Modeling hemodynamic responses to the speech envelope in superior temporal gyrus

Arielle Moore[1], Tyler Perrachione[2], & Emily P. Stephen[3]

[1]Graduate Program for Neuroscience, Boston University; [2]Department of Speech, Language, & Hearing Sciences, Boston University;
[3]Department of Mathematics & Statistics, Boston University

GPN

BOSTON UNIVERSITY

Summary

The amplitude envelope of speech provides important information about its underlying phonetic content, and behavioral evidence suggests these amplitude variations are critical for speech recognition. Measuring neural tracking of the speech envelope is an increasingly popular method for investigating the brain bases of speech recognition, with cortical activity in Heschl's gyrus (HG) and superior temporal gyrus (STG) showing correlations with changes in the speech envelope.[1-3] fMRI is one of the most accessible technologies for obtaining in vivo high spatial resolution neural data over large parts of the brain. Here, we used fMRI listening data to compare traditional and novel modeling approaches for speech envelope-related changes in STG. We found that an integrative reduced rank regression (iRRR) with a finite impulse response (FIR) model demonstrated better predictive performance than HRF and FIR mass univariate models in most subjects. These findings suggest that iRRR may provide a more general framework for capturing speech envelope responses and reveal shared patterns of speech envelope tracking that are not detectable with traditional univariate approaches.

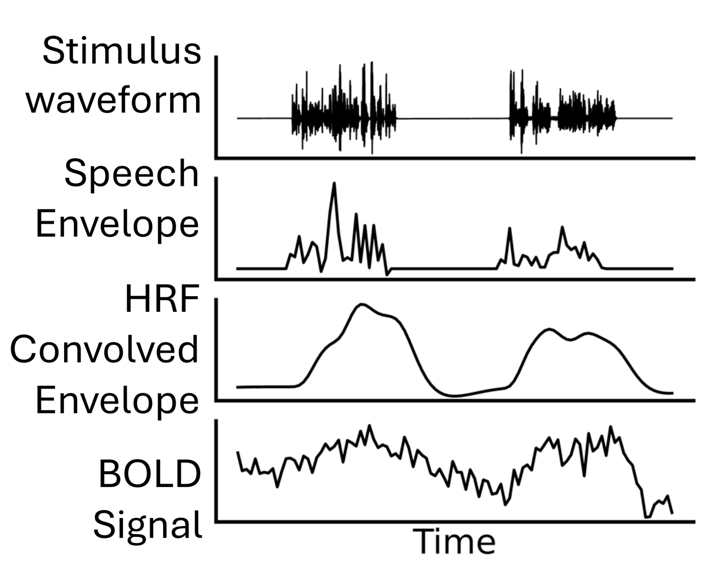
Data

Participants: 24 control native English speakers with normal speech, language, hearing, and reading abilities.

Task Design: Participants listened passively to audio recordings of natural speech (e.g., podcasts, interviews, etc.) in intact and degraded speech conditions in 18-second blocks[7].

fMRI Data Acquisition: Whole-brain structural and functional data were acquired using a 3T Siemens Trio Scanner Block design; TR=750ms, 45 axial slices, f-slice SMS, 3mm^3 voxels, 484 TRs.

Analysis



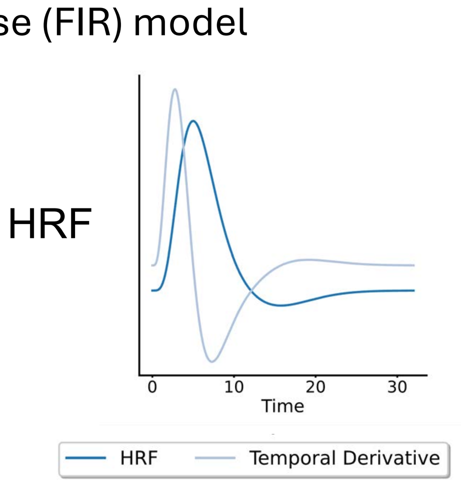
Envelope responses are represented as regressors via two ways: HRF convolution or through a finite impulse response (FIR) model

Mass Univariate Regression

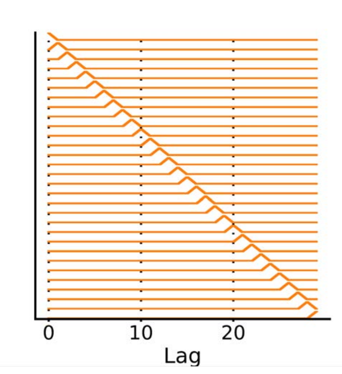
$$Y_{voxel} = X\beta_{voxel} + \epsilon_{voxel}$$

$$\hat{\beta}_{voxel} = (X^T X)^{-1} X^T y_{voxel}$$

HRF



FIR



iRRR: Low-Rank Multivariate Regression

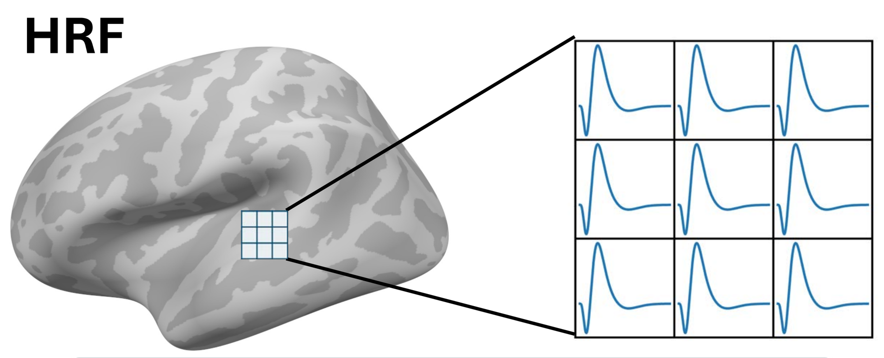
$$\mathbf{Y} = \underbrace{\mathbf{X}_{\mathbf{FIR}}\boldsymbol{\beta}_{\mathbf{FIR}} + \mathbf{X}_{other}\boldsymbol{\beta}_{other}}_{\mathbf{XB}} + \epsilon$$

$$\underset{\text{(Lag x Voxels)}}{\hat{\boldsymbol{\beta}}_{\mathbf{FIR}}} = \underset{\text{(Lag x Components)}}{\mathbf{U}}\ \underset{\text{(Components)}}{\mathbf{D}}\ \underset{\text{(Components x Voxels)}}{\mathbf{V}^T}$$

$$\underset{\beta_{FIR}\beta_{other}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{XB}\|_F^2 + \lambda \sum_i w_i \|\boldsymbol{\beta}_i\|_*$$

$*$= nuclear norm

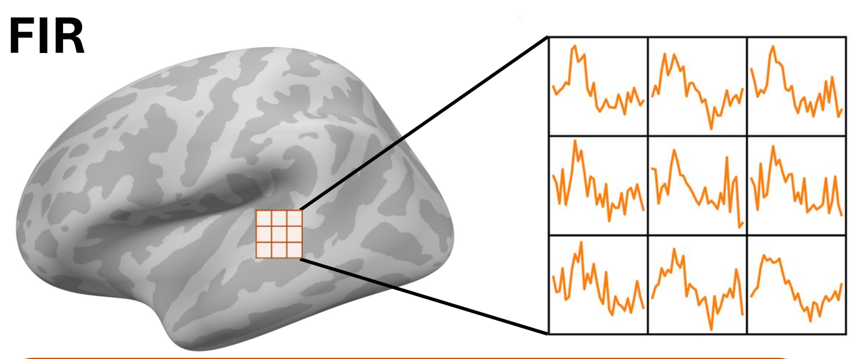
HRF



"Which voxels exhibit responses that match canonical hemodynamic responses to speech envelope tracking?"

Pros	Cons
Efficient and easily interpretable Clear interpretable model coefficients	Highly biased Assumes independence between voxels

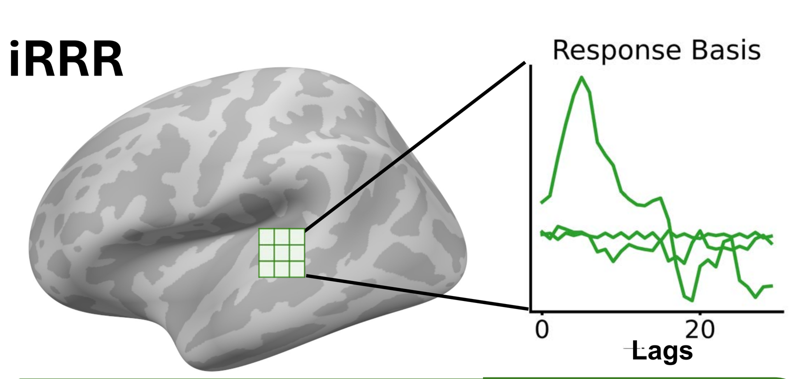
FIR



"How do the dynamics of voxel responses to speech envelope tracking vary across voxels?"

Pros	Cons
Most flexible for capturing HRF shape Clear interpretable model coefficients	Increased variability of estimates Not appropriate for studies focused on detecting activation

iRRR

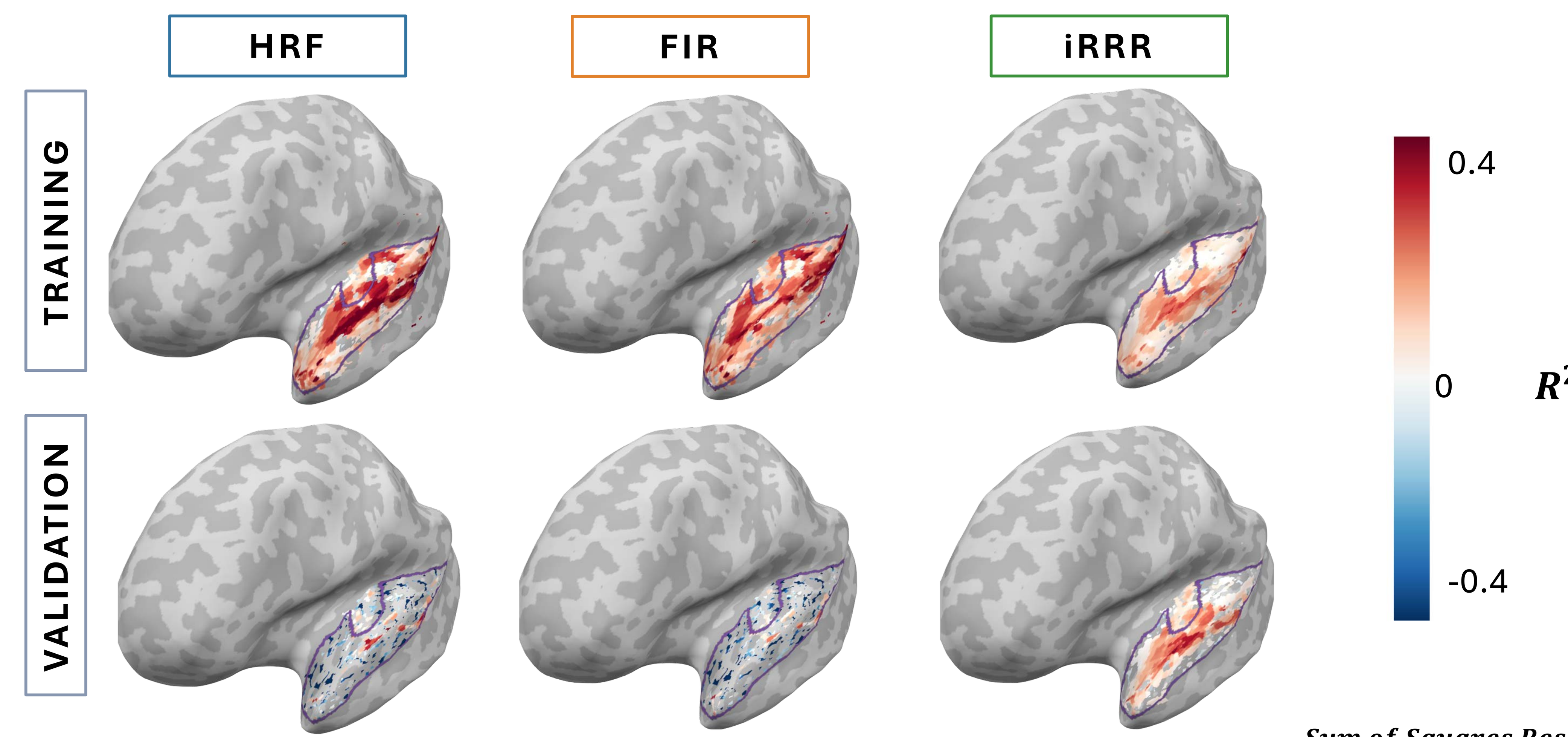


"What distributed patterns of voxel activity best encode speech information?"

Pros	Cons
Compromise between FIR and HRF models Reveals shared responses across voxels Reduces noise	Harder to interpret model coefficients iRRR not normally adapted to fMRI data analysis

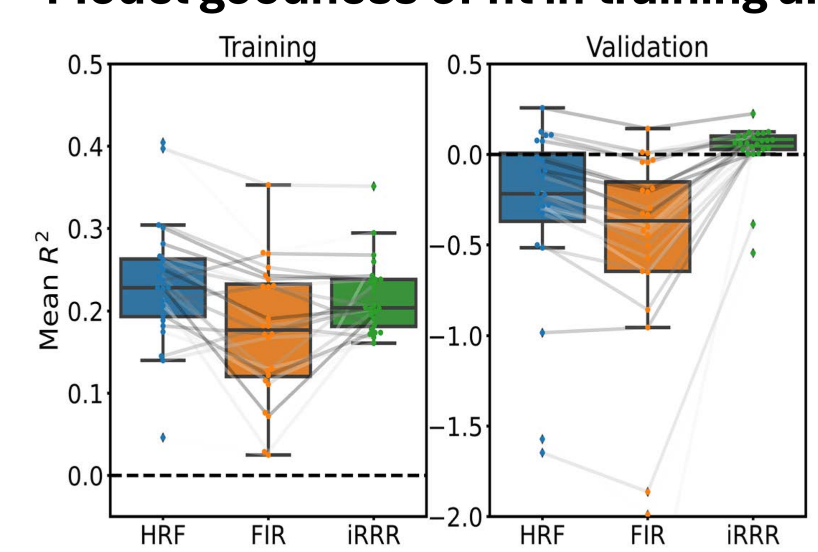
iRRR model demonstrates increased predictive performance

Model goodness of fit in training and validation data in an example subject



$$R^2 = 1 - \frac{\text{Sum of Squares Residuals}}{\text{Total Sum of Squares}}$$

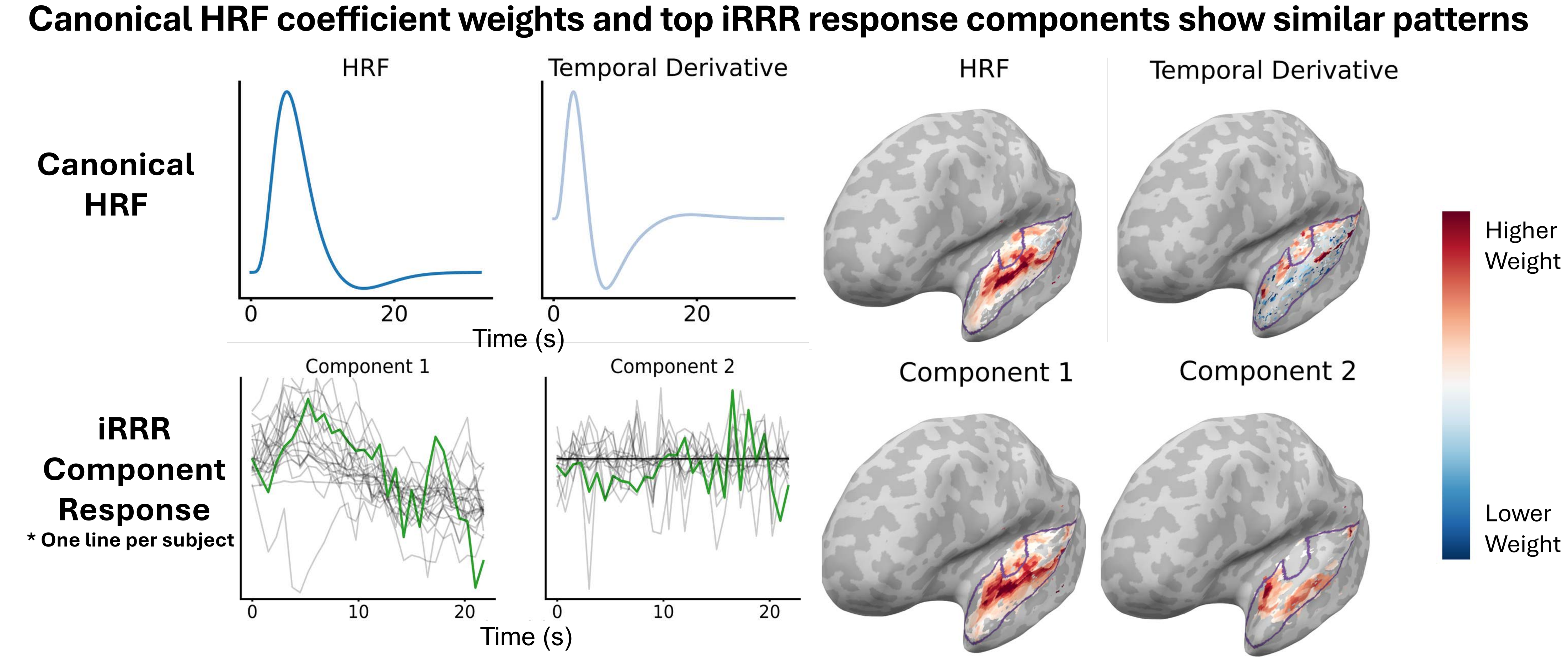
Model goodness of fit in training and validation data across subjects



Hold-out validation within a single subject (above) and across most subjects (left) demonstrates increased predictive performance for the iRRR model.

The HRF and top iRRR component show similar weight patterns across STG (below)

Canonical HRF coefficient weights and top iRRR response components show similar patterns



References

1. Pasley et al. (2012). *PLoS Biology.*
2. Hausfeld et al. (2024) *Communications Biology*
3. Oganian & Chang (2019). *Science Advances*
4. Lewis et al. (2016). *Proceedings of the National Academy of Sciences*
5. Li et al. (2019). *Biometrics*
6. Poldrack et al. (2024). *Handbook of Functional MRI Data Analysis*.
7. Scott et al. (2017). *Cognitive Neuroscience*

Funding Sources: R03 HD096098 (Perrachione), R03 DC014045 (Perrachione); F32 DC019531 (Scott); NSF GRFP (Moore); Boston University Kilachand Award, Department of Mathematics & Statistics (BU), Center for Systems Neuroscience (Stephen)

Contact: aamoore@bu.edu | https://sites.bu.edu/cnrlab/

Discussion

Our findings reveal that iRRR outperforms the predictive power of both canonical HRF and FIR approaches to modelling speech envelope encoding in STG. This suggests that iRRR's ability to capture shared response patterns across voxels and create low-dimensional representations of those patterns creates a model that generalizes better to unobserved data. To understand the functional contributions of areas within and beyond cortex on speech processing, a flexible model of speech envelope encoding is needed. iRRR's ability to reveal latent structures and adapt to the complexity of fMRI data makes it well-positioned to characterize speech encoding.