

Supplementary Materials for

Attentional Modulation of Hierarchical Speech Representations in a Multi-Talker Environment

Ibrahim Kiremitçi^{1,2}, Özgür Yılmaz^{2,3}, Emin Çelik^{1,2}, Mo Shahdloo^{2,4}, Alexander G. Huth^{5,6,7}, and Tolga Çukur^{1,2,3,7}

¹Neuroscience Program, Sabuncu Brain Research Center, Bilkent University, Ankara, TR-06800, Turkey

²National Magnetic Resonance Research Center (UMRAM), Bilkent University, Ankara, TR-06800, Turkey

³Department of Electrical and Electronics Engineering, Bilkent University, Ankara, TR-06800, Turkey

⁴Department of Experimental Psychology, Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford, OX3 9DU, U.K.

⁵Department of Neuroscience, The University of Texas at Austin, Austin, TX 78712, USA

⁶Department of Computer Science, The University of Texas at Austin, Austin, TX 78712, USA

⁷Helen Wills Neuroscience Institute, University of California, Berkeley, CA 94702, USA

Table of Contents:

- Supplementary Results 2
- Supplementary Discussion 4
- Supplementary References 5
- Supplementary Tables 6
 - Table S1: Inter-run head motion in the cocktail-party experiment.
 - Table S2a: Number of significantly predicted voxels within ROIs in the left hemisphere.
 - Table S2b: Number of significantly predicted voxels within ROIs in the right hemisphere.
 - Table S3: Hemispheric asymmetries in CI and gAI.
- Supplementary Figures 11
 - Figure S1: Prediction scores of voxelwise speech models.
 - Figure S2: Model-specific selectivity indices for non-perisylvian ROIs.
 - Figure S3a-e: Model-specific selectivity indices for subjects S1-S5.
 - Figure S4: Intrinsic selectivity profiles.
 - Figure S5: Representational complexity.
 - Figure S6: Cortical hierarchy of representational complexity.
 - Figure S7: Model-specific attention indices for non-perisylvian ROIs.
 - Figure S8a-e: Model-specific attention indices for subjects S1-S5.
 - Figure S9: Attentional modulation profiles for individual subjects.
 - Figure S10: Global attention indices.
 - Figure S11: Modulation hierarchies for individual subjects.
 - Figure S12a-e: Representation of unattended speech in subjects S1-S5.

Supplementary Results

Hierarchies in speech representations.

An emerging view is that speech features represented in the cortex grow progressively more complex towards downstream areas (Scott 2005; Hickok and Poeppel 2007; Rauschecker and Scott 2009; DeWitt and Rauschecker 2012; de Heer et al. 2017). Yet, a broad quantitative characterization of these selectivity gradients is lacking. To assess the selectivity in each stage of processing, we first measured a complexity index, (*CI*), that reflects whether an ROI is relatively tuned for low-level spectral or high-level semantic features (see Methods). *CI* across perisylvian and non-perisylvian cortex is displayed in Supplementary Fig. S5. Early auditory regions such as HG/HS have *CI* close to 0, whereas higher-order regions like PTR have *CI* tending to 1 as expected.

Next, we explored in detail finer scale gradients across the main auditory streams: dorsal and ventral stream (Rauschecker and Tian 2000; Hickok and Poeppel 2004, 2007; Scott 2005; Saur et al. 2008; Rauschecker and Scott 2009; Rauschecker 2011; Friederici 2012).

a) Dorsal stream. The dorsal stream emanates from primary auditory cortex (PAC) (Ueno et al. 2011; Friederici 2012; Bornkessel-Schlesewsky and Schlesewsky 2013), traverses planum temporale (Hickok and Poeppel 2004, 2007; Rauschecker and Scott 2009), and then projects to pars-operculum/BA44 (Friederici 2011, 2015; Bornkessel-Schlesewsky et al. 2015) and precentral-gyrus/BA6 (Bernal and Ardila 2009; Friederici 2011, 2015) in prefrontal cortex. The projections onto prefrontal cortex are suggested to be either direct or relayed via supramarginal-gyrus/BA40 in inferior parietal cortex (Hickok and Poeppel 2000; Catani et al. 2005; Frey et al. 2008; Scott et al. 2009; Friederici 2011). In the light of these reports, we examined variation of *CI* across three trajectories as shown in Figure S6a: left dorsal-1 (HG/HS_L → PT_L → (SMG_L) → POP_L), left dorsal-2 (HG/HS_L → PT_L → (SMG_L) → PreG_L) and right dorsal (HG/HS_R → PT_R → SMG_R). Left dorsal-1 and dorsal-2 were considered with or without inclusion of SMG. As shown in Figure S6b, we find significant increase in *CI* consistently in all subjects across the following left dorsal subtrajectories ($p < 0.05$): $CI_{HG/HS} < CI_{PT} < CI_{POP}$ and $CI_{HG/HS} < CI_{PT} < CI_{PreG}$. In contrast, we find no consistent pattern in *CI* across the right dorsal stream ($p > 0.05$). Note that PreG has greater articulatory selectivity in all subjects compared to POP ($p < 0.05$; see Supp. Fig. S3a-e). Furthermore, articulatory selectivity is dominant in PreG, while semantic selectivity is dominant in POP ($p < 0.05$; see Supp. Fig. S3a-e and Supp. Fig. S4). These results indicate that left dorsal-1 and left dorsal-2 differ in that the former carries relatively stronger semantic representations whereas the latter hosts articulatory representations.

b) Ventral stream. The ventral stream emanates from PAC (Ueno et al. 2011; Bornkessel-Schlesewsky and Schlesewsky 2013), passes through middle-anterior superior temporal cortex (Ueno et al. 2011; DeWitt and Rauschecker 2012; Bornkessel-Schlesewsky and Schlesewsky 2013), and then projects to pars-triangularis/BA45 (Ueno et al. 2011; Bornkessel-Schlesewsky et al. 2015) in prefrontal cortex. It is also suggested to traverse superior-inferior paths in temporal lobe (Hickok and Poeppel 2004, 2007). As such, we examined variation of *CI* across four trajectories (Figure S6a): left ventral-1 (HG/HS_L → mSTG_L → mSTS_L → MTG_L), left ventral-2 (HG/HS_L → mSTG_L → aSTG_L → PTR_L), right ventral-1 (HG/HS_R → mSTG_R → mSTS_R → MTG_R) and right ventral-2 (HG/HS_R → mSTG_R → aSTG_R → PTR_R). As shown in Figure S6b, we find significant increase in *CI* consistently in all subjects across the following subtrajectories ($p < 0.05$): $CI_{HG/HS} < CI_{mSTG} < CI_{aSTG}$ and $CI_{HG/HS} < CI_{mSTG} < CI_{mSTS}$ in the bilateral ventral stream. In contrast, there are no consistent differences between aSTG and PTR, and between MTG and mSTS ($p > 0.05$). Towards the end of ventral trajectories, semantic representations become dominant in all subjects (bilateral PTR and MTG; $p < 0.05$; see Supp. Fig. S3a-e and Supp. Fig. S4). These results indicate that representations in the ventral stream might be more symmetric across the two hemispheres compared to the dorsal stream.

Hemispheric asymmetries in speech representations.

Prior studies report right lateralization in spectral representations (Zatorre and Belin 2001; Obleser et al. 2008; Ding and Simon 2012; McGettigan and Scott 2012) and left lateralization in phonetic and semantic representations (DeWitt and Rauschecker 2012; McGettigan and Scott 2012). Thus, it is likely that speech-related cortical regions are tuned for more complex speech features in the left versus the right hemisphere. To test this prediction, we compared *CI* between the left and right counterparts of each ROI, with consistent selectivity for speech features in both hemispheres in each individual subject (see Methods). Table S3 lists the results of this across-hemisphere comparison. No consistent hemispheric asymmetry in representational complexity is observed across cortex ($p > 0.05$) except in mSTG, with higher CI in left hemisphere ($p < 0.05$). These results suggest a slight left-hemispheric bias in representational complexity across intermediate stages during passive-listening tasks.

Supplementary Discussion

Speech Representations

Prior studies using controlled auditory stimuli have suggested a hierarchical organization of speech representations across cortex (Binder et al. 2000; Hickok and Poeppel 2004; Liebenthal et al. 2005; Obleser et al. 2007; DeWitt and Rauschecker 2012). Recent studies using natural speech stimuli corroborate this view where downstream regions are notably selective for more complex speech features than early regions in speech-related cortex (de Heer et al. 2017; Brodbeck et al. 2018a, 2018b). Motivated by these results, here we quantitatively examined gradients in complexity of speech representations across speech-related cortex. Overall, we find that representational complexity increases gradually across the left dorsal and bilateral ventral streams.

An important question in neurolinguistics is the precise functional roles of the dorsal and ventral streams for natural speech representations in the human brain. During passive listening of natural stories, we find that dominant articulatory selectivity manifests (PreG) in one end of the dorsal stream whereas semantic selectivity manifests in both ends (MTG and PTR) of the ventral stream. These results confirm the view that the dorsal stream is implicated in sound-to-articulation mapping, whereas the ventral stream is primarily implicated in sound-to-meaning mapping (Hickok and Poeppel 2007, 2015; Rauschecker and Scott 2009; Friederici 2011; Rauschecker 2011).

Linguistic representations are commonly considered to be left lateralized while acoustic representations (spectral) are suggested to be right lateralized (Zatorre and Belin 2001; Ding and Simon 2012; DeWitt and Rauschecker 2012; McGettigan and Scott 2012). Here we quantitatively assessed hemispheric asymmetries in multi-level speech representations along the dorsal and ventral streams. We report a left-hemispheric dominance in the dorsal stream that becomes relatively more apparent towards the end stages (PreG and POP). However, regions along the ventral stream have similar selectivity profiles in both hemispheres, with a slight left-hemispheric bias in their representational complexity in intermediate stages. Hence, our results indicate a left-lateralized organization in the dorsal stream and a predominantly bilateral organization in the ventral stream (Hickok and Poeppel 2007).

Supplementary References

- Bernal B, Ardila A. 2009. The role of the arcuate fasciculus in conduction aphasia. *Brain*. 132:2309-2316.
- Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. 2000. Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex*. 10:512-528.
- Bornkessel-Schlesewsky I, Schlesewsky M. 2013. Reconciling time, space and function: a new dorsal-ventral stream model of sentence comprehension. *Brain Lang*. 125:60–76
- Bornkessel-Schlesewsky I, Schlesewsky M, Small SL, Rauschecker JP. 2015. Neurobiological roots of language in primate audition: common computational properties. *Trends Cogn Sci*. 19:142-150
- Brodbeck C, Presacco A, Simon JZ. 2018a. Neural source dynamics of brain responses to continuous stimuli: Speech processing from acoustics to comprehension. *NeuroImage*. 172:162-174
- Brodbeck C, Hong LE, Simon JZ. 2018b. Rapid transformation from auditory to linguistic representations of continuous speech. *Curr Biology*. 28:3976-3983.
- de Heer WA, Huth AG, Griffiths TL, Gallant JL, Theunissen FE. 2017. The hierarchical cortical organization of human speech processing. *J Neurosci*. 37:6539–6557.
- 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.
- DeWitt I, Rauschecker JP. 2012. Phoneme and word recognition in the auditory ventral stream. *Proc Natl Acad Sci*. 109:E505–E514
- Ding N, Simon JZ. 2012. Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *J Neurophysiol*. 107:78–89.
- Catani M, Jones DK, ffytche DH. 2005. Perisylvian language networks of the human brain. *Ann Neurol*. 57:8–16
- Frey S, Campbell J, Pike G, Petrides M. 2008. Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *J Neurosci*. 28:11435-11444
- Friederici AD. 2011. The brain basis of language processing: from structure to function. *Physiol Rev*. 91:1357-1392
- Friederici AD. 2012. The cortical language circuit: from auditory perception to sentence comprehension. *Trends Cogn Sci*. 16:262-268
- Friederici AD. 2015. White-matter pathways for speech and language processing. In: *Handbook of clinical neurology*. Elsevier. 129:177-186
- Hickok G, Poeppel D. 2000. Towards a functional neuroanatomy of speech perception. *Trends Cogn Sci* 4:131-138.
- Hickok G, Poeppel D. 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*. 92:67-99
- Hickok G, Poeppel D. 2007. The cortical organization of speech processing. *Nat Rev Neurosci*. 8:393–402
- Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. 2005. Neural substrates of phonemic perception. *Cereb Cortex*. 15:1621-1631.
- McGettigan C, Scott SK. 2012. Cortical asymmetries in speech perception: what's wrong, what's right and what's left? *Trends Cogn Sci*. 16:269-276
- Obleser J, Zimmermann J, Van Meter J, Rauschecker JP. 2007. Multiple stages of auditory speech perception reflected in event-related fMRI. *Cereb Cortex*. 17:2251-2257
- Obleser J, Eisner F, Kotz SA. 2008. Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J Neurosci*. 28:8116-8123.
- Rauschecker JP, Tian B. 2000. Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc Natl Acad Sci USA* 97:11800–11806
- Rauschecker JP, Scott SK. 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci*. 12:718–724.
- Rauschecker JP. 2011. An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hear Res*. 271:16-25
- Saur D, Kreher BW, Schnell S, Kümmerer D, Kellmeyer P, Vry MS, Umarova R, Musso M, Glauche V, Abel S, et al. 2008. Ventral and dorsal pathways for language. *Proc Natl Acad Sci USA*. 105:18035–40

- Scott SK. 2005. Auditory processing—speech, space and auditory objects. *Curr Opin Neurobiol.* 15:197-201
- Scott SK, McGettigan C, Eisner F. 2009. A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. *Nat Rev Neurosci.* 10: 295-302.
- Ueno T, Saito S, Rogers TT, Lambon Ralph MA. 2011. Lichtheim 2: Synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal–ventral language pathways. *Neuron.* 72:385–396.
- Zatorre RJ, Belin P. 2001. Spectral and temporal processing in human auditory cortex. *Cereb Cortex.* 11:946–953

Supplementary Tables

Table S1: Inter-run head motion in the cocktail-party experiment

		Rx	Ry	Rz	Tx	Ty	Tz
S1	Run2	0.0070	-0.0222	-0.0047	-0.8205	-1.8318	4.4034
	Run3	0.0043	-0.0193	-0.0075	0.3561	-1.1648	3.6750
	Run4	0.0081	-0.0400	-0.0122	0.2514	-2.3087	5.9718
	Run5	0.0059	-0.0258	-0.0040	-1.2676	-1.2530	4.4150
	Run6	0.0060	-0.0225	-0.0058	-0.5274	-1.9113	4.5966
	S2	Run2	-0.0099	-0.0067	-0.0099	0.6676	-0.2590
Run3		-0.0088	-0.0026	0.0018	-0.5063	0.6162	-0.6334
Run4		-0.0128	-0.0118	-0.0103	-0.2114	0.3525	-0.3772
Run5		-0.0119	-0.0044	-0.0047	-0.2807	0.7589	-1.1399
Run6		-0.0101	-0.0066	-0.0118	0.7884	-0.1214	-0.4955
S3		Run2	0.0321	0.0118	0.0060	-0.3023	-0.1689
	Run3	0.0280	0.0160	0.0059	-0.1685	-0.2995	1.9887
	Run4	0.0273	0.0139	0.0092	-0.7141	0.5719	2.3134
	Run5	0.0252	0.0129	-0.0008	0.3123	-0.5163	2.1007
	Run6	0.0263	0.0135	0.0055	-0.2489	-0.1244	2.4375
	S4	Run2	-0.0009	0.0062	-0.0028	0.7449	-0.1245
Run3		-0.0014	0.0042	0.0005	0.4494	0.5528	-0.5016
Run4		0.0101	-0.0076	-0.0030	-0.0453	-0.7291	3.1879
Run5		-0.0006	0.0019	-0.0051	1.0548	-0.5008	-0.0154
Run6		-0.0010	0.0032	-0.0051	1.2785	-0.3303	-0.3824
S5		Run2	0.0106	-0.0133	-0.0164	3.3092	-1.8717
	Run3	0.0181	-0.0245	-0.0150	3.8181	-1.6553	5.9240
	Run4	0.0097	-0.0330	-0.0067	2.0298	-0.6607	5.2784
	Run5	0.0156	-0.0299	-0.0011	0.7572	-0.5795	5.8197
	Run6	0.0079	-0.0250	-0.0202	5.3338	-1.7013	3.8040

Absolute motion between temporal-mean volume of each run (Run2-Run6) and temporal-mean volume of the first run (Run1) is calculated during the cocktail-party experiment. Motion estimates are obtained in subjects S1-S5 using FMRIB's Linear Image Registration Tool (FLIRT) (Jenkinson and Smith 2001). Rotational motion (Rx, Ry, Rz) and translational motion (Tx, Ty, Tz) are reported along the X, Y and Z axes. Rotations are expressed in radians, and translations are expressed in mm.

Table S2a: Number of significantly predicted voxels within ROIs in the left hemisphere.

	Spectrally-sel.		Articulatorily-sel.		Semantically-sel.		Speech-sel.	
HG/HS _L	59.0	(±12.9)	63.9	(±8.3)	0.0	(±0.0)	101.5	(±15.0)
PT _L	29.7	(±7.4)	45.5	(±6.4)	20.0	(±3.5)	66.3	(±7.7)
pSF _L	22.0	(±5.5)	34.0	(±5.7)	6.1	(±3.8)	49.2	(±6.5)
STG _L	32.3	(±8.6)	163.8	(±27.6)	100.7	(±30.1)	280.2	(±40.2)
aSTG _L	6.4	(±3.4)	26.6	(±8.6)	42.4	(±9.8)	63.4	(±9.5)
mSTG _L	7.6	(±3.0)	82.5	(±14.0)	14.0	(±7.7)	92.7	(±18.3)
pSTG _L	12.3	(±4.9)	41.8	(±8.2)	30.4	(±9.9)	66.4	(±13.5)
STSL	30.9	(±15.7)	200.2	(±22.8)	330.4	(±55.3)	426.7	(±59.1)
aSTSL	10.2	(±3.9)	45.6	(±9.8)	75.0	(±12.3)	97.1	(±10.8)
mSTSL	12.0	(±6.5)	70.8	(±11.4)	87.6	(±16.6)	123.1	(±17.2)
pSTSL	12.3	(±8.1)	88.6	(±21.3)	165.6	(±22.7)	205.7	(±30.4)
MTG _L	7.1	(±4.8)	38.5	(±7.2)	137.3	(±41.3)	160.6	(±34.3)
SMG _L	24.9	(±7.70)	33.5	(±6.4)	72.7	(±22.9)	111.6	(±20.4)
AG _L	2.1	(±1.9)	4.2	(±2.4)	156.6	(±39.1)	157.6	(±39.7)
IPS _L	0.0	(±0.00)	0.0	(±0.0)	65.0	(±10.9)	65.0	(±10.9)
SPS _L	2.2	(±1.8)	5.8	(±1.5)	44.8	(±11.3)	49.2	(±10.6)
PrCL	0.0	(±0.00)	9.1	(±5.1)	126.9	(±30.8)	129.9	(±31.7)
PCC _L	0.0	(±0.00)	0.0	(±0.0)	34.1	(±6.8)	34.1	(±6.8)
POS _L	0.0	(±0.00)	13.3	(±8.4)	58.7	(±18.2)	66.2	(±20.0)
POP _L	15.0	(±8.5)	24.6	(±6.7)	42.7	(±10.1)	64.8	(±9.9)
PTR _L	13.8	(±5.2)	31.8	(±7.5)	74.7	(±17.4)	98.3	(±17.2)
IFS _L	9.4	(±6.9)	26.3	(±9.1)	109.1	(±28.4)	124.4	(±30.0)
MFG _L	12.5	(±6.5)	38.2	(±10.5)	183.2	(±60.0)	216.1	(±56.4)
MFS _L	10.6	(±5.9)	25.1	(±9.1)	30.5	(±10.9)	57.6	(±8.3)
SFS _L	0.0	(±0.00)	2.4	(±2.1)	110.7	(±29.1)	111.5	(±29.1)
SFG _L	44.9	(±19.8)	59.9	(±12.7)	286.7	(±59.9)	327.6	(±62.6)
PreG _L	7.4	(±2.2)	41.9	(±7.5)	26.5	(±8.4)	60.4	(±12.2)
mOTS _L	0.0	(±0.0)	0.0	(±0.00)	26.3	(±3.1)	26.3	(±3.1)

Significantly predicted voxels by spectral, articulatory and semantic models are listed ($q(\text{FDR}) < 10^{-5}$, t-test; mean \pm sem across subjects). The union of these voxels are taken as speech-selective voxels. Only ROIs in the left hemisphere and with at least 10 speech-selective voxels in each individual subject are listed. Subscripted “L” indicates left hemisphere.

Table S2b: Number of significantly predicted voxels within ROIs in the right hemisphere.

	Spectrally-sel.		Articulatorily-sel.		Semantically-sel.		Speech-sel.	
HG/HS _R	47.4	(±5.6)	45.5	(±7.8)	0.0	(±0.0)	77.1	(±6.4)
PT _R	27.0	(±4.2)	29.6	(±3.7)	2.2	(±1.8)	41.7	(±3.4)
STG _R	56.9	(±8.5)	165.8	(±26.8)	55.1	(±8.7)	201.5	(±21.2)
aSTG _R	8.6	(±4.1)	30.5	(±11.8)	19.7	(±7.5)	47.4	(±12.6)
mSTG _R	12.3	(±4.0)	52.7	(±11.7)	7.0	(±2.0)	61.4	(±9.0)
pSTG _R	21.4	(±4.2)	57.2	(±11.0)	14.4	(±1.9)	71.6	(±10.2)
STS _R	39.9	(±12.1)	140.1	(±23.1)	338.3	(±45.6)	423.4	(±42.8)
aSTS _R	4.7	(±2.2)	30.0	(±8.0)	55.6	(±12.0)	71.1	(±13.1)
mSTS _R	24.6	(±8.9)	88.5	(±15.4)	86.1	(±24.1)	140.4	(±23.3)
pSTS _R	10.8	(±2.8)	25.0	(±4.6)	209.4	(±13.8)	224.2	(±12.5)
MTG _R	16.7	(±8.3)	29.5	(±9.1)	93.5	(±20.0)	114.7	(±20.3)
SMG _R	21.0	(±4.0)	28.2	(±6.8)	34.4	(±14.0)	71.0	(±13.7)
AG _R	2.1	(±1.8)	13.4	(±7.5)	137.1	(±35.6)	146.9	(±39.4)
IPS _R	0.0	(±0.0)	0.0	(±0.0)	33.4	(±4.0)	33.4	(±4.0)
SPS _R	2.0	(±1.7)	8.3	(±3.3)	66.0	(±17.1)	68.6	(±17.7)
PrC _R	0.0	(±0.0)	15.0	(±8.1)	107.7	(±29.9)	113.9	(±33.6)
PCC _R	0.0	(±0.0)	2.0	(±1.8)	36.2	(±8.4)	37.2	(±8.0)
POS _R	0.0	(±0.0)	18.3	(±9.5)	59.1	(±10.0)	73.5	(±15.5)
PTR _R	7.4	(±3.2)	18.3	(±6.1)	68.5	(±7.5)	80.7	(±7.2)
IFS _R	23.8	(±2.7)	41.5	(±12.0)	100.3	(±20.0)	138.3	(±21.6)
MFG _R	15.6	(±2.7)	51.1	(±15.3)	162.5	(±58.4)	218.2	(±50.6)
MFS _R	11.8	(±8.4)	29.4	(±9.7)	33.7	(±8.3)	70.5	(±11.4)
SFS _R	0.0	(±0.0)	12.6	(±5.5)	102.5	(±21.9)	109.3	(±21.6)
SFG _R	38.6	(±11.2)	75.4	(±21.7)	230.1	(±58.6)	295.3	(±60.5)

Significantly predicted voxels by spectral, articulatory and semantic models are listed ($q(\text{FDR}) < 10^{-5}$, t-test; mean \pm sem across subjects). The union of these voxels are taken as speech-selective voxels. Only ROIs in the right hemisphere and with at least 10 speech-selective voxels in each individual subject are listed. Subscripted “R” indicates right hemisphere.

Table S3: Hemispheric asymmetries in CI and gAI.

	CI	gAI
HG/HS	N, 0.0	N, 0.1
PT	N, 0.1	N, 0.1
aSTG	N, 0.2	N, 0.1
mSTG	L, 0.1	L, 0.2
pSTG	N, 0.0	N, 0.2
aSTS	N, 0.0	N, 0.1
mSTS	N, 0.1	N, 0.2
pSTS	N, 0.1	N, 0.1
MTG	N, 0.1	N, 0.1
SMG	N, 0.2	N, 0.1
AG	N, 0.0	N, 0.0
IPS	N, 0.0	N, 0.2
SPS	N, 0.0	N, 0.0
PrC	N, 0.0	N, 0.0
PCC	N, 0.0	N, 0.1
POS	N, 0.0	N, 0.0
PTR	N, 0.1	N, 0.1
IFS	N, 0.1	N, 0.2
MFG	N, 0.1	N, 0.1
MFS	N, 0.0	N, 0.0
SFS	N, 0.0	N, 0.0
SFG	N, 0.0	N, 0.0

"R", "L" and "N" indicate right-hemispheric, left-hemispheric and no-hemispheric dominance in indices respectively. The numbers show the effect-size of the significance tests, based on the difference between the left and right hemispheres. Only differences that are consistently significant in each individual subject are taken as significant ($p < 0.05$). ROIs are reported with at least 10 speech-selective voxels consistently in each hemisphere and in each subject. ($p < 0.05$).

Supplementary Figures

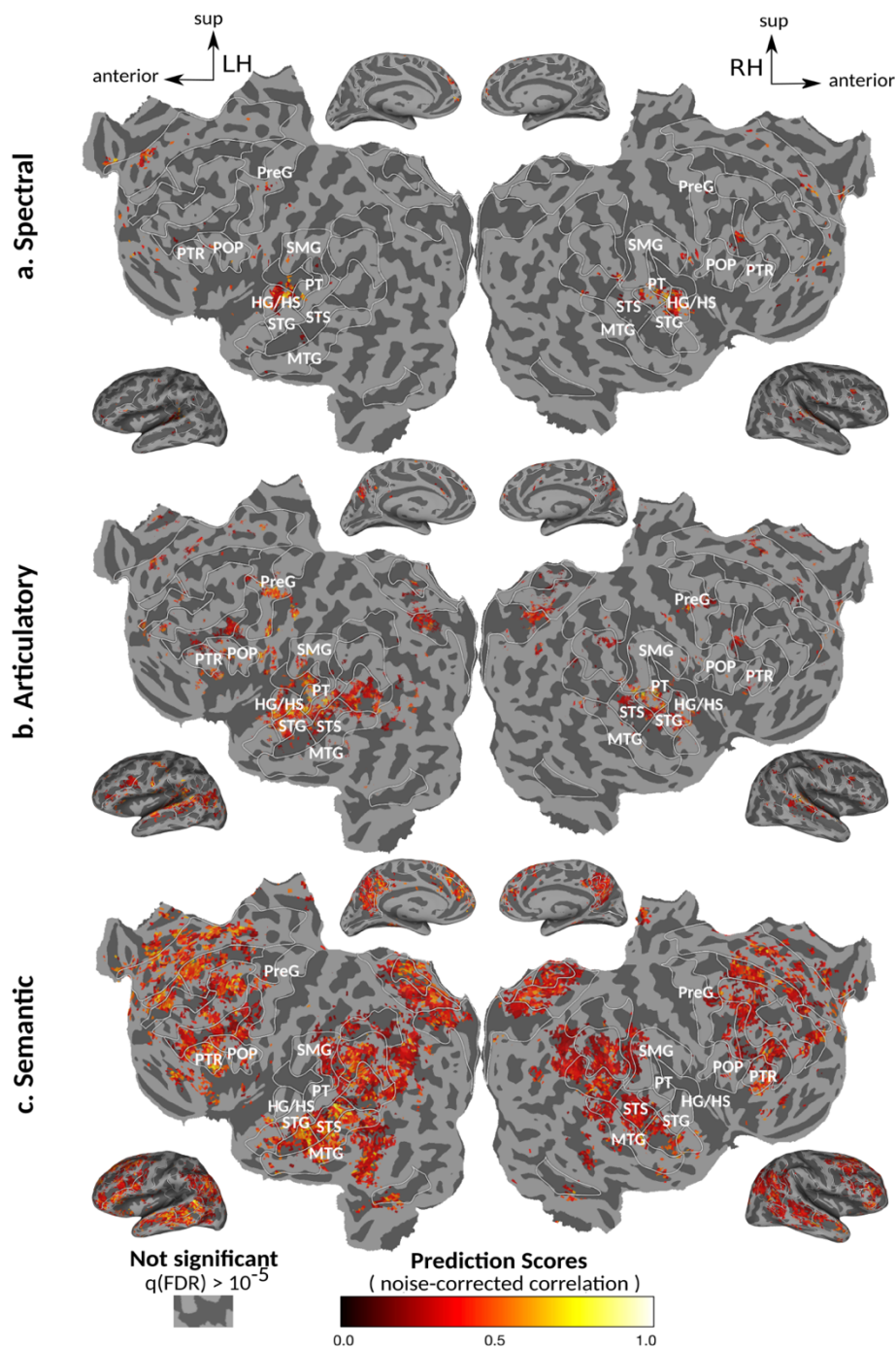


Figure S1: Prediction scores of voxelwise speech models. **a. Spectral model.** Prediction scores for voxels significantly predicted by the spectral model are plotted on the flattened cortical surface of subject S4. Medial and lateral views of the inflated hemispheres are also shown above and below the flatmaps, respectively. Colors indicate the value of the noise-corrected prediction scores (see legend). White lines encircle ROIs that are found based on an automatic atlas-based cortical parcellation. Labels of some relevant ROIs are shown (see Methods for ROI abbreviations). In this subject, significantly predicted voxels by the spectral model lie mainly in the dorsal aspect of the superior temporal cortex. **b. Articulatory model.** Significantly predicted voxels by the articulatory model are located mainly in the dorsal and lateral aspects of the superior temporal cortex extending to STS, and in inferior frontal regions. **c. Semantic model.** Significantly predicted voxels by the semantic model are broadly distributed across cortex covering most of the temporal, prefrontal and parietal cortical regions except the early auditory cortex.

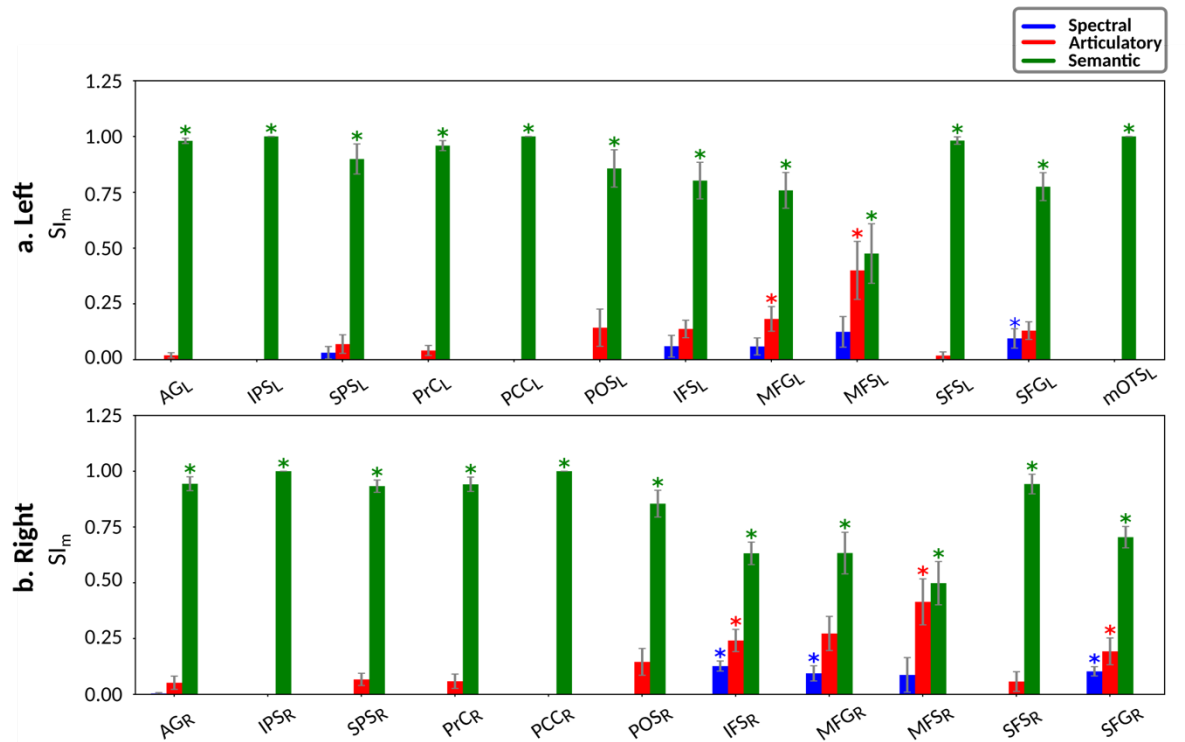


Figure S2: Model-specific selectivity indices for non-perisylvian ROIs. Single-voxel prediction scores on passive-listening data were used to quantify the degree of selectivity of each ROI to the underlying model features under passive listening. Prediction scores for a given model were averaged across speech-selective voxels within each ROI, and then normalized such that the cumulative prediction score from all models summed to 1. The resultant measure was taken as a model-specific selectivity index, (SI_m). SI_m is in the range of [0, 1], where higher values indicate stronger selectivity for the underlying model. Bar plots display SI_m for spectral, articulatory and semantic models (mean \pm sem across subjects). Significant indices are marked with *, colored according to the model ($p < 0.05$, bootstrap test; see Supp. Fig. S3a-e for selectivity indices of individual subjects). Only ROIs in non-perisylvian cortex are displayed. **a.** ROIs in LH. **b.** ROIs in RH. mOTS_R that did not have consistent speech selectivity in individual subjects was excluded (see Methods). These results show that selectivity is primarily semantic in most of the non-perisylvian ROIs.

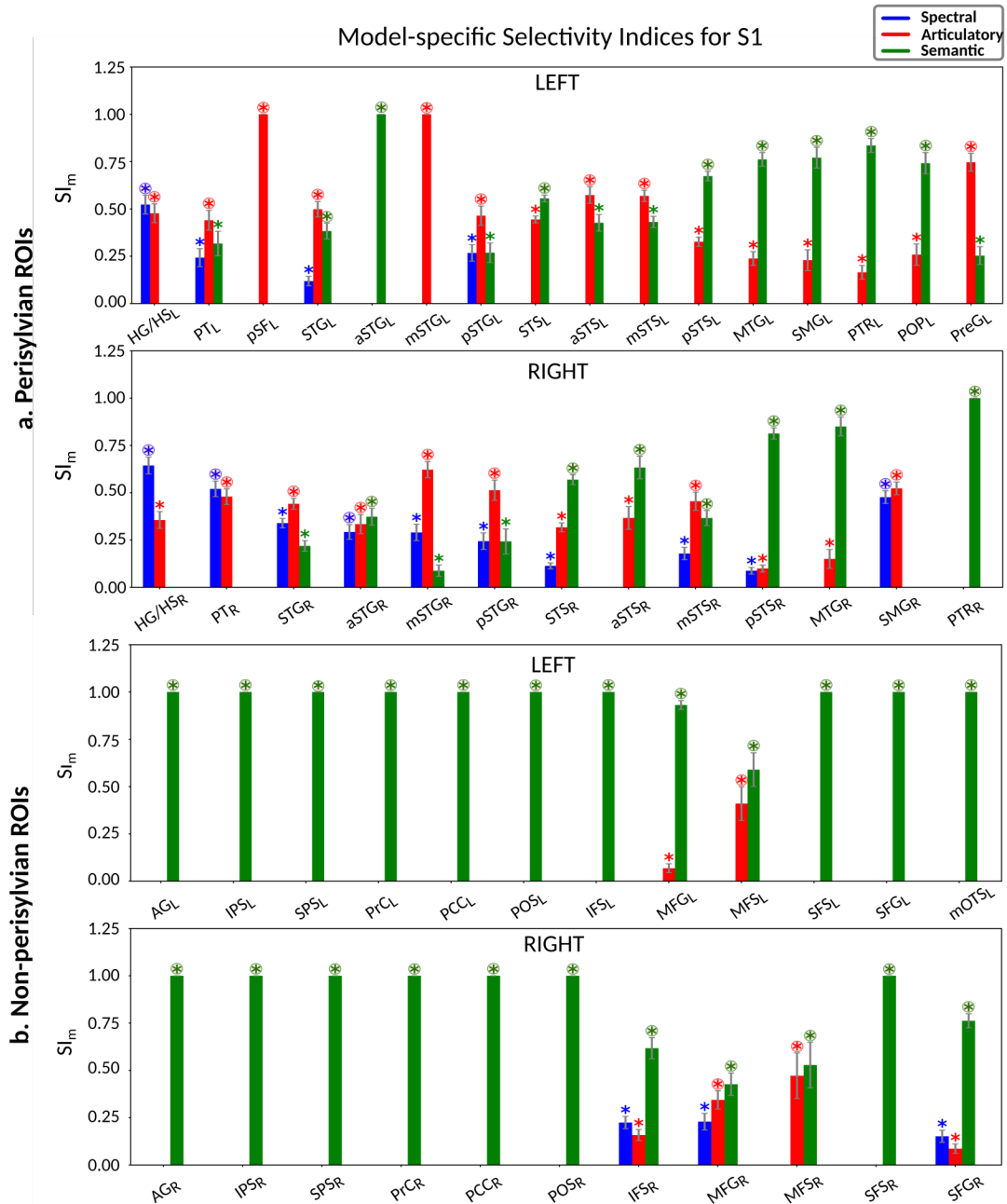


Figure S3a: Model-specific selectivity indices for subject S1. Single-voxel prediction scores on passive-listening data were used to quantify the degree of selectivity of each ROI to the underlying model features under passive listening. Prediction scores for a given model were averaged across speech-selective voxels within each ROI, and then normalized such that the cumulative prediction score from all models summed to 1. The resultant measure was taken as a model-specific selectivity index, (SI_m). SI_m is in the range of [0, 1], where higher values indicate stronger selectivity for the underlying model. Bar plots display SI_m for spectral, articulatory and semantic models (mean \pm sem across ROI's speech-selective voxels). Significant indices are marked with *, colored according to the model and encircled if they are dominant within the ROI ($p < 0.05$, bootstrap test). **a. Perisylvian ROIs.** ROIs in perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively. **b. Non-perisylvian ROIs.** ROIs in non-perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively.

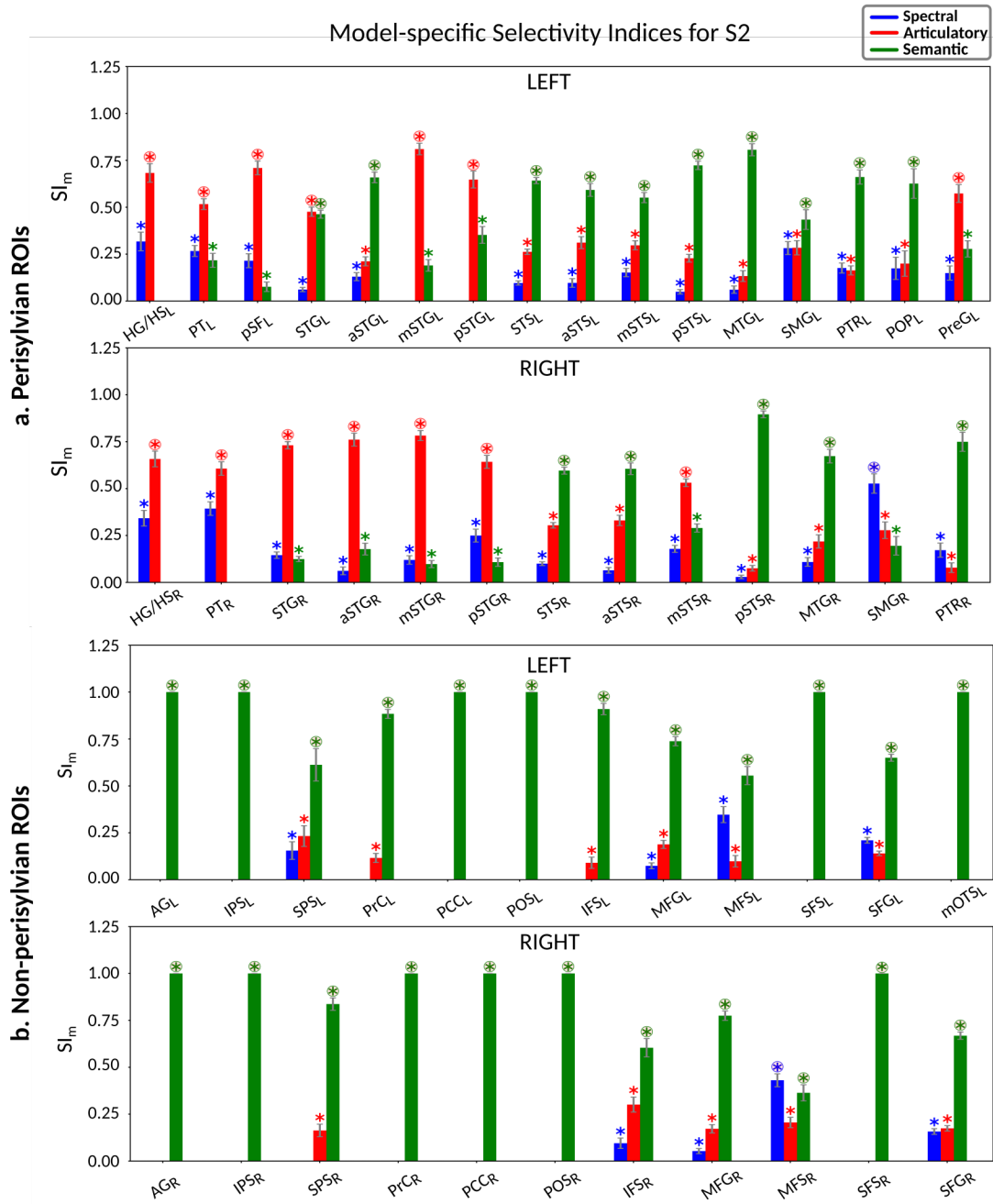


Figure S3b: Model-specific selectivity indices for subject S2. Single-voxel prediction scores on passive-listening data were used to quantify the degree of selectivity of each ROI to the underlying model features under passive listening. Prediction scores for a given model were averaged across speech-selective voxels within each ROI, and then normalized such that the cumulative prediction score from all models summed to 1. The resultant measure was taken as a model-specific selectivity index, (SI_m). SI_m is in the range of [0, 1], where higher values indicate stronger selectivity for the underlying model. Bar plots display SI_m for spectral, articulatory and semantic models (mean \pm sem across ROI's speech-selective voxels). Significant indices are marked with *, colored according to the model and encircled if they are dominant within the ROI ($p < 0.05$, bootstrap test). **a. Perisylvian ROIs.** ROIs in perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively. **b. Non-perisylvian ROIs.** ROIs in non-perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively.

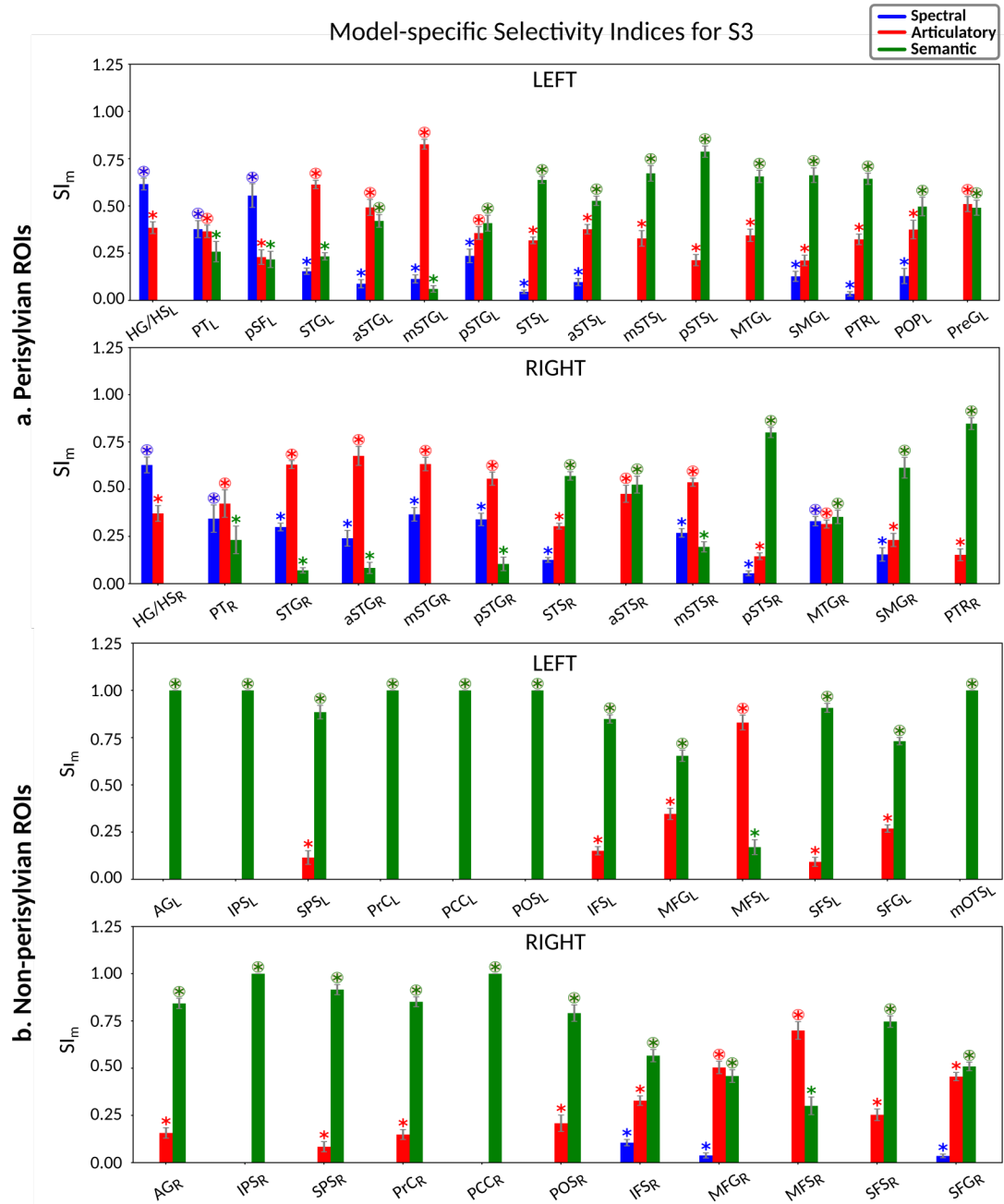


Figure S3c: Model-specific selectivity indices for subject S3. Single-voxel prediction scores on passive-listening data were used to quantify the degree of selectivity of each ROI to the underlying model features under passive listening. Prediction scores for a given model were averaged across speech-selective voxels within each ROI, and then normalized such that the cumulative prediction score from all models summed to 1. The resultant measure was taken as a model-specific selectivity index, (SI_m). SI_m is in the range of [0, 1], where higher values indicate stronger selectivity for the underlying model. Bar plots display SI_m for spectral, articulatory and semantic models (mean \pm sem across ROI's speech-selective voxels). Significant indices are marked with *, colored according to the model and encircled if they are dominant within the ROI ($p < 0.05$, bootstrap test). **a. Perisylvian ROIs.** ROIs in perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively. **b. Non-perisylvian ROIs.** ROIs in non-perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively.

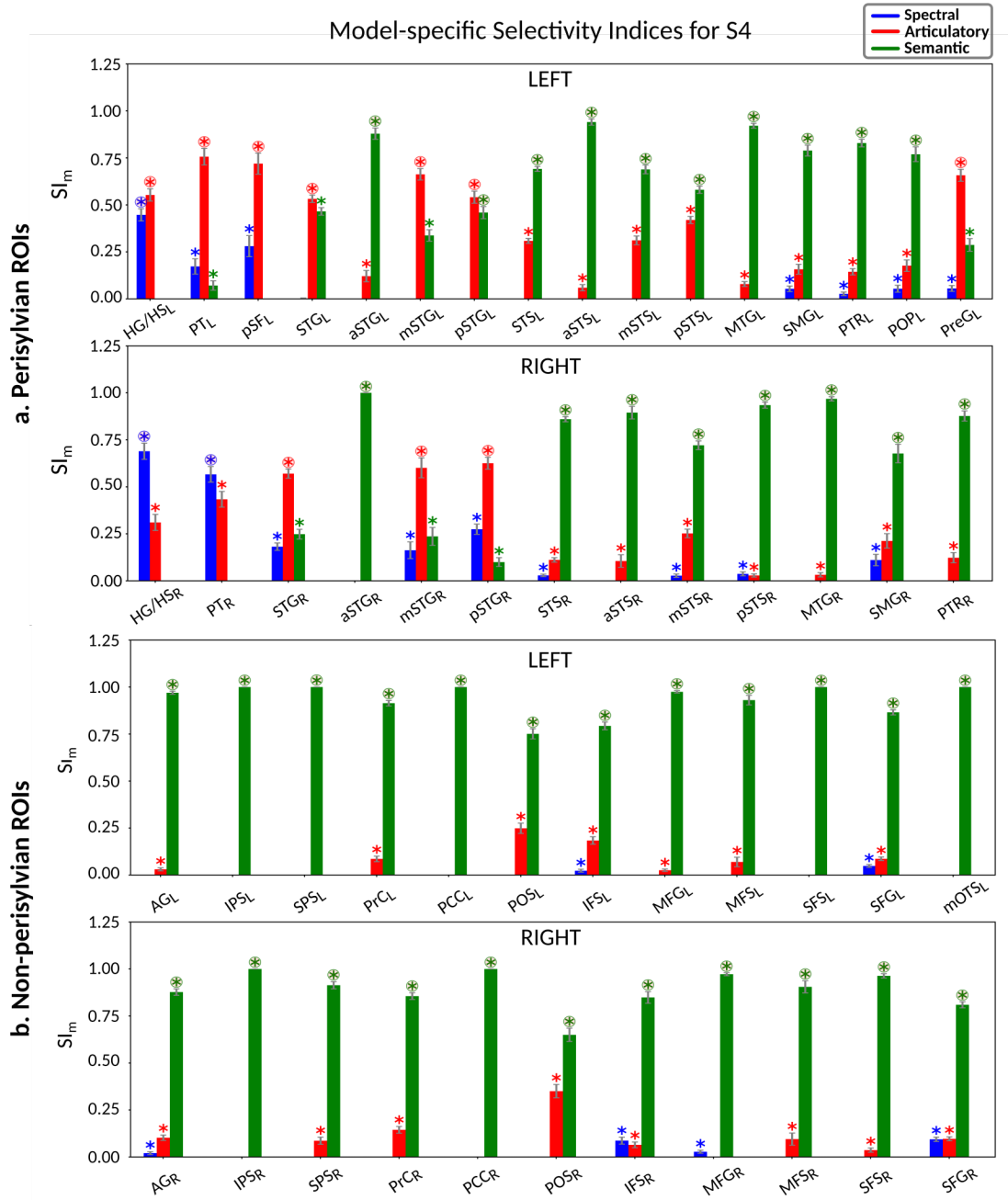


Figure S3d: Model-specific selectivity indices for subject S4. Single-voxel prediction scores on passive-listening data were used to quantify the degree of selectivity of each ROI to the underlying model features under passive listening. Prediction scores for a given model were averaged across speech-selective voxels within each ROI, and then normalized such that the cumulative prediction score from all models summed to 1. The resultant measure was taken as a model-specific selectivity index, (SI_m). SI_m is in the range of [0, 1], where higher values indicate stronger selectivity for the underlying model. Bar plots display SI_m for spectral, articulatory and semantic models (mean \pm sem across ROI's speech-selective voxels). Significant indices are marked with *, colored according to the model and encircled if they are dominant within the ROI ($p < 0.05$, bootstrap test). **a. Perisylvian ROIs.** ROIs in perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively. **b. Non-perisylvian ROIs.** ROIs in non-perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively.

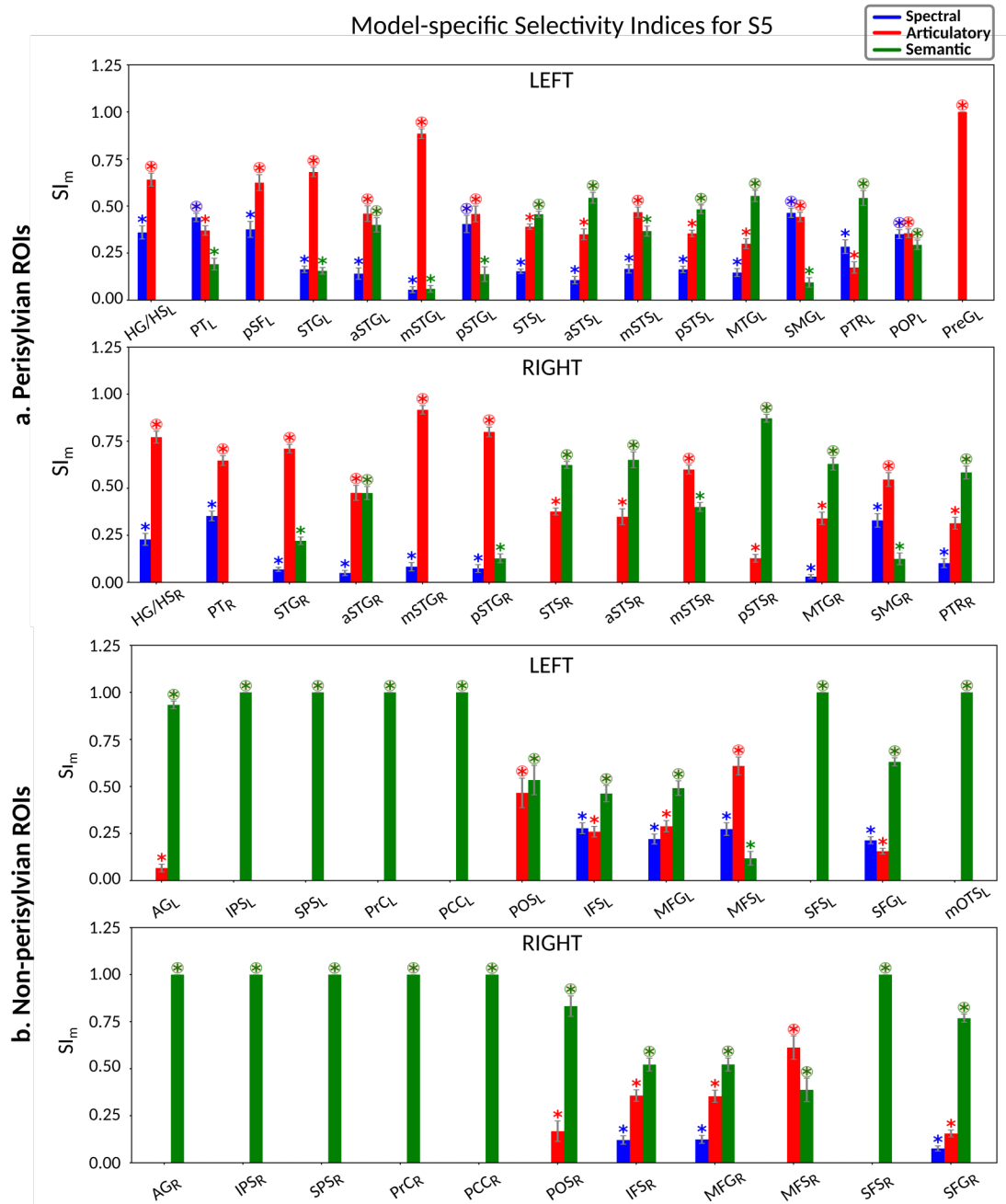


Figure S3e: Model-specific selectivity indices for subject S5. Single-voxel prediction scores on passive-listening data were used to quantify the degree of selectivity of each ROI to the underlying model features under passive listening. Prediction scores for a given model were averaged across speech-selective voxels within each ROI, and then normalized such that the cumulative prediction score from all models summed to 1. The resultant measure was taken as a model-specific selectivity index, (SI_m). SI_m is in the range of [0, 1], where higher values indicate stronger selectivity for the underlying model. Bar plots display SI_m for spectral, articulatory and semantic models (mean \pm sem across ROI's speech-selective voxels). Significant indices are marked with *, colored according to the model and encircled if they are dominant within the ROI ($p < 0.05$, bootstrap test). **a. Perisylvian ROIs.** ROIs in perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively. **b. Non-perisylvian ROIs.** ROIs in non-perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively.

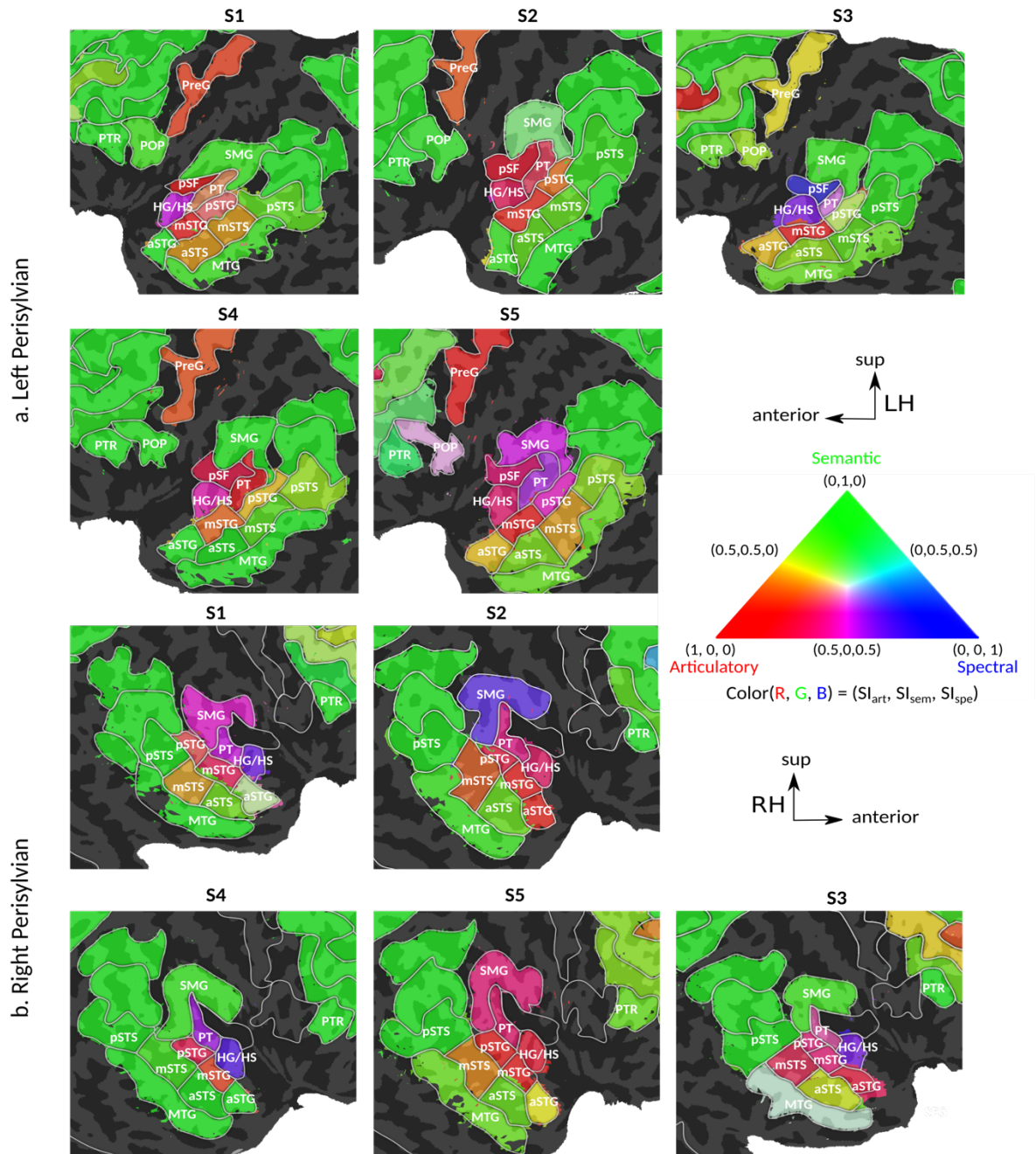


Figure S4: Intrinsic selectivity profiles. Selectivity profiles of perisylvian ROIs in subjects S1-S5 are shown on the cortical flatmaps. Significant articulatory, semantic, and spectral selectivity indices are projected onto the red, green, and blue channels of the RGB colormap (see Methods). **a.** *Left perisylvian ROIs.* **b.** *Right perisylvian ROIs.* Consistently across subjects, A progression from low- and intermediate-level to high-level speech representations are apparent across bilateral temporal cortex in superior-inferior direction (HG/HS -> mSTG -> mSTS -> MTG). These results support the view that speech representations are hierarchically organized across processing hierarchy of speech with partial overlap between spectral, articulatory and semantic representations in early to intermediate stages of auditory processing.

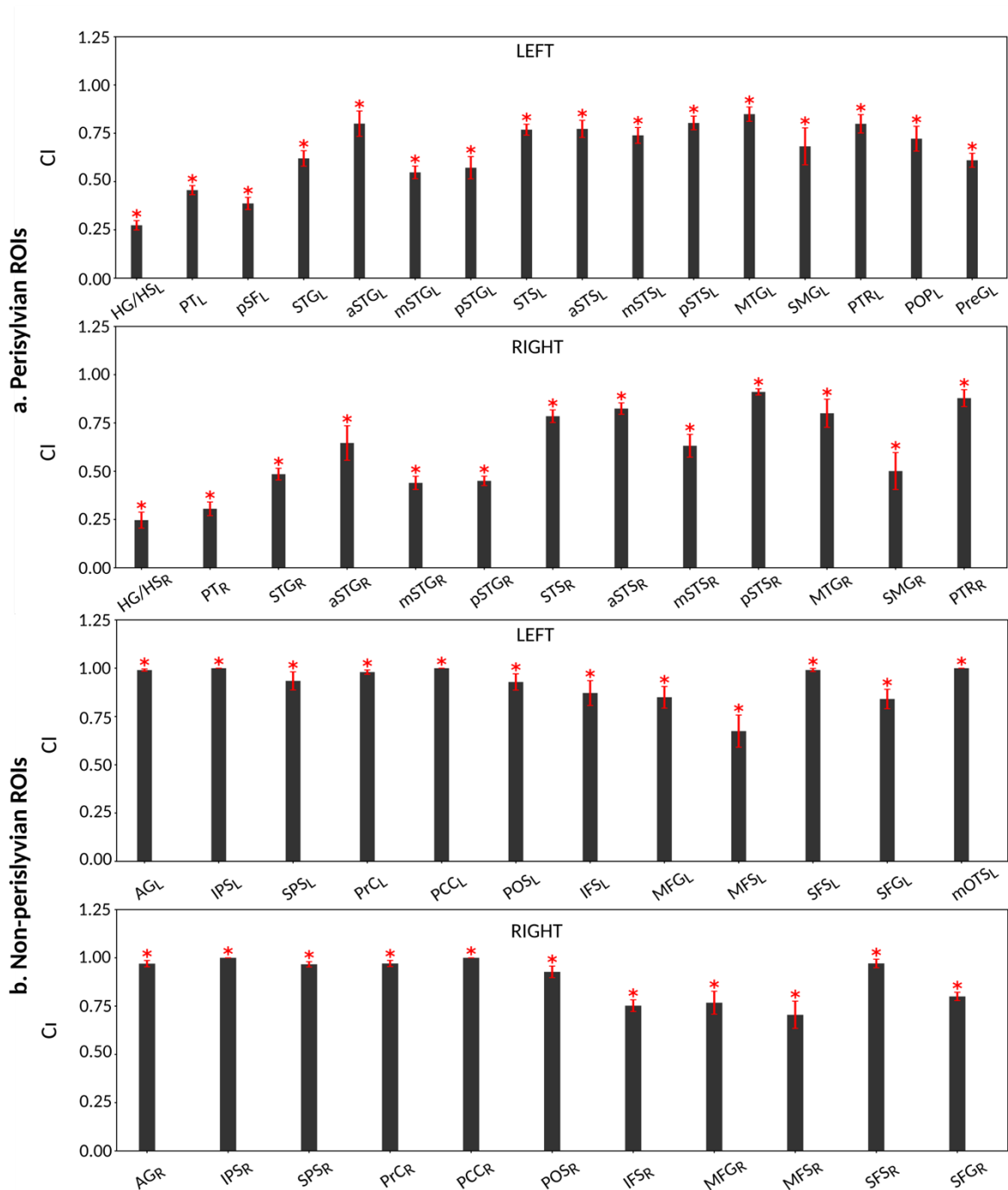


Figure S5: Representational complexity. To characterize the complexity of speech representations, a complexity index, (CI), was taken as the relative tuning of an ROI for low- versus high-level speech features. CI is in the range of $[0, 1]$, where higher values indicate stronger tuning for semantic features and lower values indicate stronger tuning for spectral features. Bar plots indicate CI (mean \pm sem across subjects). Indices that are consistently significant in each individual subject are marked with * ($p < 10^{-4}$, bootstrap test). **a. Perisylvian ROIs.** ROIs in perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively. Early auditory regions such as HG/HS have CI close to 0, whereas higher-order regions like PTR have CI tending to 1. These results indicate that more complex features are represented in higher-order regions compared to the earlier ones. **b. Non-perisylvian ROIs.** ROIs in non-perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively. Higher-order regions in non-perisylvian cortex mostly have similar representational complexity that is close to 1.

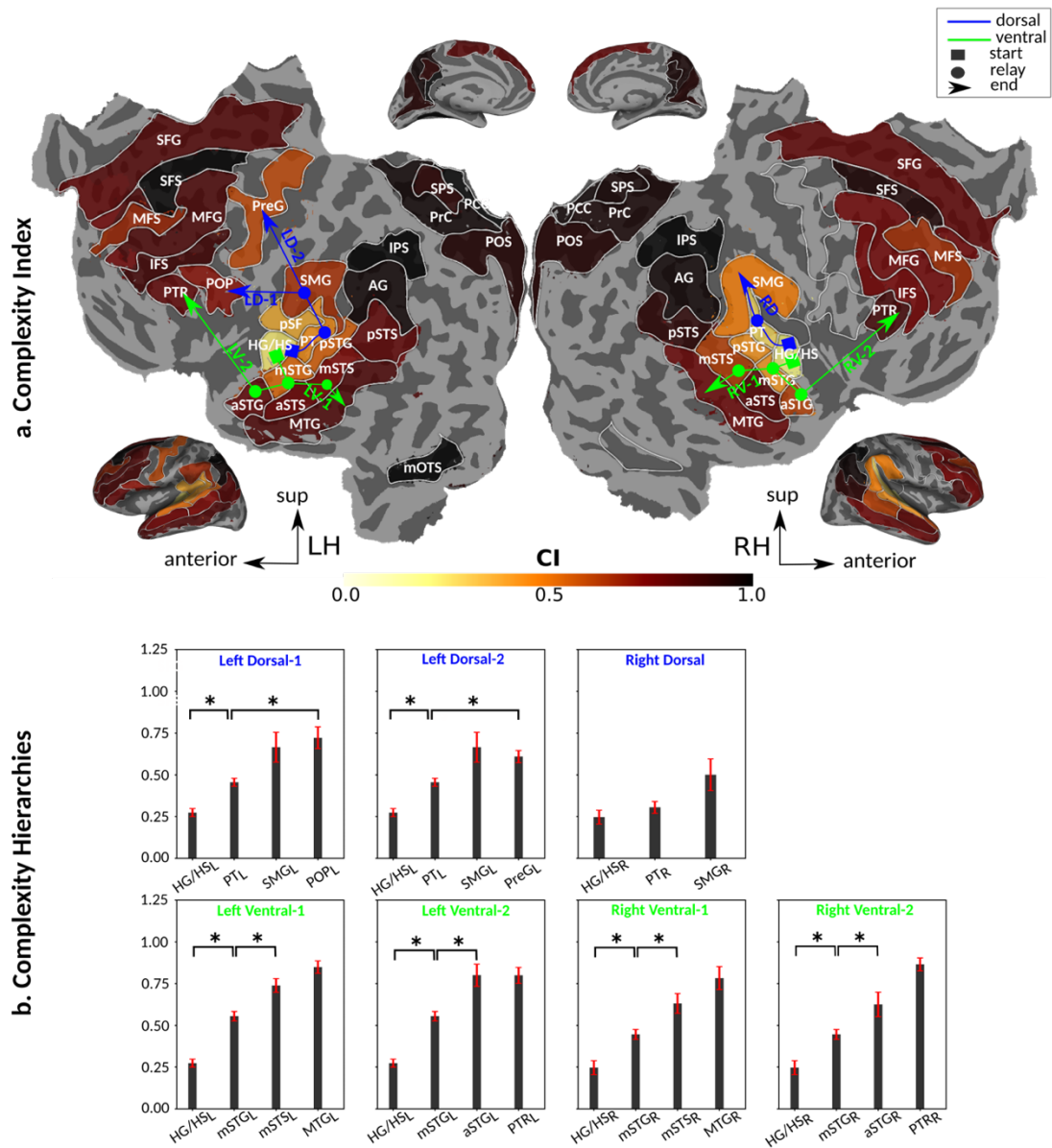


Figure S6: Cortical hierarchy of representational complexity. **a.** Complexity index and auditory pathways. To visualize how representational complexity distributes across cortex, CI of cortical ROIs averaged across subjects are displayed on the flatmap of a representative subject (S4; see legend). To illustrate the gradients in CI across the hierarchy of speech processing, dorsal and ventral pathways are shown with blue and green lines, respectively: left dorsal-1 (LD-1), left dorsal-2 (LD-2) and right dorsal (RD), left ventral-1 (LV-1), left ventral-2 (LV-2), right ventral-1 (RV-1) and right ventral-2 (RV-2). Squares mark regions where pathways begin; arrows mark regions where pathways end; and circles mark relay regions in between. **b.** Complexity hierarchies. Bar plots display CI (mean \pm sem across subjects) along LD-1, LD-2, RD, LV-1, LV-2, RV-1 and RV-2, shown in separate panels. Significant differences in CI between consecutive ROIs are marked with brackets ($p < 0.05$, bootstrap test). Significant gradients in CI are: $CI_{HG/HS} < CI_{PT} < CI_{POP}$ in LD-1, $CI_{HG/HS} < CI_{PT} < CI_{PreG}$ in LD-2, $CI_{HG/HS} < CI_{mSTG} < CI_{mSTS}$ in LV-1 and RV-1, and $CI_{HG/HS} < CI_{mSTG} < CI_{aSTG}$ in LV-2 and RV-2. These results suggest that hierarchical speech representations are systematically organized in multiple gradients across cortex.

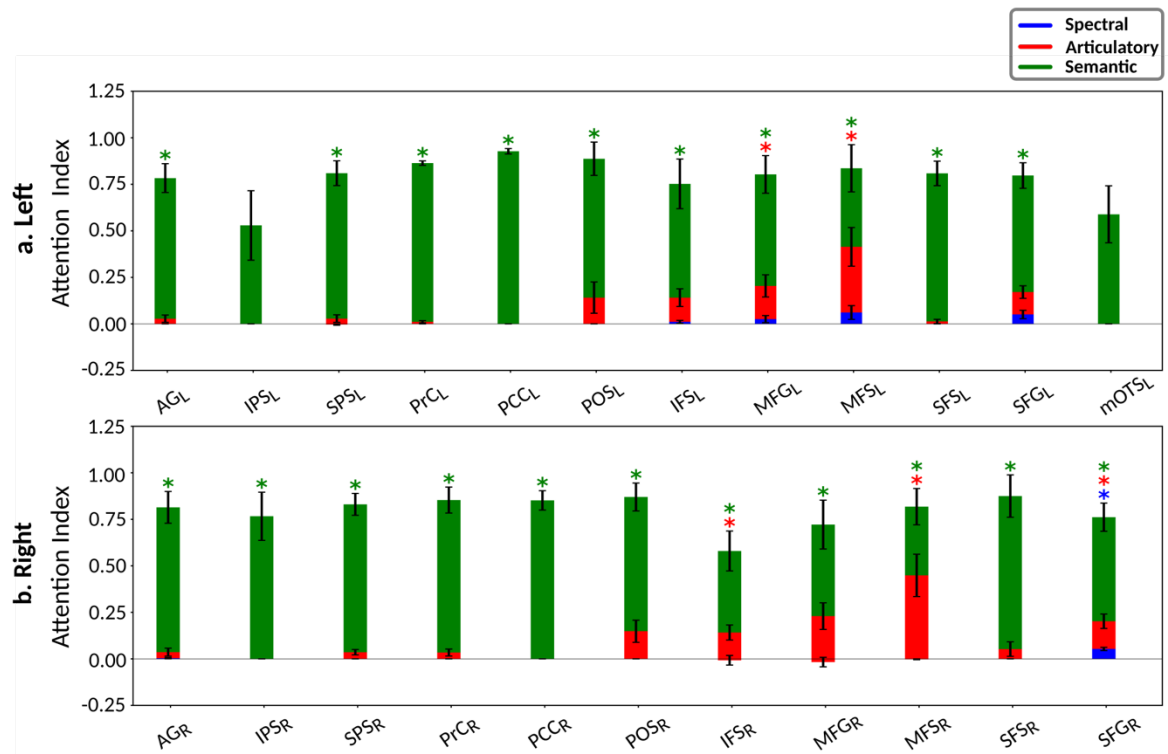


Figure S7: Model-specific attention indices for non-perisylvian ROIs. A model-specific attention index (AI_m) was computed based on the difference in model prediction scores when the stories were attended versus unattended (see Methods). AI_m is in the range of $[-1,1]$, where a positive index indicates modulation in favor of the attended stimulus and a negative index indicates modulation in favor of the unattended stimulus. For each ROI in non-perisylvian cortex, spectral, articulatory, and semantic attention indices are given (mean \pm sem across subjects), and their sum yields the overall modulation. Significant indices are marked with * ($p < 0.05$, bootstrap test; see Supp. Fig. S8a-e for attention indices of individual subjects). **a.** ROIs in LH. Modulations in left IPS and mOTS are not consistently significant in all subjects ($p > 0.05$). **b.** ROIs in RH. mOTSR that did not have consistent speech selectivity in individual subjects was excluded (see Methods). These results indicate that selectivity modulations manifest primarily at the semantic level in most of the non-perisylvian ROIs.

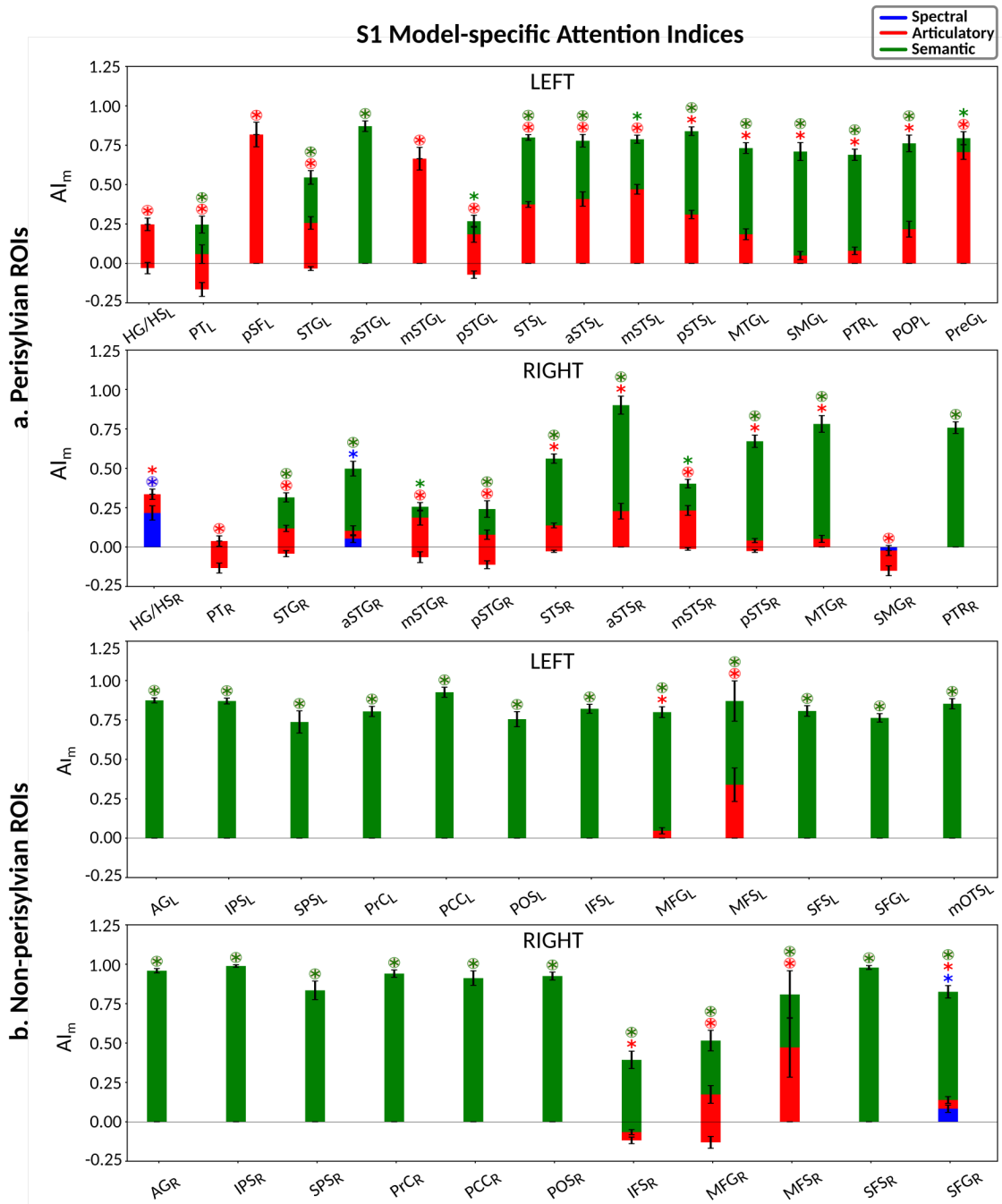


Figure S8a: Model-specific attention indices for subject S1. A model-specific attention index (AI_m) was computed based on the difference in model prediction scores when the stories were attended versus unattended (see Methods). AI_m is in the range of $[-1, 1]$, where a positive index indicates modulation in favor of the attended stimulus and a negative index indicates modulation in favor of the unattended stimulus. For each ROI, spectral, articulatory, and semantic attention indices are given (mean \pm sem across speech-selective voxels in each ROI), and their sum yields the overall modulation. Significantly positive modulations are marked with *, colored according to the corresponding model and encircled if they are dominant within the ROI ($p < 0.05$, bootstrap test). **a. Perisylvian ROIs.** ROIs in perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively. **b. Non-perisylvian ROIs.** ROIs in non-perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively.

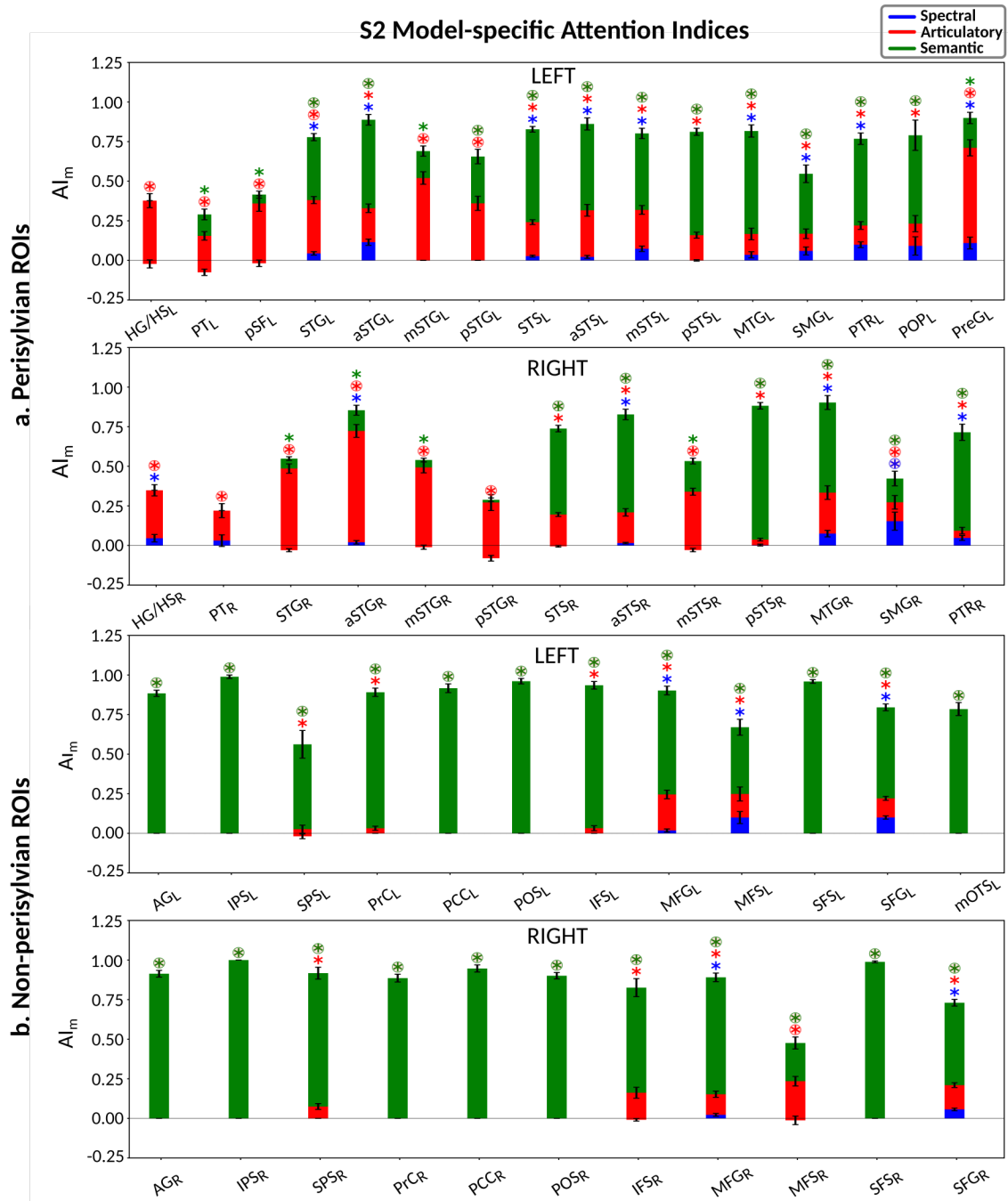


Figure S8b: Model-specific attention indices for subject S2. A model-specific attention index (AI_m) was computed based on the difference in model prediction scores when the stories were attended versus unattended (see Methods). AI_m is in the range of $[-1, 1]$, where a positive index indicates modulation in favor of the attended stimulus and a negative index indicates modulation in favor of the unattended stimulus. For each ROI, spectral, articulatory, and semantic attention indices are given (mean \pm sem across speech-selective voxels in each ROI), and their sum yields the overall modulation. Significantly positive modulations are marked with *, colored according to the corresponding model and encircled if they are dominant within the ROI ($p < 0.05$, bootstrap test). **a. Perisylvian ROIs.** ROIs in perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively. **b. Non-perisylvian ROIs.** ROIs in non-perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively.

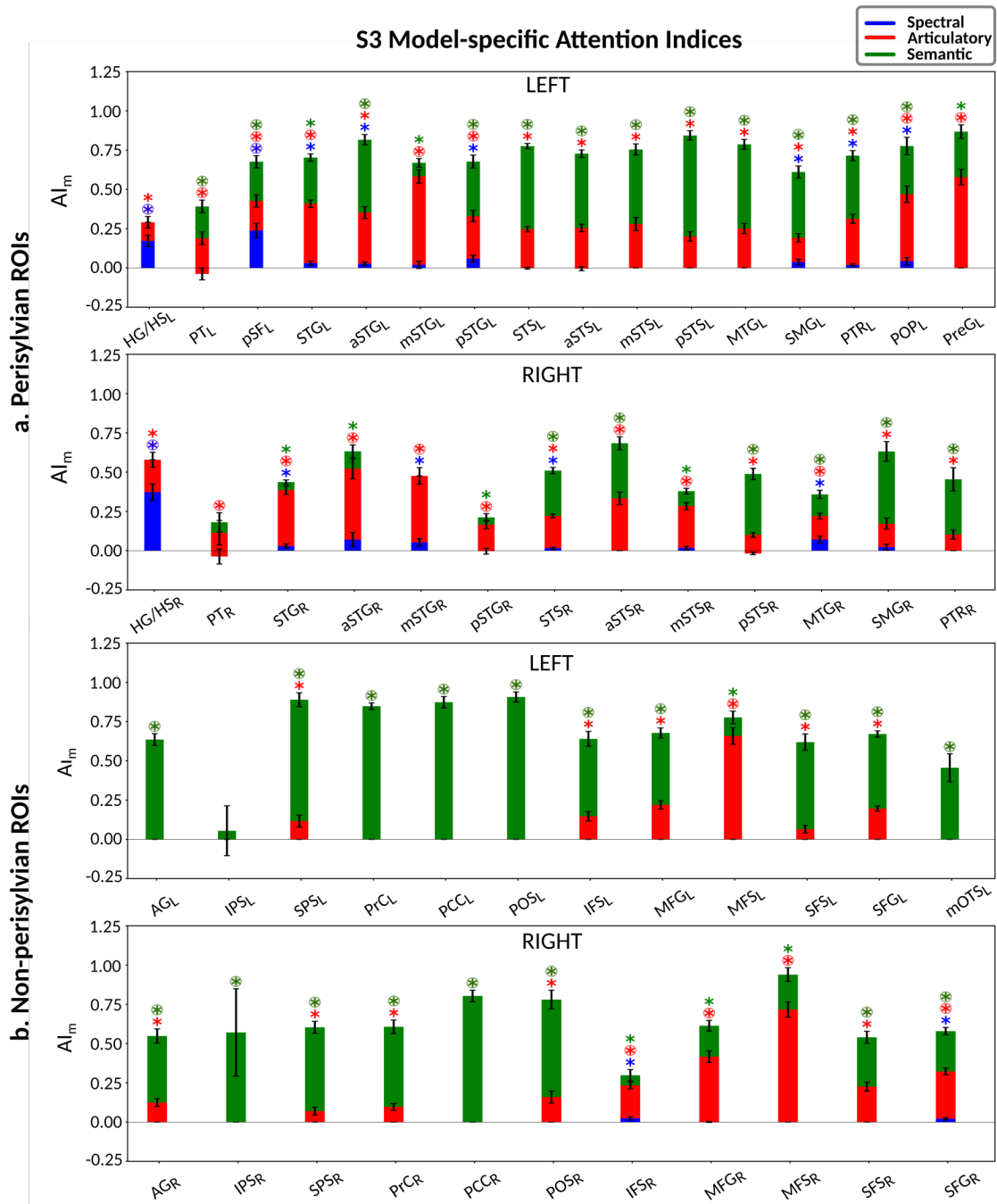


Figure S8c: Model-specific attention indices for subject S3. A model-specific attention index (AI_m) was computed based on the difference in model prediction scores when the stories were attended versus unattended (see Methods). AI_m is in the range of $[-1, 1]$, where a positive index indicates modulation in favor of the attended stimulus and a negative index indicates modulation in favor of the unattended stimulus. For each ROI, spectral, articulatory, and semantic attention indices are given (mean \pm sem across speech-selective voxels in each ROI), and their sum yields the overall modulation. Significantly positive modulations are marked with *, colored according to the corresponding model and encircled if they are dominant within the ROI ($p < 0.05$, bootstrap test). **a. Perisylvian ROIs.** ROIs in perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively. **B. Non-perisylvian ROIs.** ROIs in non-perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively.

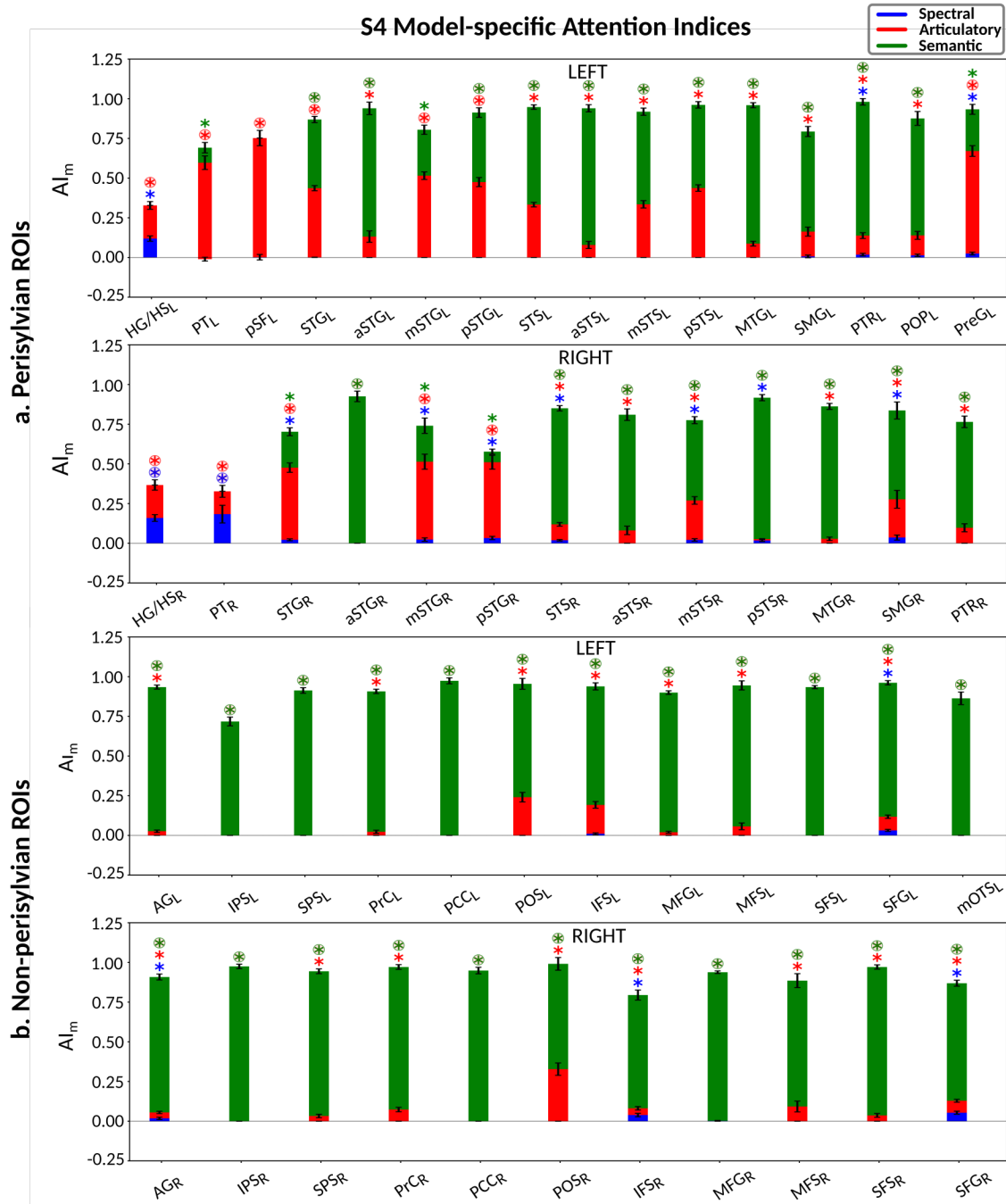


Figure S8d: Model-specific attention indices for subject S4. A model-specific attention index (AI_m) was computed based on the difference in model prediction scores when the stories were attended versus unattended (see Methods). AI_m is in the range of $[-1,1]$, where a positive index indicates modulation in favor of the attended stimulus and a negative index indicates modulation in favor of the unattended stimulus. For each ROI, spectral, articulatory, and semantic attention indices are given (mean \pm sem across speech-selective voxels in each ROI), and their sum yields the overall modulation. Significantly positive modulations are marked with *, colored according to the corresponding model and encircled if they are dominant within the ROI ($p < 0.05$, bootstrap test). **a. Perisylvian ROIs.** ROIs in perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively. **B. Non-perisylvian ROIs.** ROIs in non-perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively.

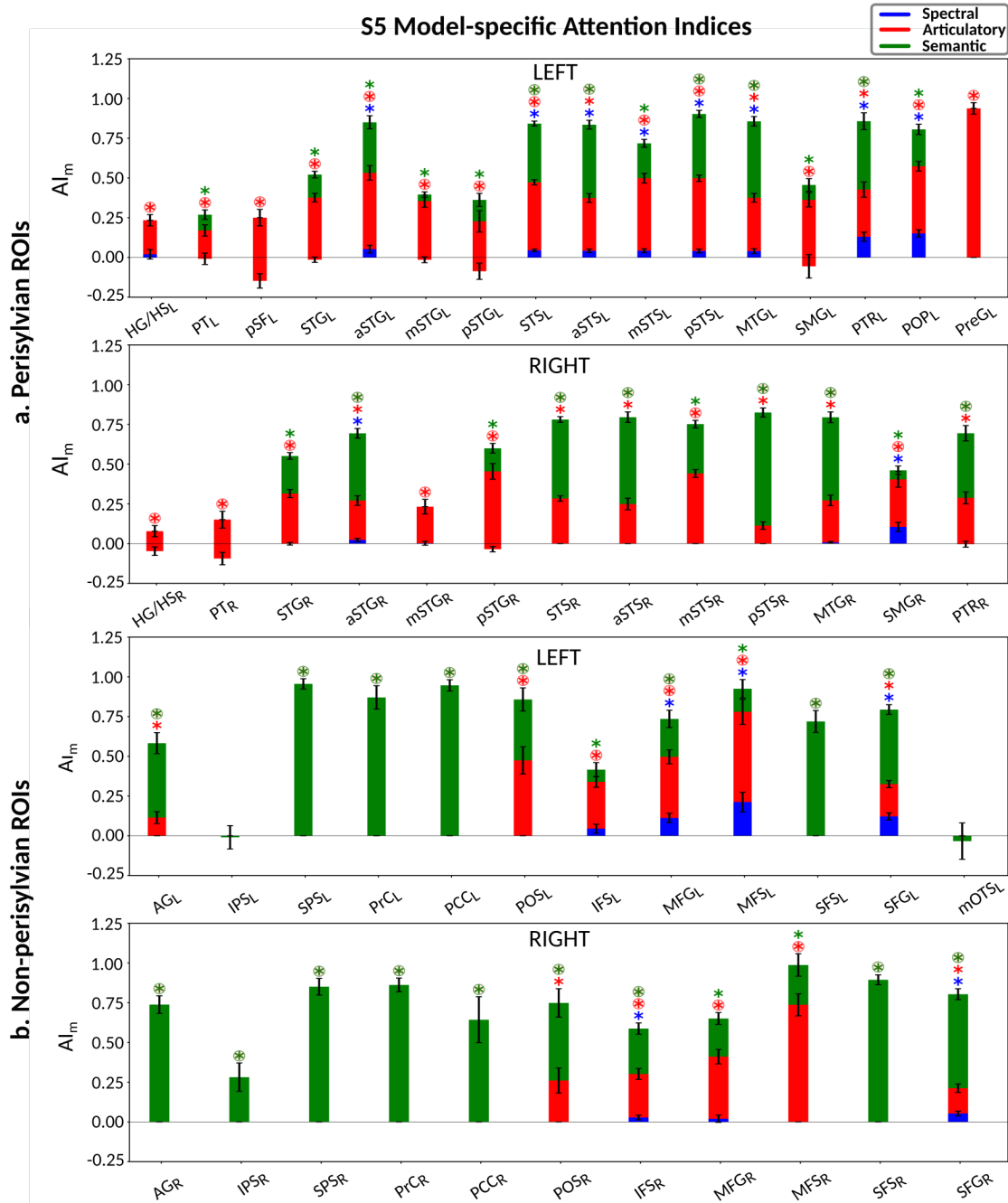


Figure S8e: Model-specific attention indices for subject S5. A model-specific attention index (AI_m) was computed based on the difference in model prediction scores when the stories were attended versus unattended (see Methods). AI_m is in the range of $[-1, 1]$, where a positive index indicates modulation in favor of the attended stimulus and a negative index indicates modulation in favor of the unattended stimulus. For each ROI, spectral, articulatory, and semantic attention indices are given (mean \pm sem across speech-selective voxels in each ROI), and their sum yields the overall modulation. Significantly positive modulations are marked with *, colored according to the corresponding model and circled if they are dominant within the ROI ($p < 0.05$, bootstrap test). **a. Perisylvian ROIs.** ROIs in perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively. **b. Non-perisylvian ROIs.** ROIs in non-perisylvian cortex are displayed. ROIs in the LH and RH are shown in top and bottom panels, respectively.

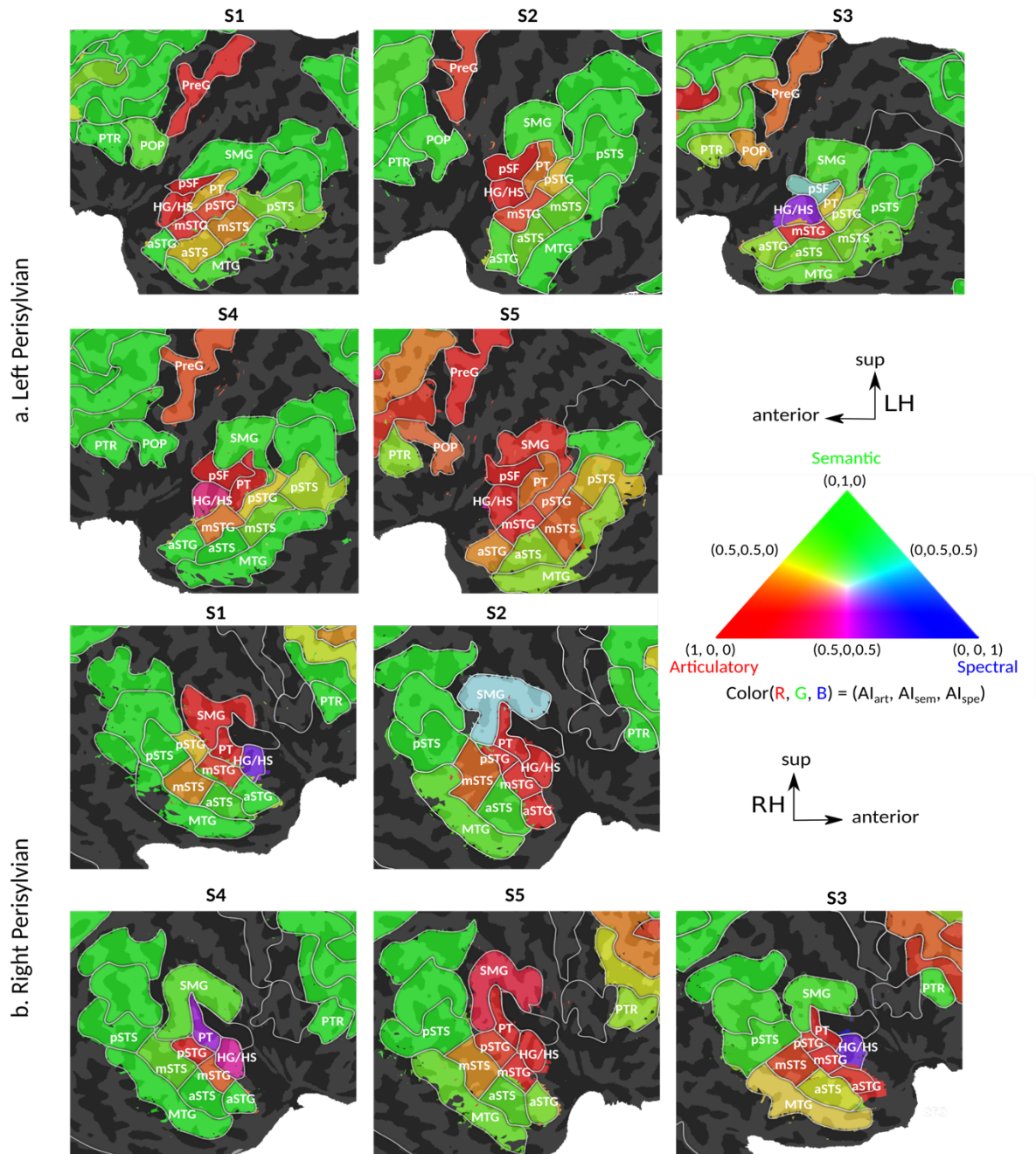


Figure S9: Attentional modulation profiles for individual subjects. Modulation profiles of perisylvian ROIs in subjects S1-5 are shown on the cortical flatmaps. Significantly positive articulatory, semantic, and spectral attention indices were projected to the red, green, and blue channels of the RGB colormap (see Methods). **a.** *Left perisylvian ROIs.* **b.** *Right perisylvian ROIs.* In all subjects, a progression in the level of speech representations dominantly modulated is apparent across bilateral temporal cortex in the superior-inferior direction (HG/HS → mSTG → mSTS → MTG). Semantic modulation is dominant at both ends (MTG and PTR) of bilateral ventral stream, while articulatory modulation is dominant at one end (PreG) of left dorsal stream ($p < 0.05$, bootstrap test; see also Supp. Fig. S8a-e). There is no consistent tendency towards dominant articulatory or semantic modulation at the other end of the left dorsal stream (POP).

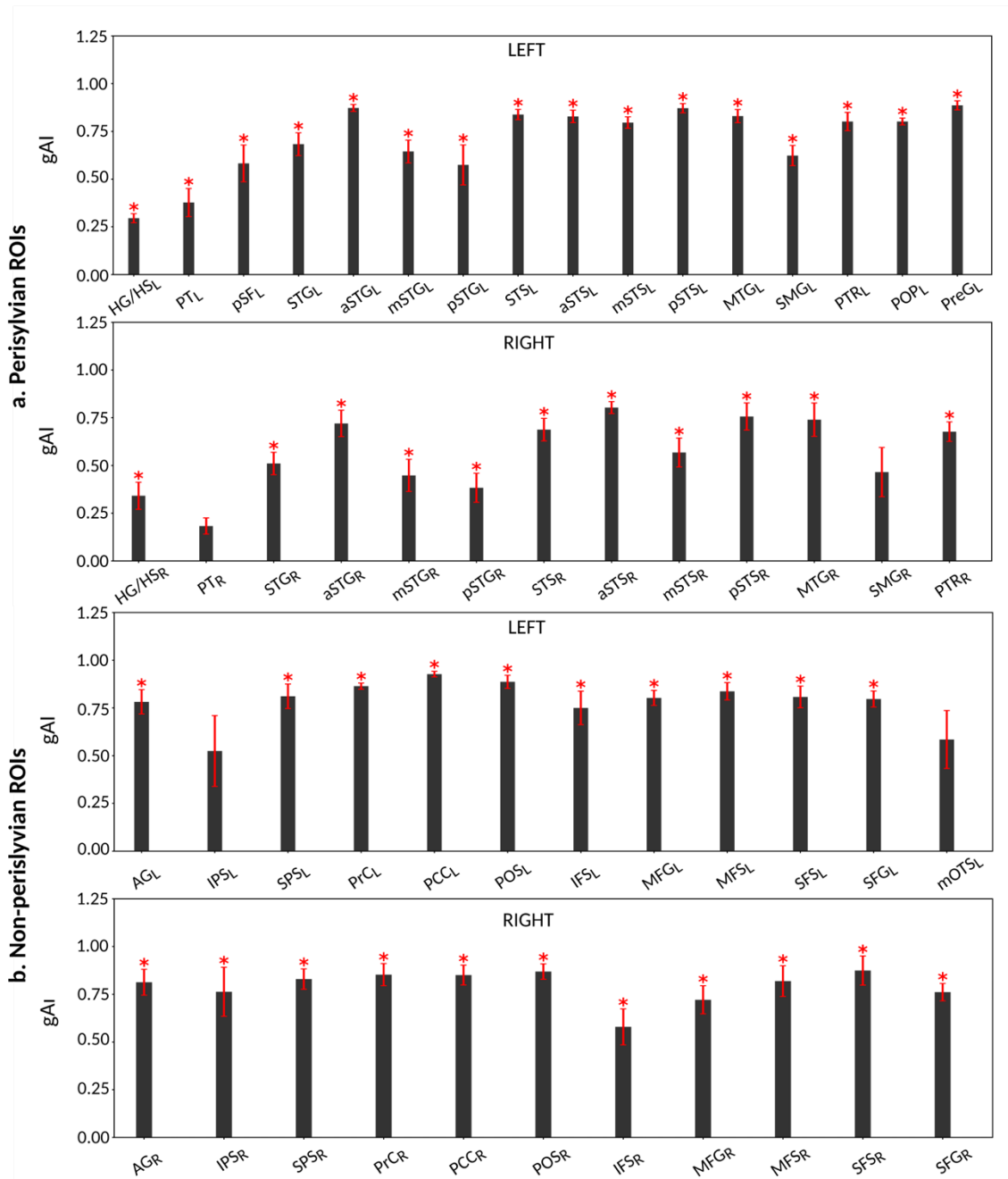


Figure S10: Global attention indices. To quantify overall modulatory effects on selectivity across all examined feature levels, global attentional modulation (gAI) was computed by summing spectral, articulatory, and semantic attention indices (see Methods). gAI is in the range of $[-1,1]$, where a positive index indicates attentional modulation in favor of the attended stimuli and a negative index indicates modulation in favor of the unattended stimuli. A value of zero indicates no modulation. Bar plots indicate gAI (mean \pm sem across subjects). Significant indices are marked with * ($p < 0.05$, bootstrap test; see the bar plots in Supp. Fig. 8a-e for gAI in individual subjects). **a. Perisylvian ROIs.** ROIs within perisylvian cortex in LH and RH are displayed in top and bottom panels, respectively. gAI is significant in all perisylvian ROIs consistently in all subjects except in right PT and SMG ($p > 0.05$). **b. Non-perisylvian ROIs.** ROIs in non-perisylvian cortex are displayed. gAI is significant in all non-perisylvian ROIs consistently in all subjects except in left IPS and mOTS ($p > 0.05$).

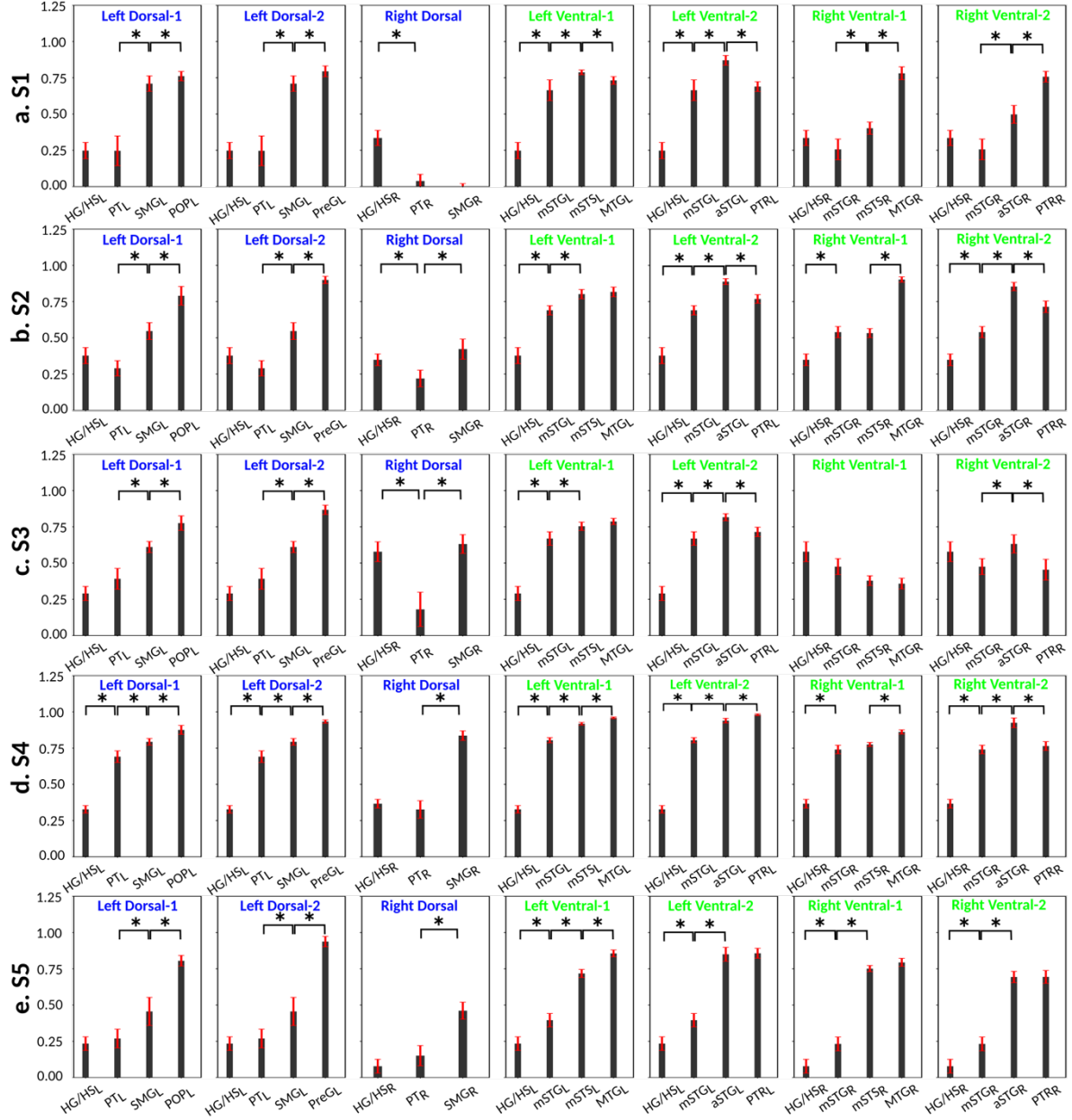


Figure S11: Modulation hierarchies for individual subjects. **a.** Modulation hierarchies for subject S1. Gradients in gAI across LD-1, LD-2, RD, LV-1, LV-2, RV-1 and RV-2 are shown in subfigures labeled accordingly. Bar plots display gAI in each ROI (mean \pm sem across speech-selective voxels in each ROI). Only ROIs within a given trajectory are included in the corresponding subfigure. Significant differences in gAI between consecutive ROIs along the trajectory are marked with brackets ($p < 0.05$, bootstrap test). **b.** Modulation hierarchies for subject S2. **c.** Modulation hierarchies for subject S3. **d.** Modulation hierarchies for subject S4. **e.** Modulation hierarchies for subject S5. Gradients consistently obtained in each individual subject are ($p < 0.05$): $gAI_{PT} < gAI_{SMG} < gAI_{POP}$ in LD-1, $gAI_{PT} < gAI_{SMG} < gAI_{PreG}$ in LD-2, $gAI_{HG} < gAI_{mSTG} < gAI_{mST5}$ in LV-1, $gAI_{HG} < gAI_{mSTG} < gAI_{aSTG}$ in LV-2, and $gAI_{mSTG} < gAI_{aSTG}$ in RV-2.

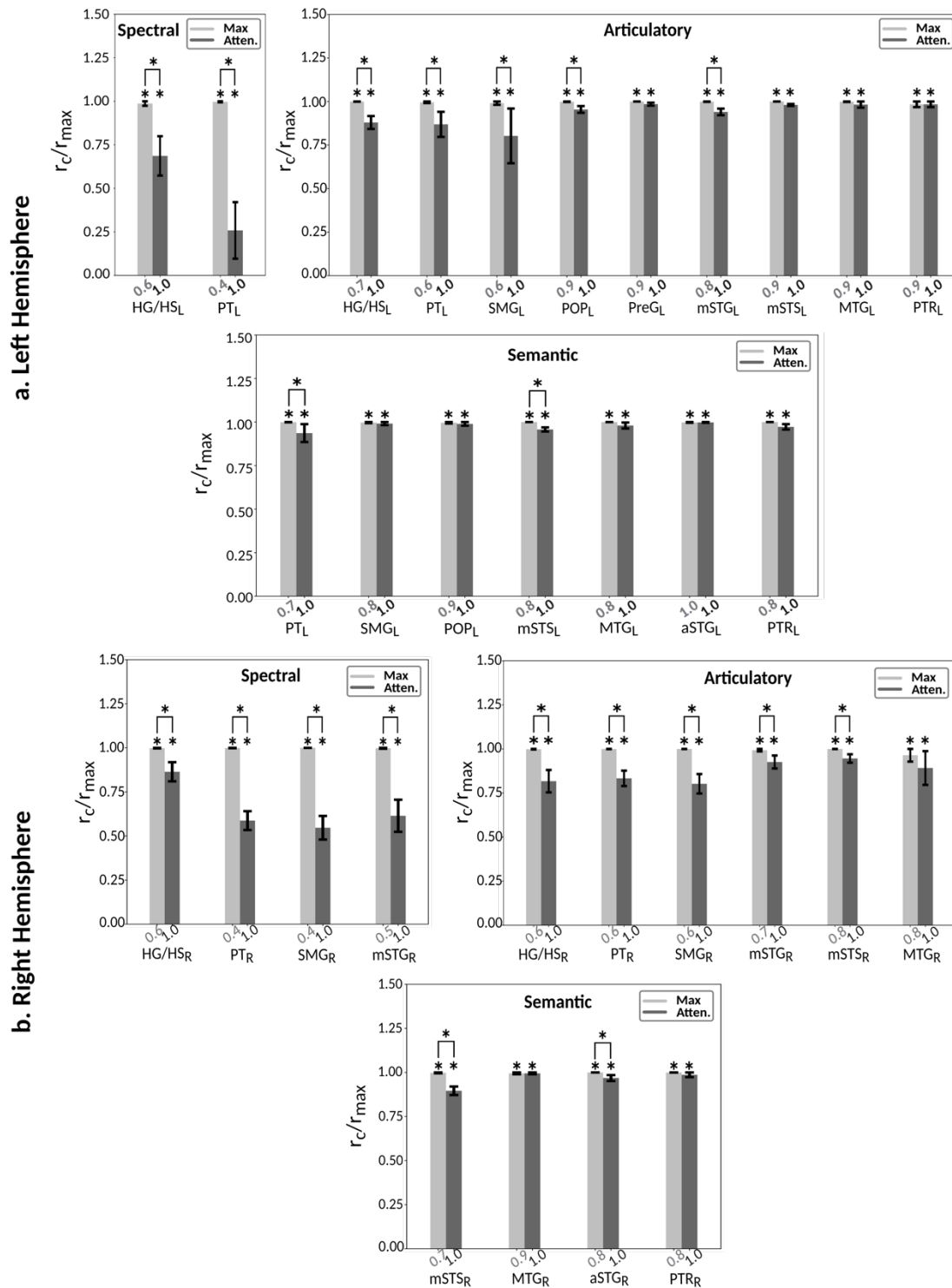


Figure S12a: Representation of unattended speech in subject S1. Passive-listening models were tested on cocktail-party data to assess representation of unattended speech during the cocktail-party task. Prediction scores were calculated separately for a combination model comprising features of both attended and unattended stories (r_{max} : optimal convex combination) and an individual model only comprising features of the attended story (r_a). Significant difference in prediction between the two models is an indication that BOLD responses carry significant information on unattended speech. Bar plots display normalized prediction scores (mean \pm sem across speech-selective voxels in each ROI; combination model in light gray and individual model in gray). Significant scores are marked with * ($p < 10^{-4}$, bootstrap test), and significant differences are marked with brackets ($p < 0.05$). Prediction scores are only displayed for ROIs in the dorsal and ventral streams, with significant selectivity for given model features. **a.** Representation of unattended speech in the left hemisphere. Spectral, articulatory and semantic representations are labeled in subplots accordingly. **b.** Representation of unattended speech in the right hemisphere. Spectral, articulatory and semantic representations are labeled in subplots accordingly.

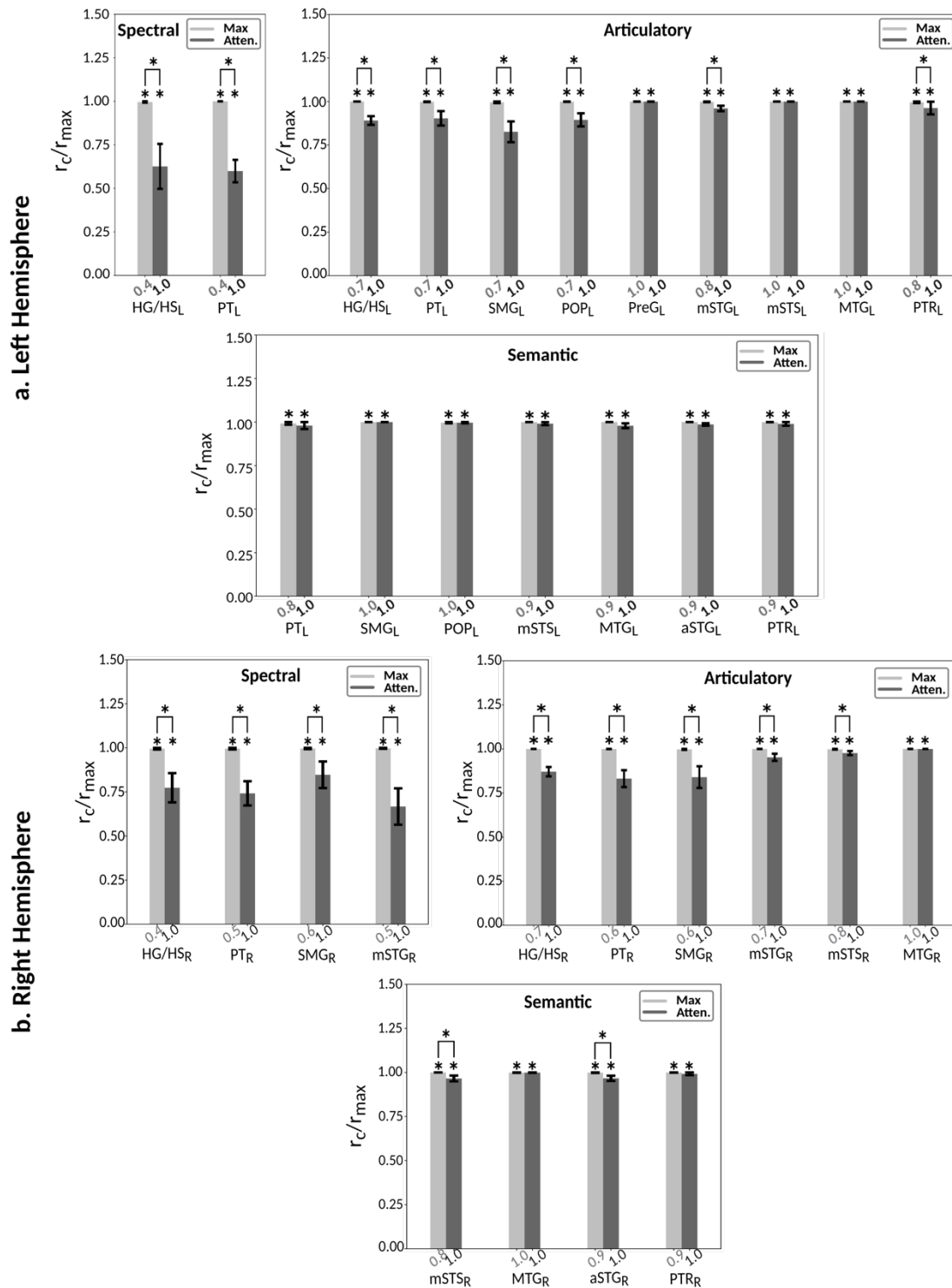


Figure S12b: Representation of unattended speech in subject S2. Passive-listening models were tested on cocktail-party data to assess representation of unattended speech during the cocktail-party task. Prediction scores were calculated separately for a combination model comprising features of both attended and unattended stories (r_{max} : optimal convex combination) and an individual model only comprising features of the attended story (r_a). Significant difference in prediction between the two models is an indication that BOLD responses carry significant information on unattended speech. Bar plots display normalized prediction scores (mean \pm sem across speech-selective voxels in each ROI; combination model in light gray and individual model in gray). Significant scores are marked with * ($p < 10^{-4}$, bootstrap test), and significant differences are marked with brackets ($p < 0.05$). Prediction scores are only displayed for ROIs in the dorsal and ventral streams, with significant selectivity for given model features. **a.** Representation of unattended speech in the left hemisphere. Spectral, articulatory and semantic representations are labeled in subplots accordingly. **b.** Representation of unattended speech in the right hemisphere. Spectral, articulatory and semantic representations are labeled in subplots accordingly.

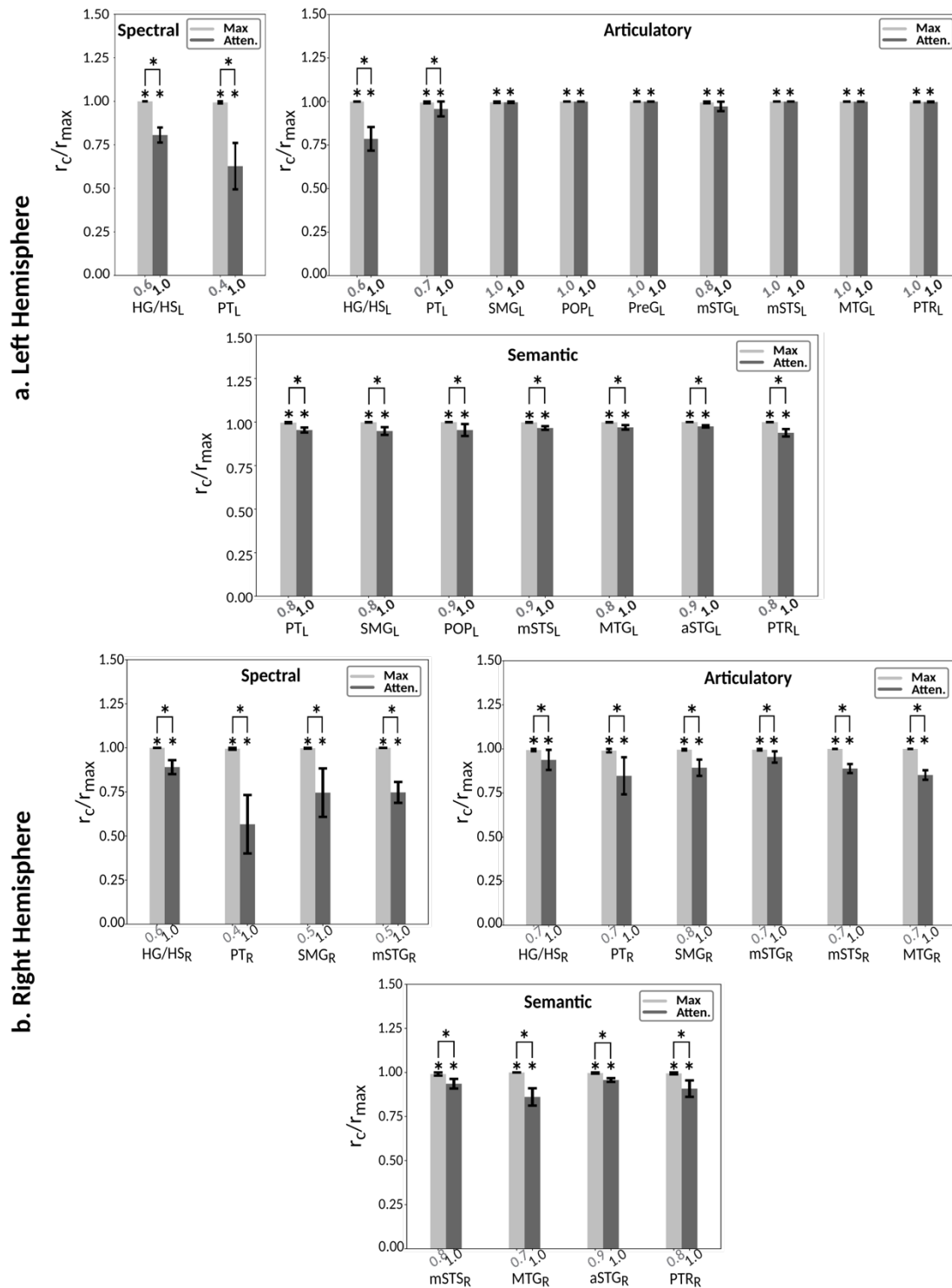


Figure S12c: Representation of unattended speech in subject S3. Passive-listening models were tested on cocktail-party data to assess representation of unattended speech during the cocktail-party task. Prediction scores were calculated separately for a combination model comprising features of both attended and unattended stories (r_{max} : optimal convex combination) and an individual model only comprising features of the attended story (r_a). Significant difference in prediction between the two models is an indication that BOLD responses carry significant information on unattended speech. Bar plots display normalized prediction scores (mean \pm sem across speech-selective voxels in each ROI; combination model in light gray and individual model in gray). Significant scores are marked with * ($p < 10^{-4}$, bootstrap test), and significant differences are marked with brackets ($p < 0.05$). Prediction scores are only displayed for ROIs in the dorsal and ventral streams, with significant selectivity for given model features. **a.** Representation of unattended speech in the left hemisphere. Spectral, articulatory and semantic representations are labeled in subplots accordingly. **b.** Representation of unattended speech in the right hemisphere. Spectral, articulatory and semantic representations are labeled in subplots accordingly.

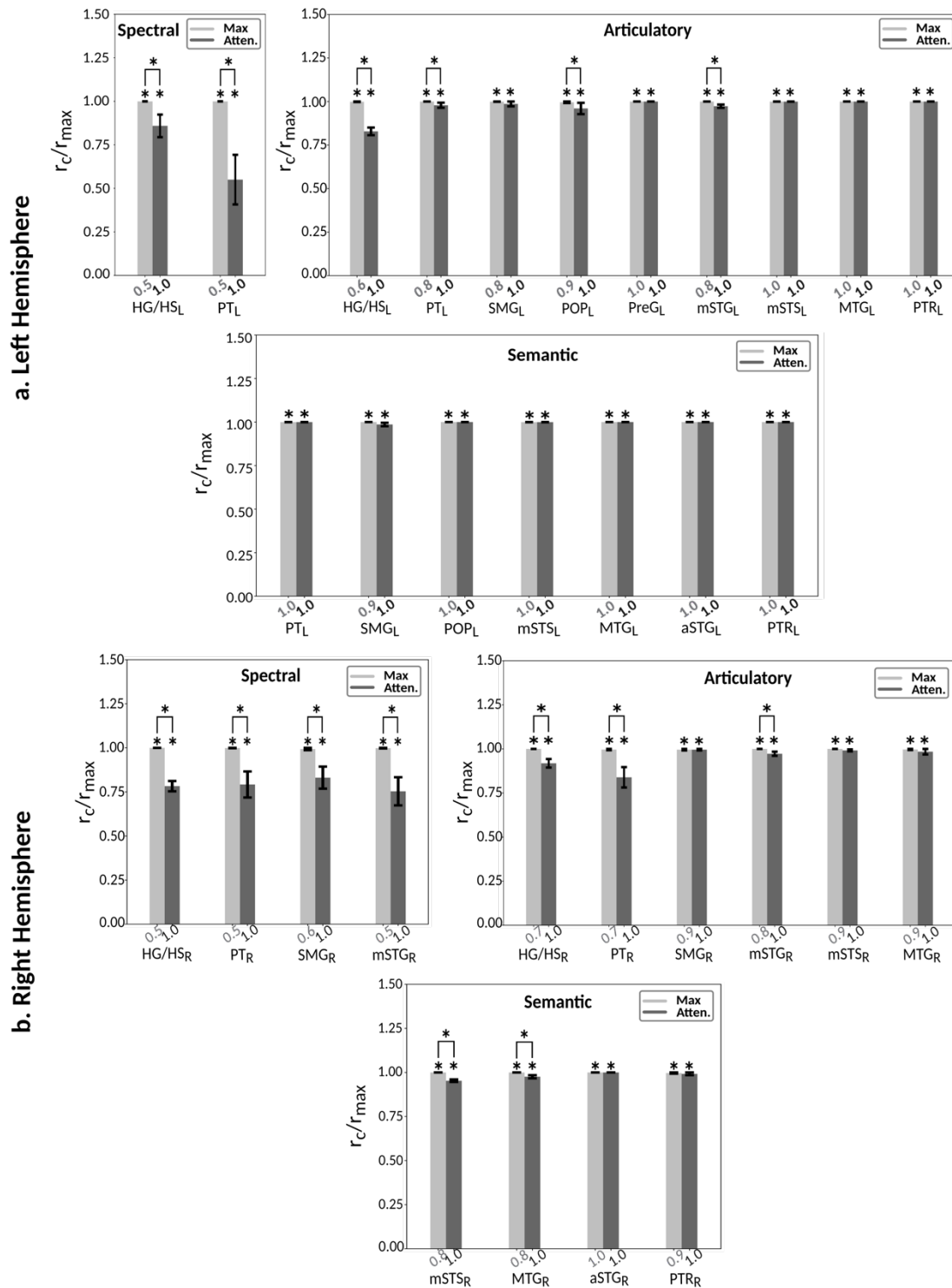


Figure S12d: Representation of unattended speech in subject S4. Passive-listening models were tested on cocktail-party data to assess representation of unattended speech during the cocktail-party task. Prediction scores were calculated separately for a combination model comprising features of both attended and unattended stories (r_{max} : optimal convex combination) and an individual model only comprising features of the attended story (r_a). Significant difference in prediction between the two models is an indication that BOLD responses carry significant information on unattended speech. Bar plots display normalized prediction scores (mean \pm sem across speech-selective voxels in each ROI; combination model in light gray and individual model in gray). Significant scores are marked with * ($p < 10^{-4}$, bootstrap test), and significant differences are marked with brackets ($p < 0.05$). Prediction scores are only displayed for ROIs in the dorsal and ventral streams, with significant selectivity for given model features. **a.** Representation of unattended speech in the left hemisphere. Spectral, articulatory and semantic representations are labeled in subplots accordingly. **b.** Representation of unattended speech in the right hemisphere. Spectral, articulatory and semantic representations are labeled in subplots accordingly.

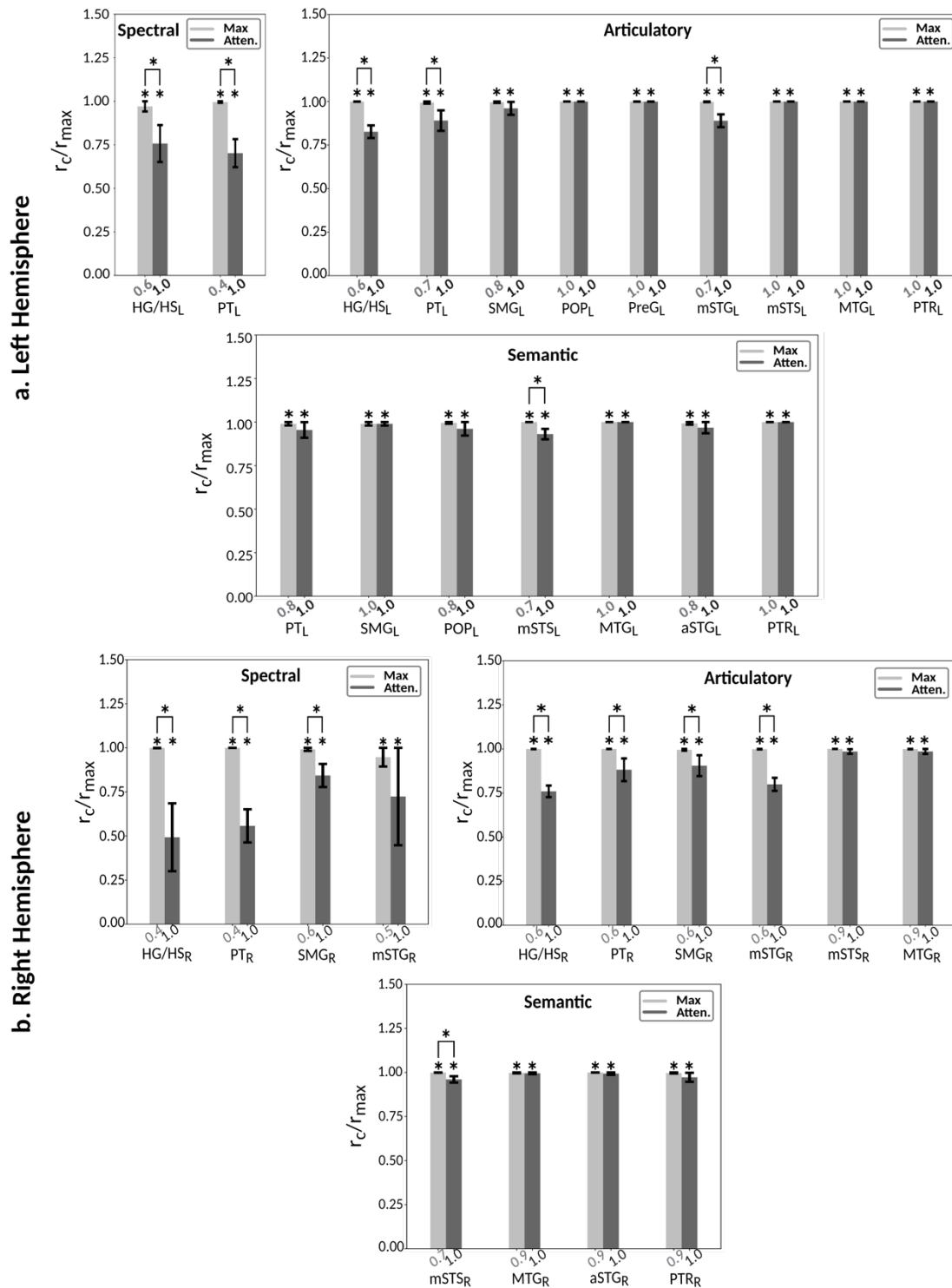


Figure S12e: Representation of unattended speech in subject S5. Passive-listening models were tested on cocktail-party data to assess representation of unattended speech during the cocktail-party task. Prediction scores were calculated separately for a combination model comprising features of both attended and unattended stories (r_{max} : optimal convex combination) and an individual model only comprising features of the attended story (r_a). Significant difference in prediction between the two models is an indication that BOLD responses carry significant information on unattended speech. Bar plots display normalized prediction scores (mean \pm sem across speech-selective voxels in each ROI; combination model in light gray and individual model in gray). Significant scores are marked with * ($p < 10^{-4}$, bootstrap test), and significant differences are marked with brackets ($p < 0.05$). Prediction scores are only displayed for ROIs in the dorsal and ventral streams, with significant selectivity for given model features. **a.** Representation of unattended speech in the left hemisphere. Spectral, articulatory and semantic representations are labeled in subplots accordingly. **b.** Representation of unattended speech in the right hemisphere. Spectral, articulatory and semantic representations are labeled in subplots accordingly.