

# Using vowels to dissociate the representations of three distinct spectrotemporal features within and beyond human auditory cortex

Yongtian Ou<sup>1</sup>, Kendrick N. Kay<sup>1,2</sup>, Andrew J. Oxenham<sup>1</sup>

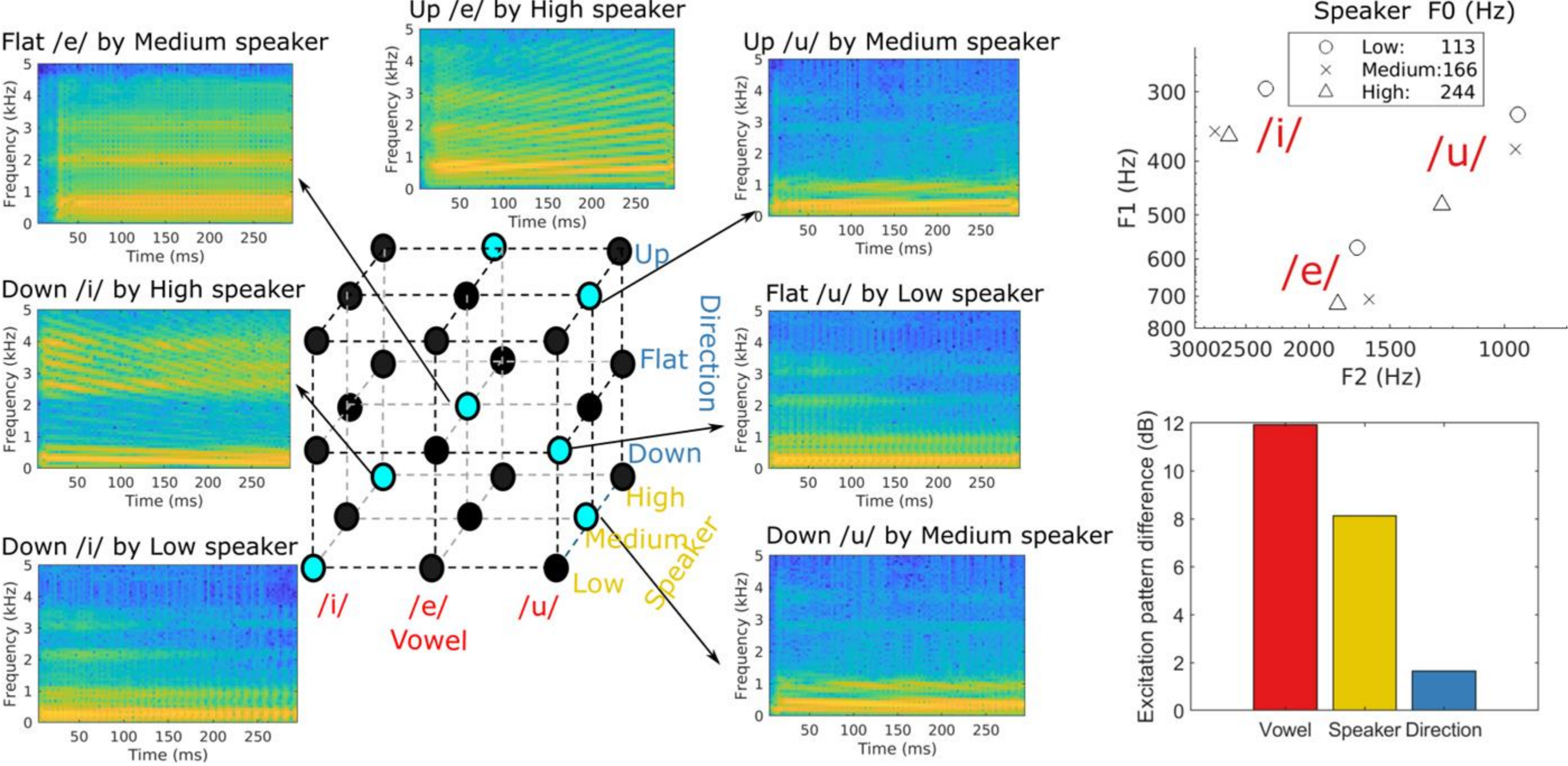
<sup>1</sup>Department of Psychology, University of Minnesota, USA <sup>2</sup>Center for Magnetic Resonance Research, Department of Radiology, University of Minnesota, USA

## Background

- Cortical representations of **spectral content**, **fundamental frequency (F0)**, and **spectrotemporal modulation**, three key acoustic features of speech, have mostly been found **within auditory cortex**<sup>1,2,3</sup>, but also **beyond auditory cortex** in some studies<sup>4,5</sup>.
- Although everyday communication often involves **concurrent processing** of all three features, the cortical representations of these features have been **mostly studied separately**.
- Using ultra-high-field (7T) fMRI, this study investigated the **relative strength** of these features' cortical representations **driven by the same stimulus** and **compared their tuning properties** within and beyond auditory cortex.

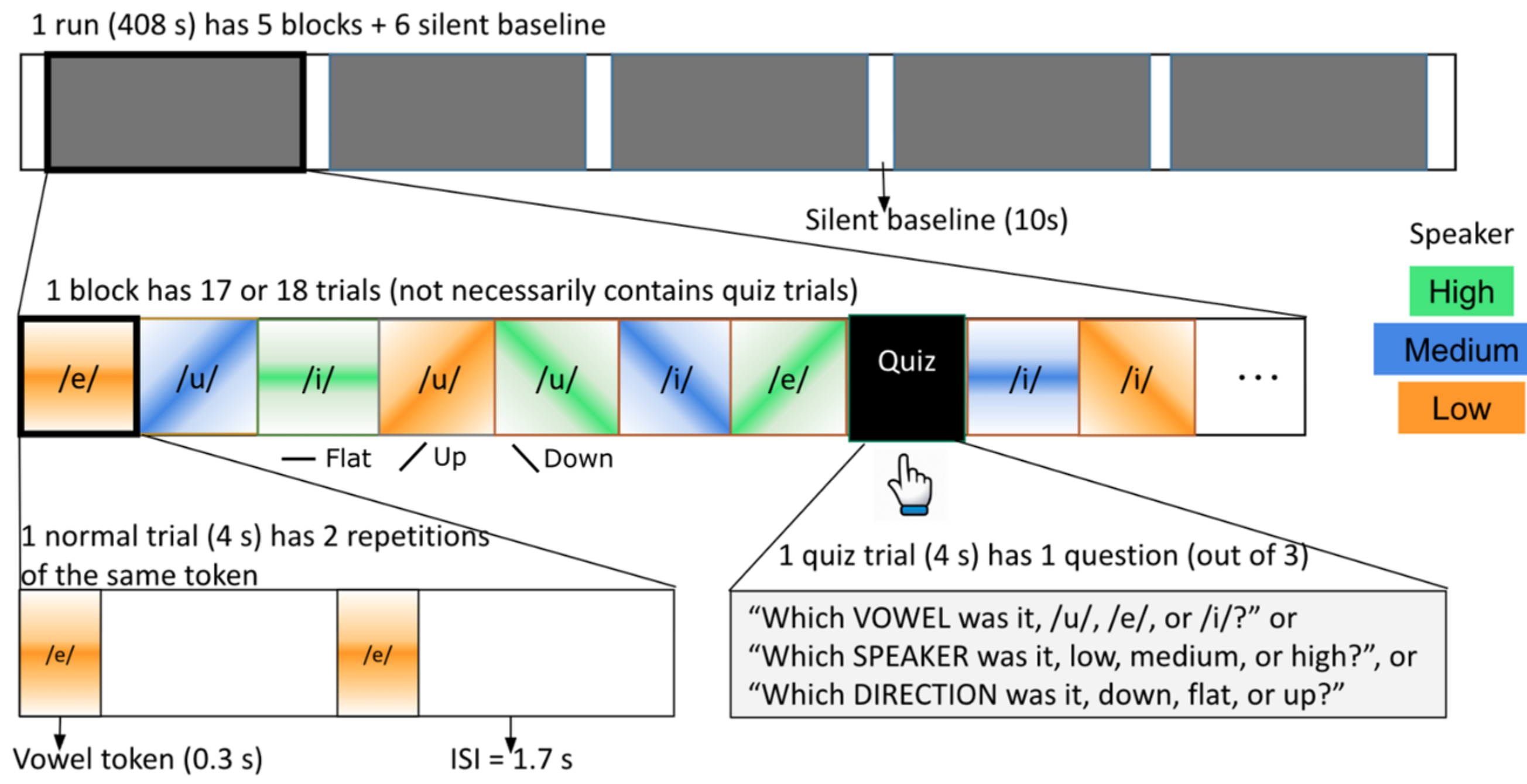
## Methods

### Experiment design & stimuli



- Vowel identity (probing spectral content): /u/, /e/, /i/
- Speaker identity (probing primarily F0): low (male), medium (female), high (female)
- Pitch glide direction (probing spectrotemporal modulation): one octave down, flat, one octave up

### Data acquisition



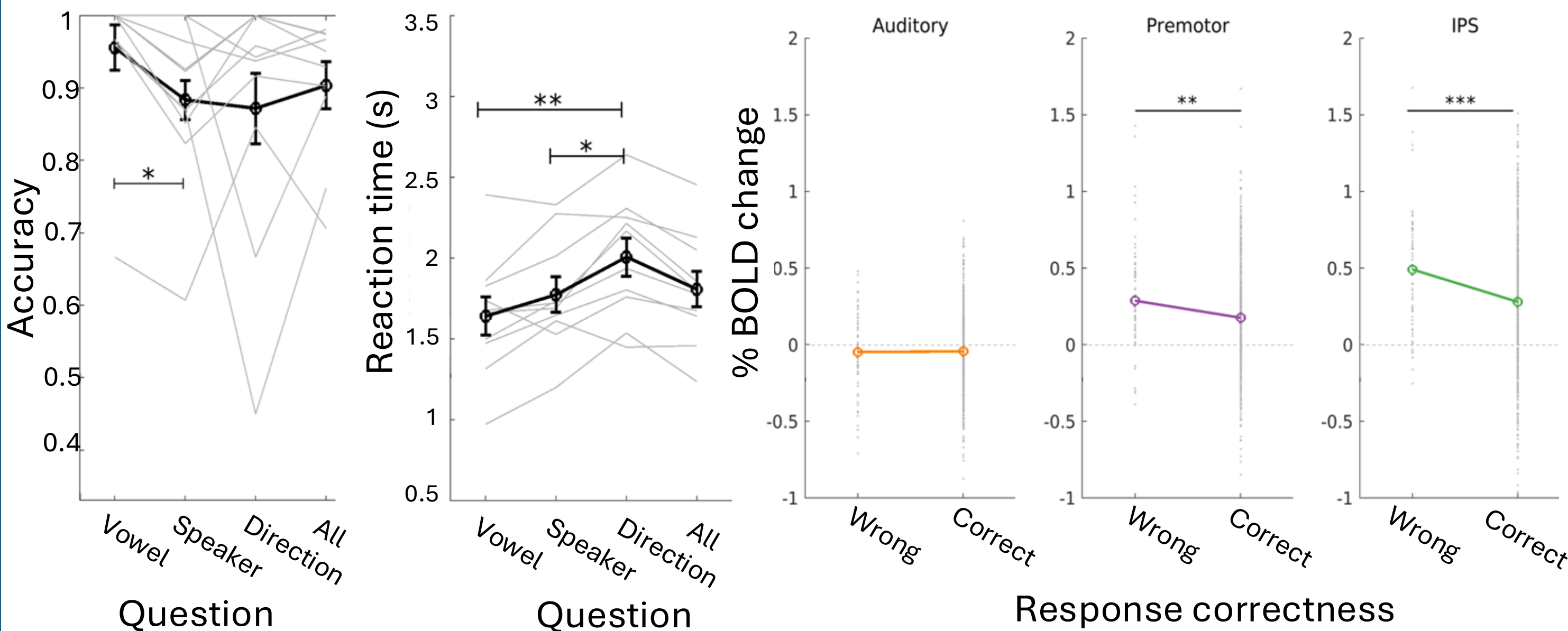
### Participants

- 10 normal-hearing participants aged 21-55, including 5 native tone-language speakers.

### Functional MRI

- 7T scanner** (Siemens Magnetom Active Shield); Gradient Echo EPI, **TR = 1.6s**, **1.8mm isotropic**, **whole-brain coverage**, continuous acquisition, partial Fourier = 7/8; iPAT = 2; multiband = 3
- 12 runs x 3 repetitions of the same condition per run = **36 repetitions per condition** per participant
- Audio: Sensimetrics S15 + custom filters; Visual: back-projected screen
- Preprocessed with custom scripts and GLMsingle<sup>6</sup> to obtain single trial beta estimates and resampled to mid-grey surface.

## Result: Quiz Trials

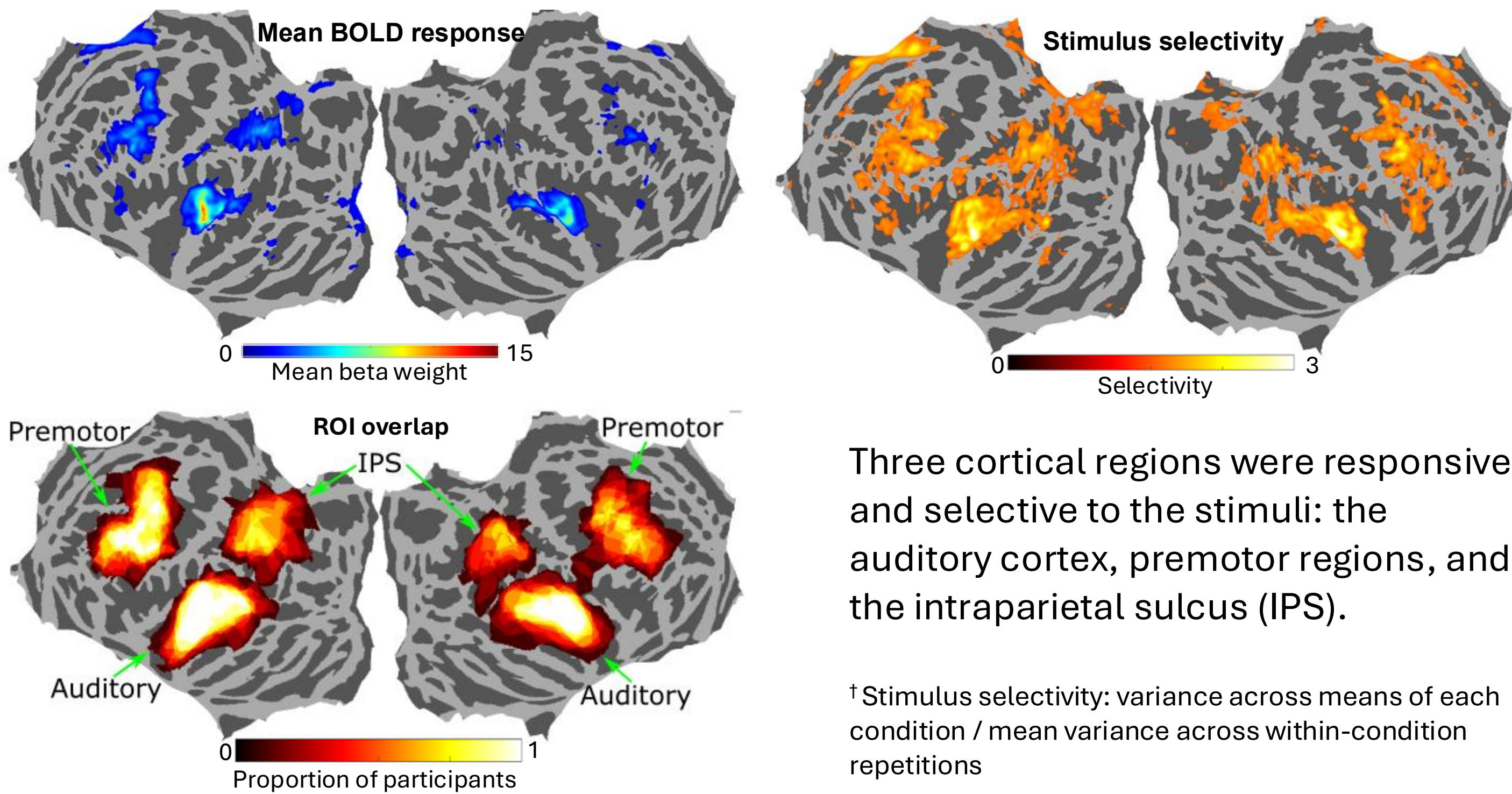


More accurate and faster vowel identification than speaker or direction identification

Higher BOLD responses in wrong trials than correct trials in premotor and IPS but not auditory cortex

## Main Results

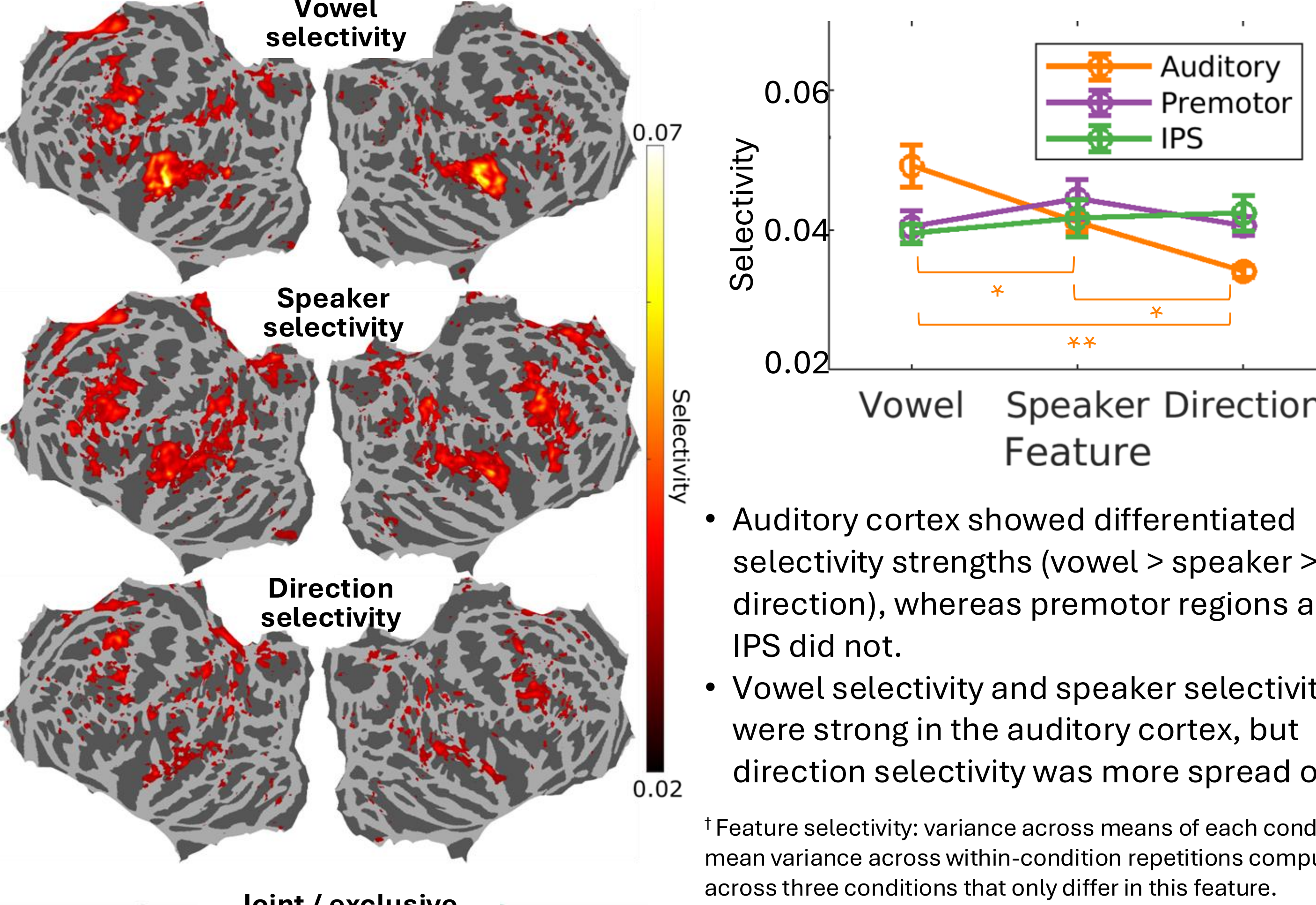
### Strong stimulus selectivity<sup>†</sup> found within and beyond auditory cortex



Three cortical regions were responsive and selective to the stimuli: the auditory cortex, premotor regions, and the intraparietal sulcus (IPS).

<sup>†</sup> Stimulus selectivity: variance across means of each condition / mean variance across within-condition repetitions

### Distinct strengths and spatial distributions of feature selectivity<sup>†</sup>



- Auditory cortex showed differentiated selectivity strengths (vowel > speaker > direction), whereas premotor regions and IPS did not.
- Vowel selectivity and speaker selectivity were strong in the auditory cortex, but direction selectivity was more spread out.

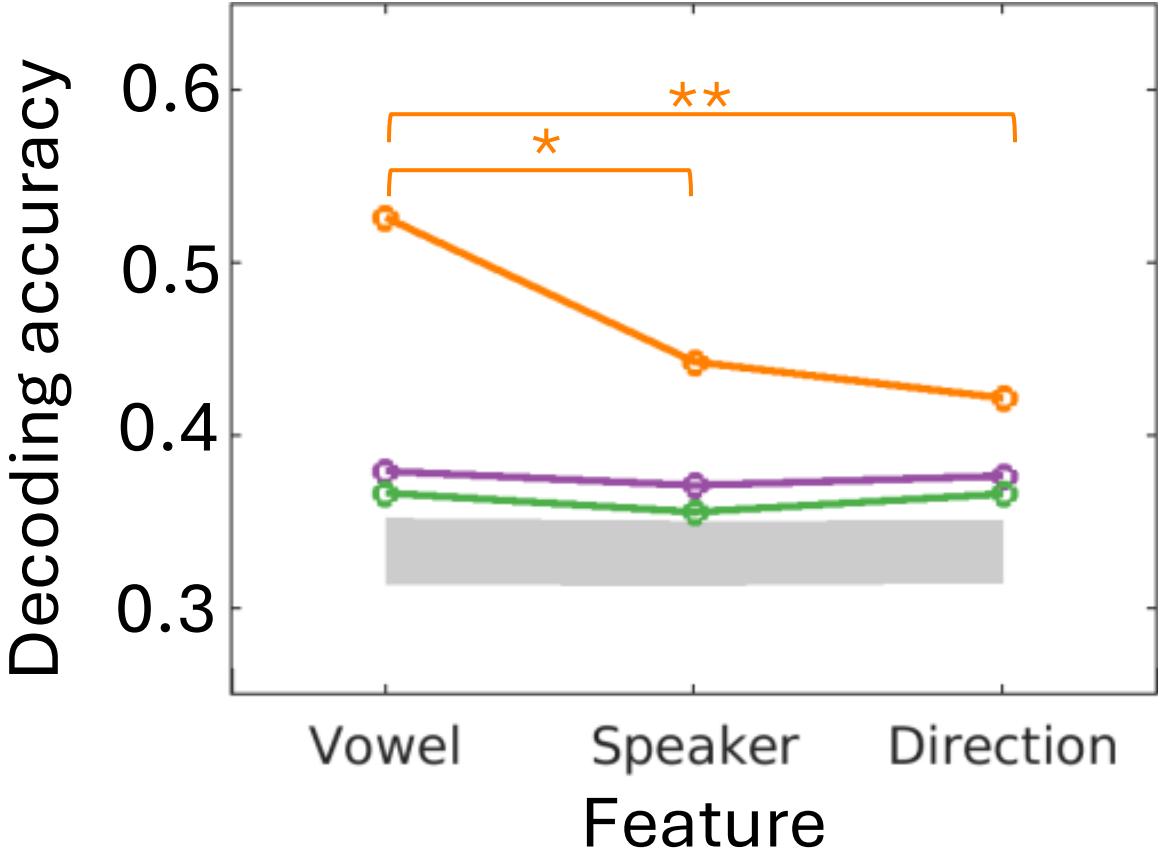
<sup>†</sup> Feature selectivity: variance across means of each condition / mean variance across within-condition repetitions computed across three conditions that only differ in this feature.

- Auditory cortex mainly showed joint tuning to both vowel and speaker identity.
- Superior temporal gyrus (STG) exhibited tuning to all three features.
- Exclusive and joint tuning involving speaker identity was observed in premotor regions.

<sup>†</sup> Color of the voxel indicates which feature(s) among the three it is significantly selective to, based on permutation test.

### Decoding accuracies<sup>†</sup> differ across the three features in auditory cortex but not beyond

- All decoding was significantly above chance.
- Auditory cortex showed the highest decoding accuracy, especially with vowel identity.
- Decoding accuracies in the premotor regions and IPS were uniform across features.



<sup>†</sup> Decoding accuracy is defined as how accurate a diagonal linear discriminant classifier trained on 11 runs can classify the vowel / speaker / direction of a trial in the left-out run. Chance level is 0.33.

\*p<.05, \*\*p<.01, \*\*\*p<.001

## Summary

- Auditory cortex demonstrated **differentiated** encoding across features (spectral content > F0 > spectrotemporal modulation), which may indicate its **acoustic-specific** nature that emphasizes spectral representation.
- By contrast, frontoparietal regions encoded these features **uniformly**, which may be associated with the features' equal importance in the identification task, suggesting their **goal-driven** property.

**References**

1. Saenz, M., & Langers, D. R. M. (2014). Tonotopic mapping of human auditory cortex. *Hearing Research*, 307, 42-52.
2. Allen, E. J., Mesik, J., Kay, K. N., & Oxenham, A. J. (2022). Distinct representations of tonotopy and pitch in human auditory cortex. *Journal of Neuroscience*, 42(3), 416-434.
3. Schönwiesner, M., & Zatorre, R. J. (2009). Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. *Proceedings of the National Academy of Sciences*, 106(34), 14611-14616.
4. Venezia, J. H., Richards, V. M., & Hickok, G. (2021). Speech-Driven Spectrotemporal Receptive Fields Beyond the Auditory Cortex. *Hearing Research*, 408, 108307.
5. Cheung, C., Hamilton, L. S., Johnson, K., & Chang, E. F. (2016). The auditory representation of speech sounds in human motor cortex. *eLife*, 5, e12577. <https://doi.org/10.7554/eLife.12577>
6. Prince, J. S., Charest, I., Kurzawski, J. W., Pyles, J. A., Tarr, M. J., & Kay, K. N. (2022). Improving the accuracy of single-trial fMRI response estimates using GLMsingle. *eLife*, 11, e77599.

Supported by  
NIH grant  
R01 DC005216



Scan QR code for digital copy