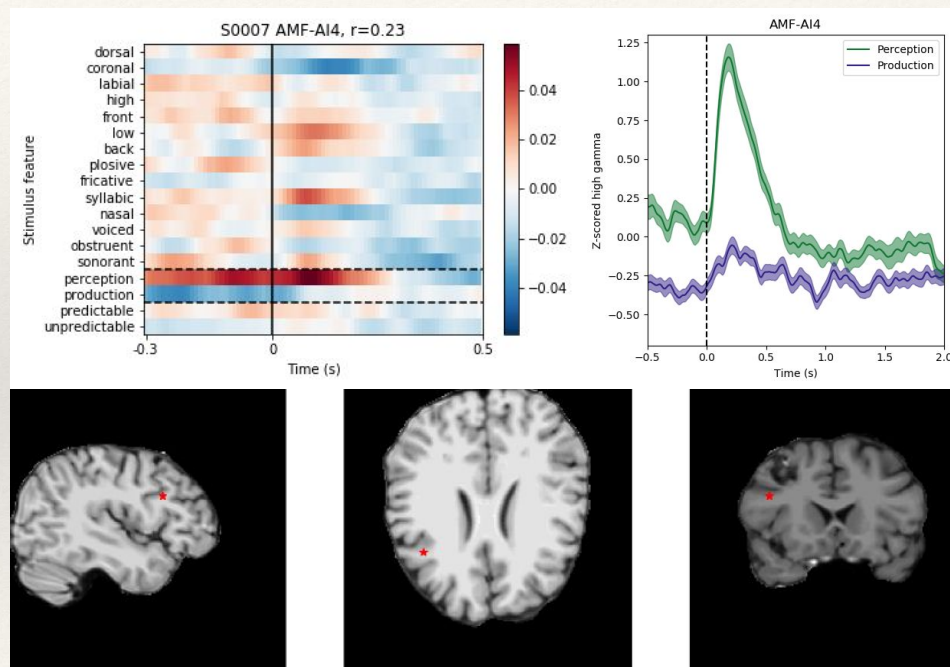
The insula encodes speech production on a faster timescale than speech perception

- Onset responses are present during speech production *and* speech perception in the insula
- Phonological feature encoding did not differ between perception and production
- Predictability not differentially encoded
- Percent of insula electrodes encoding speech production:
  - Superior insula: 93%
  - Inferior insula: 60%
  - Long gyrus: 83%
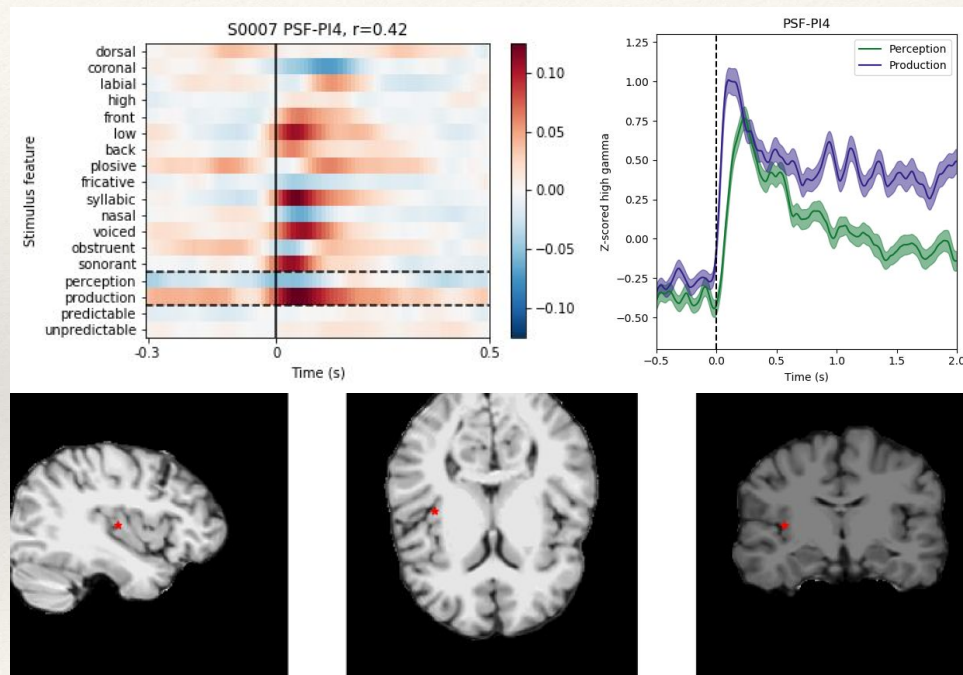
Predictability encoded in temporal & frontal regions, but less in frontal

Discussion

Summary

- Speaker-induced suppression is not the result of differential encoding of phonological features
  - Onset responses *are* differentially encoded
- The insula encodes phonological features during speech production *and* speech perception
  - Responses to speech production occur on a faster timescale than perception
  - Onset responses during speech production are not suppressed in the insula

The University of Texas at Austin
Moody College of Communication

Acknowledgements

- Our wonderful participants
- The Hamilton Lab
- Transcription volunteers: Jade, Nicole, Amanda, Chris, Val, Cassie, Paranjaya
- Dell Children's Medical Center
- Moody College of Communication

Supplemental

Feature selection split by model

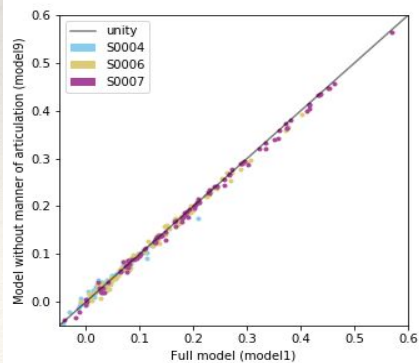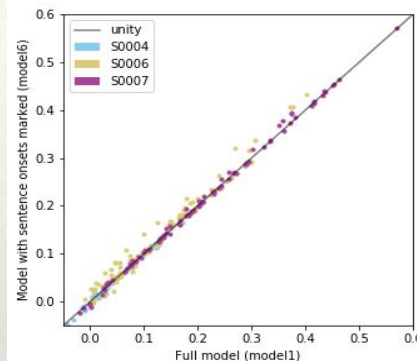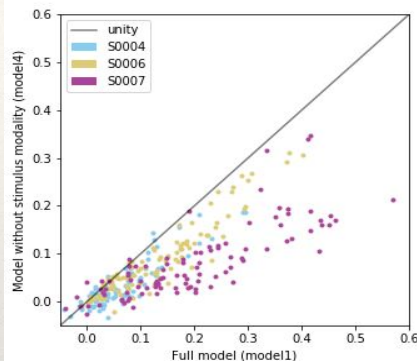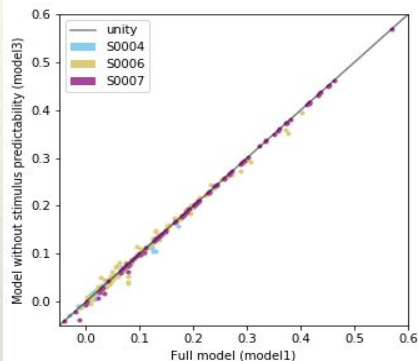| Model number	Model name	Place of articulation	Manner of articulation	Split phonological features between perception and production	Information about modality: perception vs. production	Information about predictability: predictable vs. unpredictable	Onset responses
1	"Full"	green	green	red	green	green	red
2	"Modality features"	green	green	green	green	green	red
3	"Modality only"	green	green	red	green	red	red
4	"Predictability only"	green	green	red	red	green	red
5	"Onset full"	green	green	red	green	green	green
6	"Onset modal"	green	green	green	green	green	green

mTRF equation

$$\text{EEG}(t, n) = \sum_f \sum_t w(f, \tau, n) s(f, t - \tau) + \epsilon(t, n)$$

- The neural response `(EEG)` at a certain time t and electrode n is a convolution between two matrices:
- `w(f,τ,n)` is the temporal receptive field
- `s(f, t - τ)` is the input stimulus
- `ε(t, n)` is the residual response not explained by the model

Di Liberto et al. 2015

The University of Texas at Austin
Moody College of Communication

Feature space comparison scatterplots
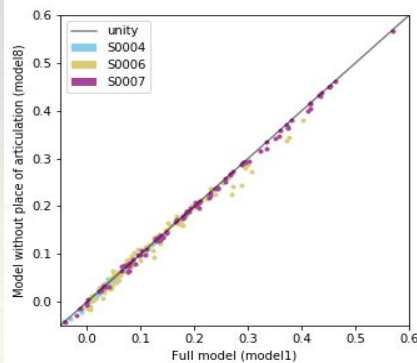
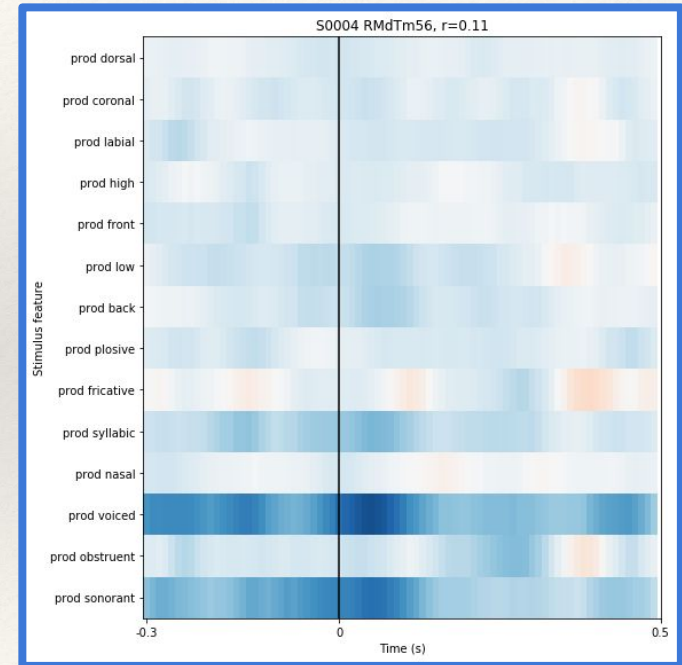Omitting stimulus modality impairs correlation between predicted response and actual response

The University of Texas at Austin
Moody College of Communication

Similar phonological feature encoding across modalities in perception-selective STG
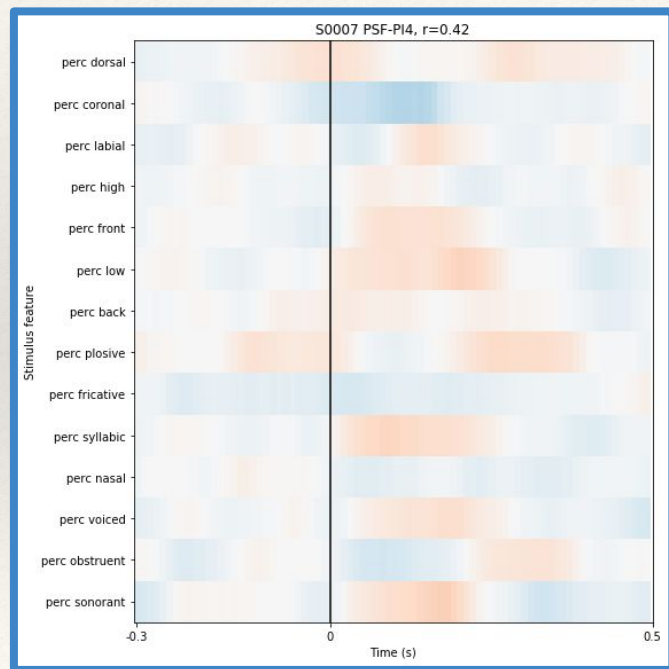
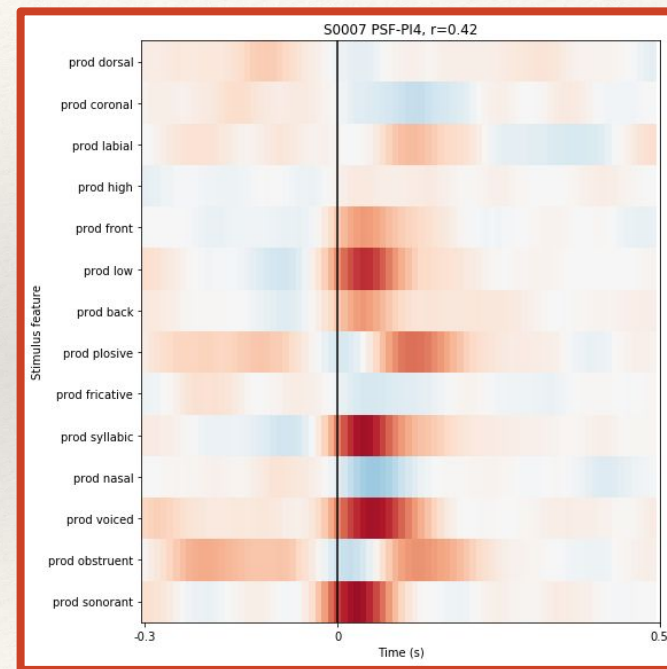

PERCEPTION
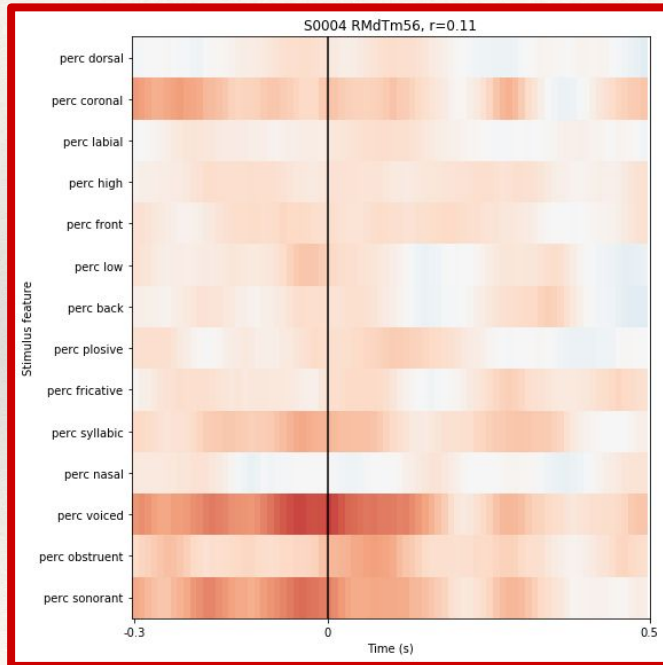
PRODUCTION

Similar phonological feature encoding across modalities in production-selective insula
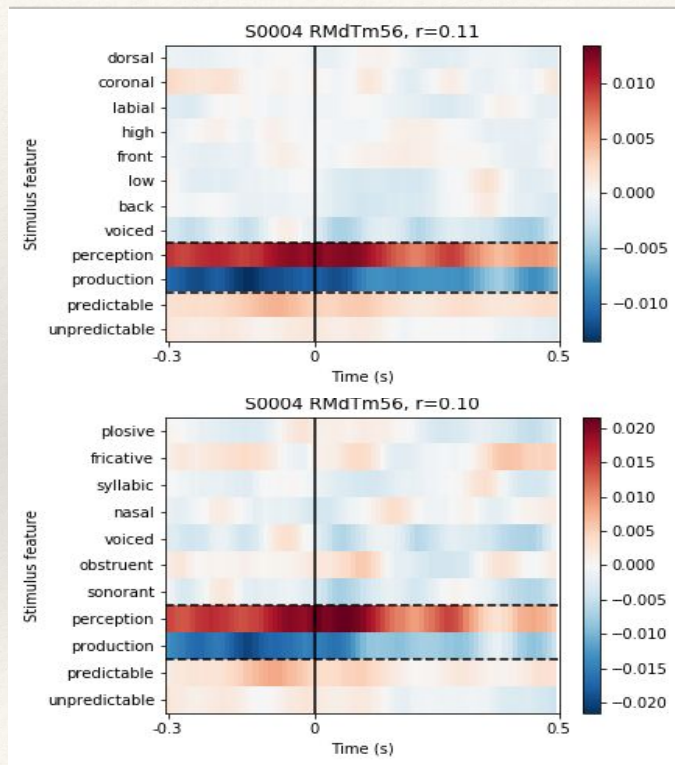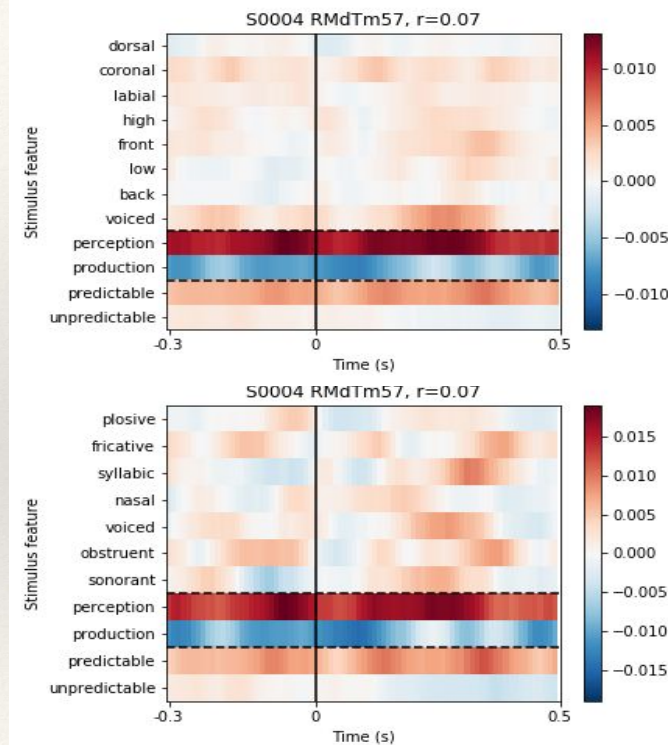

PERCEPTION

PRODUCTION

Encoding of place vs. manner in STG, STS

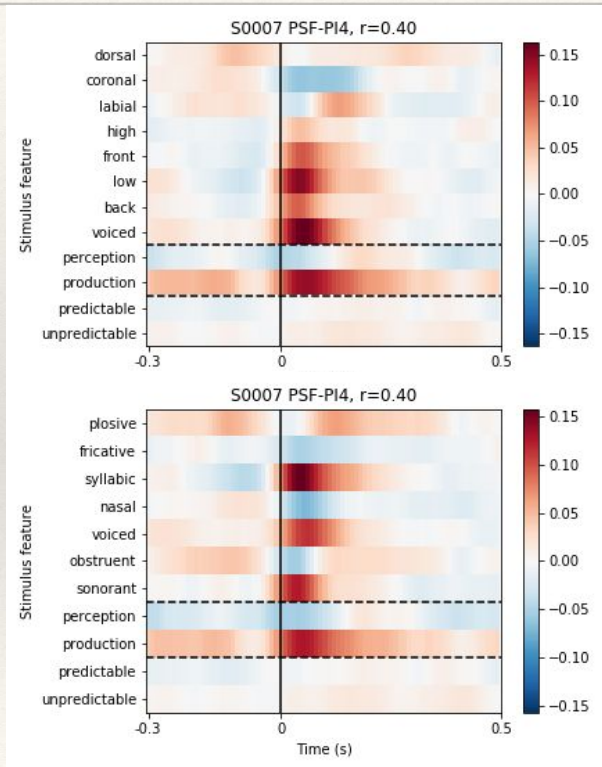Place of
articulation

Manner of
articulation

Encoding of place vs. manner in the insula

Place of articulation

Manner of articulation