Voxelwise encoding models show that cerebellar language representations are highly conceptual

Amanda LeBel[1], Shailee Jain[3], Alexander G. Huth[2,3]

[1]Helen Wills Neuroscience Institute, University of California Berkeley, Berkeley, CA 94720, USA
[2]Department of Neuroscience, The University of Texas at Austin, Austin, TX 78712, USA
[3]Department of Computer Science; The University of Texas at Austin, Austin, TX 78712, USA

Correspondence should be addressed to:
      Alexander G. Huth
      100 E 24th St., NHB 2.504
      The University of Texas at Austin
      Austin, TX 78712
      Email: <huth@cs.utexas.edu>

Running Title: Language encoding models of the cerebellum

Manuscript Summary:
      Pages: 40
      Abstract Words: 202
      Introduction Words: 898 (not including figure caption)
      Discussion Words: 1158
      Figures: 6
      Supplementary Figures: 12
      References: 55

LeBel et al. Language encoding models of the cerebellum

32

33

34 **Abstract**
35 There is a growing body of research demonstrating that the cerebellum is involved in language
36 understanding. Early theories assumed that the cerebellum is involved in low-level language
37 processing. However, those theories are at odds with recent work demonstrating cerebellar
38 activation during cognitive tasks. Using natural language stimuli and an encoding model
39 framework, we performed an fMRI experiment where subjects passively listened to five hours of
40 natural language stimuli which allowed us to analyze language processing in the cerebellum
41 with higher precision than previous work. We used this data to fit voxelwise encoding models
42 with five different feature spaces that span the hierarchy of language processing from acoustic
43 input to high-level conceptual processing. Examining the prediction performance of these
44 models on separate BOLD data shows that cerebellar responses to language are almost
45 entirely explained by high-level conceptual language features rather than low-level acoustic or
46 phonemic features. Additionally, we found that the cerebellum has a higher proportion of voxels
47 that represent social semantic categories, which include "social" and "people" words, and lower
48 representations of all other semantic categories, including "mental", "concrete", and "place"
49 words, than cortex. This suggests that the cerebellum is representing language at a conceptual
50 level with a preference for social information.

51

52 **Significance Statement**
53 Recent work has demonstrated that, beyond its typical role in motor planning, the cerebellum is
54 implicated in a wide variety of tasks including language. However, little is known about the
55 language representations in the cerebellum, or how those representations compare to cortex.
56 Using voxelwise encoding models and natural language fMRI data, we demonstrate here that
57 language representations are significantly different in the cerebellum as compared to cortex.
58 Cerebellum language representations are almost entirely semantic, and the cerebellum contains
59 over-representation of social semantic information as compared to cortex. These results
60 suggest that the cerebellum is not involved in language processing per se, but cognitive
61 processing more generally.

62

63 **Introduction**
64 The cerebellum is known to be involved in a diverse set of cognitive processes including
65 attention(Allen et al., 1997), working memory(Brissenden et al., 2018), object recognition(Liu et
66 al., 1999), and language processing(Booth et al., 2007; Stoodley & Schmahmann, 2009).
67 Evidence for the cognitive function of the cerebellum in healthy subjects has come largely from
68 neuroimaging studies, which have found that certain cognitive tasks elicit consistently localized
69 BOLD responses across cerebellum(King et al., 2018) and that resting-state BOLD fluctuations
70 in cerebellum align to known resting-state networks in cortex(Buckner et al., 2011; Marek et al.,
71 2018). However, little is known about what role the cerebellum plays in cognitive processes, or
72 how representations in the cerebellum might differ from those found in cortex.

73

74 Language understanding is a highly complex cognitive process, which makes it a rich area of
75 research to study cognitive processing. Hierarchically organized networks for language

76 processing are widely distributed across much of cortex(J. R. Binder et al., 1997; de Heer et al.,
77 2017; Dronkers et al., 2004; Hickok & Poeppel, 2007; Poeppel et al., 2012). These networks
78 include some putative "language specific" areas in temporal and inferior frontal
79 cortex(Fedorenko et al., 2011), as well as non-language specific conceptual areas in temporal,
80 parietal, and prefrontal cortex(Fedorenko et al., 2013). However, it is unclear whether these
81 networks are also reflected in the cerebellum. Clinical evidence for a cerebellar role in language
82 processing is found in work on cerebellar cognitive affective syndrome (CCAS), which shows
83 that patients with acquired cerebellar damage experience language degradation which can
84 include agrammatism, dysprosody, and anomia(Schmahmann & Sherman, 1998). However, the
85 subtlety and variability of these effects have made it difficult to form a complete picture. Early
86 work into language deficits from cerebellar lesions has often conflicted with cases suggesting a
87 degradation in grammar while preserving semantic content(Frank et al., 2008; Justus, 2004; M.
88 C. Silveri et al., 1994) and other work suggesting a more uniform degradation in language
89 processing that includes semantic content(Cook et al., 2004; Fiez et al., 1992; Maria Caterina
90 Silveri & Misciagna, 2000). However, it is unclear if the standard aphasia tests used in these
91 studies are sensitive enough to detect deficits from cerebellar damage (Cook et al., 2004;
92 Murdoch, 2010). Our goal is to determine how language perception is localized in the
93 cerebellum, what aspects of language are represented in the cerebellum, and how this
94 compares to language processing systems in cortex.
95
96 Here we modeled cortical and cerebellar representations of natural speech using three different
97 categories of features that span the putative language processing hierarchy (de Heer et al.,
98 2017; Hickok & Poeppel, 2007): modality-specific, language-specific, and conceptual. Modality-
99 specific features capture information specific to how people perceive the language stimulus. In
100 this study, subjects listened to audio recordings of naturally spoken narrative stories, so we
101 used a feature space that captures frequency information in sound(Cheung et al., 2016). This
102 feature space is known to be represented in auditory cortex(de Heer et al., 2017). Building upon
103 modality-specific features, language-specific features capture information that only exists in
104 language, like phoneme articulations and syntax. These feature spaces are known to be
105 represented in STG(de Heer et al., 2017; Fedorenko et al., 2011) and inferior frontal cortex(de
106 Heer et al., 2017). Finally, conceptual features capture information about the meaning conveyed
107 by language, which is known to be represented across broad regions of cortex, overlapping with
108 other cognitive tasks(de Heer et al., 2017; Fedorenko et al., 2013). Previous work used similar
109 methods to demonstrate that there is a hierarchy across these feature categories in cortex,
110 where modality-specific information feeds into language-specific and then conceptual
111 representations(de Heer et al., 2017). Here we investigated whether this hierarchy is replicated
112 in the cerebellum, or if the cerebellum is specifically involved in only some aspects of language
113 processing. For ease of language, "cortex" here refers exclusively to the cerebral cortex and
114 "cerebellum" refers to the whole cerebellum, as cerebellar white matter was not excluded from
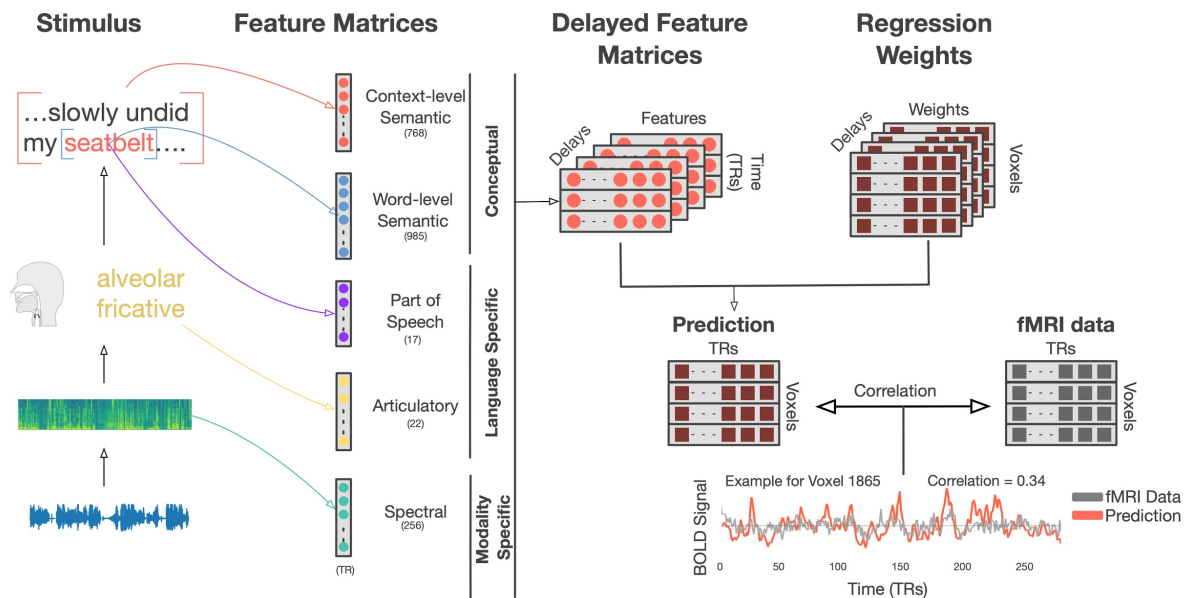115 analysis.
116
117

Figure 1. Voxelwise encoding model construction. To localize different stages of language processing across the cerebellum we used five feature spaces to predict voxelwise BOLD responses in each subject: spectral, articulatory, part-of-speech, word-level semantic, and context-level semantic. Each 10-15 minute stimulus story was transcribed and temporally aligned to the audio recording at the word and phoneme level. Features were then extracted for each of the five feature spaces. The features for the spectral model are 256 bands of a mel-frequency spectrogram, the features for the articulatory model are a 22 length n-hot vector, the features for the part-of-speech model are a 1-hot 17 length vector, the features for the word-level semantic model are a 985-dimensional vector based on statistical word co-occurrence, and the features for the context-level semantic model are a 768 dimensional vector based on GPT(Radford et al., 2018), a neural network language model that incorporates context (preceding words) into the representation of the current word. Features were extracted for each timepoint, word, or phoneme, and concatenated into a feature matrix. The feature matrix was then resampled to the rate of the BOLD signal (0.5 Hz) and delayed to form an FIR model that accounts for hemodynamics. Then regularized linear regression was used to fit weights that predict each voxel's BOLD signal from the stimulus matrix. Finally, models were used to predict responses on a held out test dataset that was not used for model fitting. Model performance was assessed as the linear correlation between held out BOLD data and model predictions for each voxel.

118
119   To determine which aspects of language the cerebellum is involved in processing or
120   representing, we conducted a functional MRI (fMRI) experiment where subjects passively
121   listened to 27 natural, narrative stories (5.4 hours) about a diverse set of topics. We then used
122   voxelwise encoding models (**Figure 1**) to determine how well each set of speech-related
123   features could predict each voxel in each subject. The stimuli were first transformed into 5
124   different feature spaces: spectral, articulatory, part-of-speech, word-level semantic, and context-
125   level (multi-word) semantic. We used ridge regression to fit voxelwise encoding models with
126   each feature space, and then tested how well these encoding models could predict responses
127   to a new story that was not used for model fitting. Finally, we used variance partitioning to
128   measure how much variance in cerebellar and cortical BOLD responses is uniquely explained
129   by each of the five feature spaces. We found substantial evidence that the cerebellum

130   represents language at a high conceptual and semantic level, and no strong evidence that the
131   cerebellum represents any language-specific or modality-specific information.

132

133   In addition, we used the word-level semantic encoding models to determine whether the
134   cerebellum represents different semantic categories than cortex. This analysis showed that all
135   semantic categories are represented in both the cerebellum and cortex, but that the cerebellum
136   has an overrepresentation of social semantic categories and an underrepresentation of mental,
137   concrete, and place-related semantic categories as compared to cortex.

138

139   **Methods**
140   *Participants.* Data was collected from three male subjects and two female subjects: UT-S-
141   01(female, age=24), UT-S-02 (male, age 34), UT-S-06 (female, age 23), UT-S-07 (male,
142   age=25), UT-S-08 (male, age=24). Three of the subjects were authors (UT-S-01:S.J, UT-S-
143   02:A.G.H, and UT-S-06:A.L). All subjects were healthy and had normal hearing. The
144   experimental protocol was approved by the Institutional Review Board at the University of Texas
145   at Austin. Written informed consent was obtained from all subjects.

146

147   *fMRI collection.* MRI data was collected on a 3T Siemens Skyra scanner at the UT Austin
148   Biomedical Imaging Center using a 64 channel Siemens volume coil. Functional scans were
149   collected using gradient echo EPI with repetition time (TR) = 2.00 s, echo time (TE) = 30.8 ms,
150   flip angle = 71°, multi-band factor (simultaneous multi-slice) = 2, voxel size = 2.6mm x 2.6mm x
151   2.6mm (slice thickness = 2.6mm), matrix size = (84, 84), and field of view = 220 mm. Field of
152   view covered both the cortex and the cerebellum in their entirety for all subjects. Anatomical
153   scans were collected using a T1-weighted multi-echo MP-RAGE sequence on the same 3T
154   scanner with voxel size = 1mm x 1mm x 1mm following the Freesurfer morphometry protocol.
155   Anatomical data for subject UT-S-02 was collected on a 3T Siemens TIM Trio at the Berkeley
156   Brain Imaging Center with a 32-Channel Seimen's volume coil using the same sequence.

157

158   Known regions of interest (ROIs) were localized separately in each subject. Three different
159   tasks were used to define ROIs, these include a visual category localizer, an auditory cortex
160   localizer, and a motor localizer.

161

162   For the visual category localizer, data were collected in six 4.5 minute scans consisting of 16
163   blocks of 16 seconds each. During each block 20 images of either places, faces, bodies,
164   household objects, or spatially scrambled objects were displayed. Subjects were asked to pay
165   attention for the same image being presented twice in a row. The corresponding ROIs defined in
166   cortex with this localizer were the fusiform face area (FFA)(Kanwisher et al., 1997), occipital
167   face area (OFA)(Kanwisher et al., 1997), extrastriate body area (EBA)(Downing et al., 2001),
168   parahippocampal place area (PPA)(Epstein & Kanwisher, 1998), and the occipital place area
169   (OPA).

170

171   Motor localizer data were collected during 2 identical 10-minute scans. The subject was cued to
172   perform six different tasks in a random order in 20-second blocks. The cues were 'hand', 'foot',
173   'mouth', 'speak', saccade, and 'rest' presented as a word at the center of the screen, except for

174   the saccade cue which was presented as a random array of dots. For the hand cue subjects
175   were instructed to make small finger-drumming movements for the entirety of the time the cue
176   was displayed. For the foot cue, the subjects were instructed to make small foot and toe
177   movements. For the mouth cue, subjects were instructed to make small vocalizations that were
178   nonsense syllables such as *balabalabala*. For the speak cue, subjects were instructed to self-
179   generate a narrative without vocalization. For the saccade cue, subjects were instructed to look
180   around for the duration of the task.
181
182   Weight maps for the motor areas were used to define primary motor and somatosensory areas
183   for the hands, feet, and mouth; supplemental motor areas for the hands and feet, secondary
184   motor areas for the hands, feet, and mouth, the ventral premotor hand area. The weight map for
185   the saccade responses was used to define the frontal eye field and intraparietal sulcus visual
186   areas. The weight map for the speech production was used to define broca's area and the
187   superior ventral premotor area (sPMV) speech area(Chang et al., 2011). In the cerebellum,
188   weight maps for each subject were resliced in SUIT space(Diedrichsen, 2006) and then the
189   resliced maps were averaged across subjects for each task. Motor areas for the hand, mouth,
190   foot, and saccade tasks were defined in the posterior and anterior lobe.
191
192   Auditory cortex localizer data were collected in one 10-minute scan. The subject listened to 10
193   repeats of 1-minute auditory stimulus each containing 20 seconds of music (Arcade Fire),
194   speech (Ira Glass, This American Life), and natural sound (a babbling brook). To determine
195   whether a voxel was responsive to auditory stimulus, the repeatability of the voxel response
196   across the 10 repeats was calculated using an *F*-statistic. This map was used to define the
197   auditory cortex (AC).
198
199   *fMRI preprocessing.* All functional data were motion corrected using the FMRIB Linear Image
200   Registration Tool (FLIRT) from FSL 5.0(Woolrich et al., 2009). FLIRT was used to align all data
201   to a template that was made from the average of all functional runs in the first story session for
202   each subject. These automatic alignments were manually checked. Low frequency voxel
203   response drift was identified using a 2nd order Savitzky-Golay filter with a 120 second window
204   and then subtracted from the signal. To avoid artifacts from onsets and poor detrending
205   performance, responses were trimmed by removing 20 seconds (10 volumes) at the beginning
206   and end of each scan. This removed the 10-second silent period as well as the first and last 10
207   seconds of each story. The mean response for each voxel was subtracted and the remaining
208   response was scaled to have unit variance.
209
210   *Cortical Surface reconstruction and Visualization.* For cortical surfaces, meshes were generated
211   from the T1-weighted anatomical scans using freesurfer(Dale et al., 1999). Before surface
212   reconstruction, anatomical surface segmentations were hand-checked and corrected. Blender
213   was used to remove the corpus callosum and make relaxation cuts for flattening. Functional
214   images were aligned to the cortical surface using boundary based registration (BBR)
215   implemented in FSL. These were checked for accuracy and adjustments were made as
216   necessary.
217

218   For the cerebellum cortical surfaces, the SUIT toolbox(Diedrichsen, 2006) was used to isolate
219   the cerebellum from the rest of the brain using the T1-weighted anatomical image. The
220   anatomical maps for the cerebellum were normalized into SUIT space using the SUIT
221   registration algorithm. After encoding model fitting, cerebellar functional results were
222   transformed into anatomical space and then resliced using SUIT. The SUIT flatmap and surface
223   was added to the pycortex database for the purpose of surface visualization.
224
225   Model maps were created by projecting the values for each voxel onto the cortical surface using
226   the 'nearest' scheme in pycortex software(Gao et al., 2015). This projection finds the location of
227   each pixel in the image in 3D space, and assigns that pixel the associated value.
228
229   *Stimulus set.* The modeling training stimulus set consisted of 26 10-15 min stories taken from
230   *The Moth Radio Hour*. In each story, a single speaker tells an autobiographical story without
231   reading from a prepared speech. Each story was played during one scan with a buffer of 10
232   seconds on either side of the story start and stop. Data collection was broken up into 6 different
233   days, the first session involving the anatomical scan and localizers, and each successive
234   session consisting of 4 to 5 stories, plus one additional story used for model prediction. This
235   additional story (which was not one of the 26 stories used for model training) was played in
236   every session and the responses to this story were averaged. Stories were played over
237   Sensimetrics S14 in-ear piezoelectric headphones. The audio for each story was filtered to
238   correct for frequency response and phase errors induced by the headphones using calibration
239   data provided by sensimetrics and custom python code
240   (https://github.com/alexhuth/sensimetrics_filter). All stimuli were played at 44.1 kHz using the
241   pygame library in Python.
242
243   Each story was manually transcribed by one listener. Certain sounds (for example, laughter and
244   breathing) were also marked to improve the accuracy of the automated alignment. The audio of
245   each story was downsampled to 11kHz and the Penn Phonetics Lab Forced Aligner
246   (P2FA)(Jiahong Yuan, 2008) was used to automatically align the audio to the transcript. Praat(P
247   Boersma, 2014) was then used to check and correct each aligned transcript manually.
248
249   *Feature Spaces.* Five feature spaces were used to cover the hierarchy of language processing.
250   Each feature space was fit separately for each subject. The spectral feature space was a mel-
251   band spectrogram(Jiahong Yuan, 2008) with frequencies ranging from approximately 0 Hz to
252   8kHz with 256 windows. The articulatory feature space was a n-hot feature space where each
253   phoneme is assigned a 1 for each articulation that is required to produce the sound and a 0 for
254   every other articulation for a total of 22 features per phoneme. For the part-of-speech feature
255   space, a one-hot vector of 17 features was assigned to each word noting the part of speech for
256   each word in each story. Part of speech tagging was done using the flair package(Akbik et al.,
257   2019). Flair is a language model that uses recurrent neural networks to tag speech into 17
258   categories (e.g. noun, verb, number, determiner, etc). The word-level semantic space was a
259   985-dimensional feature space based on word co-occurrence(Huth et al., 2016). Each word in
260   the stimulus set was assigned the vector associated with it in the original space. If the word in
261   the story was not present in the original semantic space, it was assigned a vector of length 985

262   of zeros. The contextual semantic space was based on a fine-tuned GPT language
263   model(Radford et al., n.d.). GPT is a state of the art language model that takes into account
264   previous words while generating features for the current word. To assign features to each word,
265   we extracted 768-dimensional feature vectors from layer 9 with a context length of 25 words.
266   We chose layer 9 because it is a midlayer of GPT and it has been demonstrated that middle
267   layers of recurrent language models are best able to predict brain activity(Jain & Huth, 2018;
268   Toneva & Wehbe, 2019).
269
270   *Experimental Design and Statistical Analysis*
271   *Encoding model fitting.* We used each of the 5 features spaces to fit a linearized finite impulse
272   response (FIR) model to every cortical voxel in each subject. The cerebellar models and the
273   cortical models were fit separately. The stimulus matrix for each story was downsampled using
274   a 3-lobe Lanczos filter, then z-scored and concatenated together. To fit the linear model the
275   stimulus matrix has to account for variance in the hemodynamic response function across
276   voxels. To do this we concatenate 4 delayed copies of the stimulus (using delays of 1, 2, 3, and
277   4 time points). This final stimulus matrix is then regressed with the BOLD data using ridge
278   regression. We then test the model using a held out data set. This is done by taking the dot
279   product of the weight matrix from the regression with the stimulus matrix from the held out test
280   set, resulting in a voxel by timepoint matrix. This resulting matrix is compared to the actual
281   BOLD data for the held out test set and the correlation calculated over time for each voxel to
282   give a measure of model performance. The correlation was then noise-ceiling corrected for
283   some analyses (noted in the text)(Schoppe et al., 2016). Total model performance metrics were
284   computed using the mean R^2 across voxels. Mean was used instead of summation to better
285   account for the difference in number of voxels over the cerebellum as compared to the cortex.
286   To keep the scale of the weights consistent, a single value of the regularization coefficient was
287   used for all voxels in both the cerebellum and cortex in all subjects. To find the best
288   regularization coefficient, the regression procedure was bootstrapped 50 times in each subject
289   and a regularization performance curve was obtained for each subject by averaging the
290   bootstrap sample correlations across the 50 samples, then across voxels, and finally across the
291   6 subjects and the best overall value of the regularization parameter was selected. This was
292   done separately for each feature space.
293
294   *Individual model comparison.* Encoding models with each of the feature spaces were fit in the
295   cerebellum and cortex in each subject and the regression weights were used to predict a held
296   out test set. Model performance for each voxel was estimated by taking the correlation of the
297   predicted time series for each voxel with the actual data. Then to test if the model performance
298   was significant the time series for each voxel was randomly shuffled in blocks of TRs and the
299   correlation with the predicted time series was recalculated. This was done for 10000
300   permutations to gain a null distribution of responses. Lastly, the Fisher-corrected p-value was
301   calculated and this was FDR corrected to account for all the comparisons. A threshold of
302     $q(FDR) < 0.05$ was used to test for significantly well predicted voxels. This was done
303   individually in the cerebellum and cortex in each subject for each model. The correlations were
304   also noise-ceiling corrected. Comparison was done across subjects by taking the average R^2 of
305   all voxels in each subject in the cerebellum and cortex

306
307   *Variance Partitioning.* Because 5-way variance partitioning has too many partitions to be
308   interpretable, we used two versions of variance partitioning to test specific hypotheses. The first
309   version looked at the unique variance explained by each model. This was done to test if one
310   feature space is uniquely better at predicting cerebellar or cortical voxels. The second version
311   was a pairwise variance partitioning where each model was jointly fit with the contextual
312   semantic space. This was done to test for the specific hypothesis that the contextual semantic
313   model is better predicting the same areas as the low level models in the cerebellum, i.e. are
314   there unique low-level language representations in the cerebellum or is the contextual semantic
315   model better predicting the same areas as the low-level models. To do variance partitioning,
316   joint models with the concatenated feature spaces are fit and then used to predict the held out
317   data set. To be succinct, the variance explained by the five feature spaces will be written as sets
318   A-E.
319   Unique Partition -The following nested models were fit as follows
320   $A \cup B \cup C \cup D \cup E, A \cup B \cup C \cup D, A \cup B \cup C \cup E, A \cup B \cup D \cup E, A \cup C \cup D \cup E,$
321   and $B \cup C \cup D \cup E$
322   The variance uniquely explained by each feature space without any overlap from the other
323   feature spaces, or relative complement (RC), was then calculated for each feature space as
324   follows:

$$A^{RC} = A \cup B \cup C \cup D \cup E - B \cup C \cup D \cup E,$$
$$B^{RC} = A \cup B \cup C \cup D \cup E - A \cup C \cup D \cup E$$
$$C^{RC} = A \cup B \cup C \cup D \cup E - A \cup B \cup D \cup E$$
$$D^{RC} = A \cup B \cup C \cup D \cup E - A \cup B \cup C \cup E$$
$$E^{RC} = A \cup B \cup C \cup D \cup E - A \cup B \cup C \cup D$$

330   A Fisher-corrected permutation test with 10000 permutations was done in each subject in both
331   the cerebellum and cortex for each voxel for the unique partitions using the joint $A \cup B \cup C \cup$
332   $D \cup E$ model. Multiple comparison correction was done using FDR with a threshold of $p < 0.05$.
333   Cerebellar data was resliced after the calculation of the unique partitions and the significance
334   testing. The mean of the variance explained was calculated for each subject, in each partition, in
335   the cerebellum and cortex.
336   Pairwise Variance Partitioning - The following concatenated models were fit as follows
337   (where A is the contextual semantic feature space):

$$A \cup B, A \cup C, A \cup D, AUE, A, B, C, D, E$$

339   The variance explained by the intersections were calculated as follows:

$$A \cap B = A + B - A \cup B$$
$$A \cap C = A + C - A \cup C$$
$$A \cap D = A + D - A \cup D$$
$$A \cap E = A + E - A \cup E$$

344   Then the unique contribution of each feature space in each pair can be calculated. This is the
345   unique contribution without overlap from the other feature space noted as RC/X where X is the
346   other paired feature space. These are calculated as follows:

$$A^{RC/B} = A \cup B - B$$
$$B^{RC} = A \cup B - A$$
$$A^{RC/C} = A \cup C - C$$

$$C^{RC} = A \cup C - A$$
$$A^{RC/D} = A \cup D - D$$
$$D^{RC} = A \cup D - A$$
$$A^{RC/E} = A \cup E - E$$
$$E^{RC} = A \cup E - A$$

A Fisher-corrected permutation test with 10000 permutations was done in each subject in both the cerebellum and cortex for each voxel for the unique partitions and intersections using the joint $A \cup B, A \cup C, A \cup D, A \cup E$ models. Multiple comparison correction was done using FDR with a threshold of $p < 0.05$. Cerebellar data was resliced after the calculation of the unique partitions and the significance testing.

Analysis of model weights. To assess similarity of semantic categories between cortex and cerebellum, the semantic space had to be broken into discrete categories instead of a smoothly continuous space. To do this the encoding model weights for the top 25% of voxels predicted by the word-level semantic model in each subject were concatenated together across subjects. This was done separately in cortex and cerebellum, and then those were also concatenated together. Then the model weights were normalized across voxels and PCA was used to drop the number of dimensions from 985 to 86, which we chose because it explained 80% of the variance. These data were then clustered using spherical k-means into 5 clusters.

To choose the number of clusters we calculated inertia, which is the within-cluster sum of squares criterion, of the clustering algorithm for a range of clusters between 1 and 20 clusters. From this we calculated the point where the inertia changes from an exponential drop to a linear drop in inertia. This can also be defined as the point where the inertia is farthest from a linear line connecting the inertia at cluster 1 to the inertia at cluster 20. This point occurred at 5 clusters. (Supplemental Figure 11 shows the inertia across all clusters tested.)

To test for significance in category differences between cerebellum and cortex, a permutation test was done by shuffling voxels between the cortex and the cerebellum for each subject. The difference in the ratio of each category in the cerebellum as compared to the ratio of that category in the cortex was calculated for both the permutation set and the original data. The two-tailed p-value was calculated for each category as the ratio of the permutation difference greater than the absolute value of the original data difference plus the ratio of the permutation difference less than the negative absolute value of the original data. This was multiple comparison corrected using FDR with a threshold of $p < 0.05$.

Data Availability

A dataset including the data used in this study is being prepared for public release. Before the data are publicly available, they are also available upon request.

Results

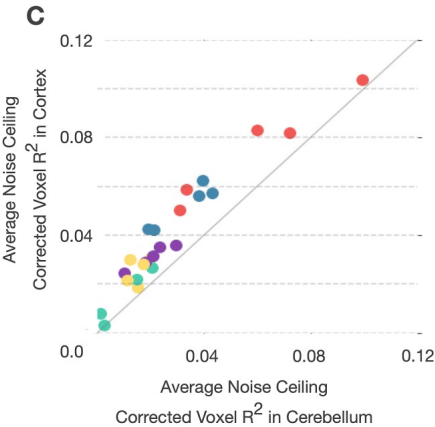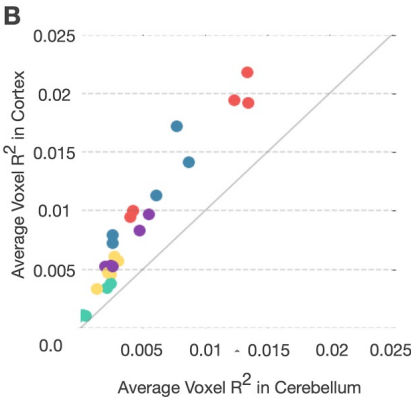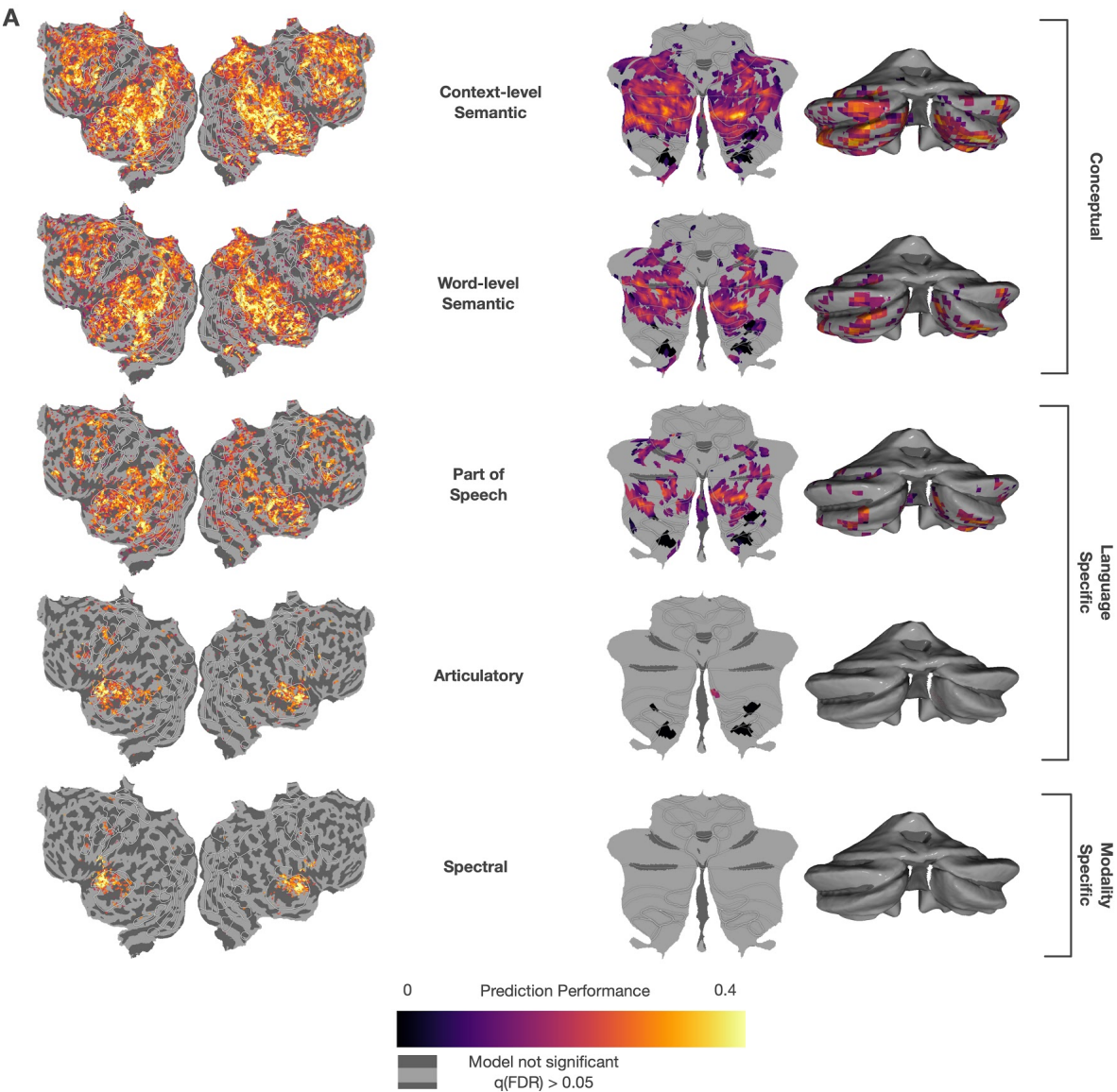*LeBel et al. Language encoding models of the cerebellum*

Figure 2. Prediction performance of encoding models based on five language feature spaces in cortex and cerebellum. Encoding models fit with 5.4 hours of BOLD data were tested against a held out story (10 minutes). (**A**) Correlation between predicted and actual BOLD response is plotted on flattened cortical and cerebellar surfaces for one subject (UT-S-02; other subjects are shown in **Supplementary Figure 1**). Significance testing for each model in each voxel was done using a one-sided FDR-corrected permutation test with a threshold of $p < 0.05$. The higher-level models have better prediction performance in both cerebellum and cortex. In cortex, the areas best predicted by each of the three feature categories are spatially distinct. However, in the cerebellum, the areas best predicted by each feature space are highly overlapping. (**B**) To compare across subjects, we plotted average R^2 across all voxels in the cerebellum and cortex for each subject and each feature space. The context-level semantic feature space has the highest predictive performance in both the cerebellum and cortex for all subjects. Performance scales roughly linearly in both cerebellum and cortex across the hierarchy of language representations, albeit with higher R^2 in cortex than cerebellum. (**C**) Because cortical and cerebellar BOLD responses might have different levels of noise, which could obscure differences in representation, we also computed noise ceiling-corrected correlations(Schoppe et al., 2016). This correction caused the average R^2 to be less biased in favor of cortex (for corrected correlation flatmaps, see **Supplemental Figure 2)** and suggests that each feature space might be represented to a similar extent in cerebellum and cortex. However, overlapping prediction performance between different feature spaces in the cerebellum suggests that the cerebellum may not be separately representing each stage of language processing.

392

393   *Encoding model performance.* To determine which aspects of language might be processed in
394   the cerebellum, we created five feature spaces that span the hierarchy of language processing
395   from sound to context-level meaning, including a spectral feature space, an articulatory space, a
396   part-of-speech space, a word-level semantic space, and a context-level semantic space that
397   combines information across words. Previous work has demonstrated that these feature spaces
398   can capture these different components of language and predict BOLD responses in cortex(de
399   Heer et al., 2017; Huth et al., 2016; Jain & Huth, 2018). We fit separate encoding models with
400   each feature space using 5.4 hours of BOLD responses recorded while subjects listened to 26
401   different natural narrative stories taken from *The Moth Radio Hour*. Then, each model was used
402   to predict responses to a *different* 10-minute story, and model performance was quantified as
403   the correlation between the predicted and actual BOLD responses (R^2). **Figure 2A** shows the
404   prediction performance values for each feature space in one subject projected onto the SUIT
405   cerebellar surface as well as prediction performance of each model in the cortex (similar maps
406   for other subjects are in **Supplemental Figure 1**).

407

408   The spectral model uses a 256-dimensional, modality-specific feature space representing a mel-
409   frequency spectrogram. This feature space is highly predictive of the primary auditory cortex
410   along the transverse temporal gyrus. It does not significantly predict any voxel in the cerebellum
411   (one-sided permutation test, q(FDR)<0.05), but it does appear to have diffuse low prediction
412   performance across lobules VIIA, VIIB, and VIIIA. Of note, this is similar to previous results that
413   showed cerebellar response to auditory stimulus along the medial portion of these
414   lobules(Snider & Stowell, 1944). However, there appears to be no clustering of spectrally-
415   selective voxels in the cerebellum, as is seen in the auditory cortex. This suggests that the
416   cerebellum has no homologous area to the primary auditory cortex.

417

418   The articulatory model uses a 22-dimensional binary, language-specific feature space with each
419   dimension representing one of the 22 articulations used in English (e.g. bilabial, back)(Levelt,

420   1993). In cortex, the articulatory space best predicts lateral, posterior temporal cortex along
421   superior temporal gyrus. In the cerebellum, this feature space has diffuse prediction
422   performance across lobules VIIIA and VIIIB, and significantly predicts a limited number of voxels
423   in the medial posterior cerebellum (one-sided permutation test, q(FDR)<0.05). These areas are
424   not traditionally considered motor speech areas(Callan et al., 2006; Manto et al., 2012) and thus
425   this is unlikely to be due to covert rehearsal. This suggests that the cerebellum is not merely
426   representing the articulations required to produce speech, and the lack of spatial clustering of
427   well-predicted voxels further supports a lack of a homologous area to the auditory cortex.
428
429   The part-of-speech model uses a 17-dimensional binary, language-specific feature space,
430   where each dimension represents one of 17 lexical classes (e.g. noun, verb, adjective). This
431   feature space weakly but significantly predicts voxels covering a wide area of the cortex (one-
432   sided permutation test, q(FDR)<0.05), including much of the frontal, temporal, and parietal
433   lobes, with peak performance along the superior temporal lobe and near the intraparietal sulcus
434   (IPS). In the cerebellum, this model significantly predicts voxels in many areas of the posterior
435   lobe, with the highest model prediction performance in Crus I and II. This is a mid-level,
436   language-specific feature space and its performance suggests that the cerebellum is largely
437   representing information at a higher level than sound or articulations.
438
439   The word-level semantic model uses a 985-dimensional conceptual feature space that is based
440   on word co-occurrence statistics across a large corpus of written English(de Heer et al., 2017;
441   Deniz et al., 2019; Huth et al., 2016). This feature space captures semantic information under
442   the assumption that *words that frequently occur in similar contexts carry similar meaning(FIRTH*
443   *& R, 1957)*. The word-level semantic model predicts cortical voxels across regions in the frontal,
444   parietal, and temporal lobes beyond core language-specific regions(Fedorenko et al., 2011,
445   2013). In the cerebellum, this model significantly predicts voxels in Crus I and II and lobules
446   VIIIA and VIIIB. This conceptual model predicts much more response variance in the cerebellum
447   and cortex than do lower-level models.
448
449   The best model in both the cerebellum and cortex is the context-level semantic model. This
450   model builds on the word-level conceptual model by combining information across words. It
451   uses the hidden state of a neural language model (LM) as a feature space. Neural LMs are
452   artificial neural networks that learn to predict the next word in a sequence from past words. As a
453   consequence, they learn a word's meaning in context, improving upon the word-level model
454   which is context-invariant(Lin et al., 2019; Radford et al., 2019; Tenney et al., 2019). Here, we
455   used GPT(Jain & Huth, 2018; Radford et al., 2018), which is a popular neural LM. The feature
456   space is 768-dimensional and the features are extracted from a middle layer of the LM that has
457   previously been shown to be highly effective at predicting brain responses(Toneva & Wehbe,
458   2019). For each word, the past 25 words are used as context in the model. The context-level
459   semantic model significantly predicts the largest number of voxels and most total variance
460   across cortex, with peak prediction performance in frontal, parietal and temporal cortex. In the
461   cerebellum, this model yields very high prediction performance across most of the posterior
462   cerebellum, including Crus I and II and lobules VIIIA and VIIIB.
463

464   To compare model performance between the cerebellum and cortex directly, we computed the

465   average performance of each model in the cerebellum and cortex for each subject. **Figure 2B**

466   shows that there is a linear relationship between model performance in the cerebellum and

467   cortex, suggesting that language might be represented similarly in these two structures. To

468   account for the possibility that BOLD signal-to-noise varies systematically between cortex and

469   cerebellum, we also adjusted the estimated correlation for each voxel using a standard

470   technique(Schoppe et al., 2016). **Figure 2C** shows these results when accounting for the

471   difference in signal-to-noise variance between cortex and cerebellum. Here, the pattern of

472   results is largely the same, but prediction performance in the cerebellum is more similar to that

473   of cortex. In both cases however, cerebellar voxels that are well predicted by each feature

474   space are highly overlapping. This could be caused by the feature spaces carrying overlapping

475   information with each other, making it difficult to interpret the results from each feature space

476   independently. To disentangle these representations and explore the differences between

477   cortex and cerebellum in more detail, we next performed a variance partitioning analysis.
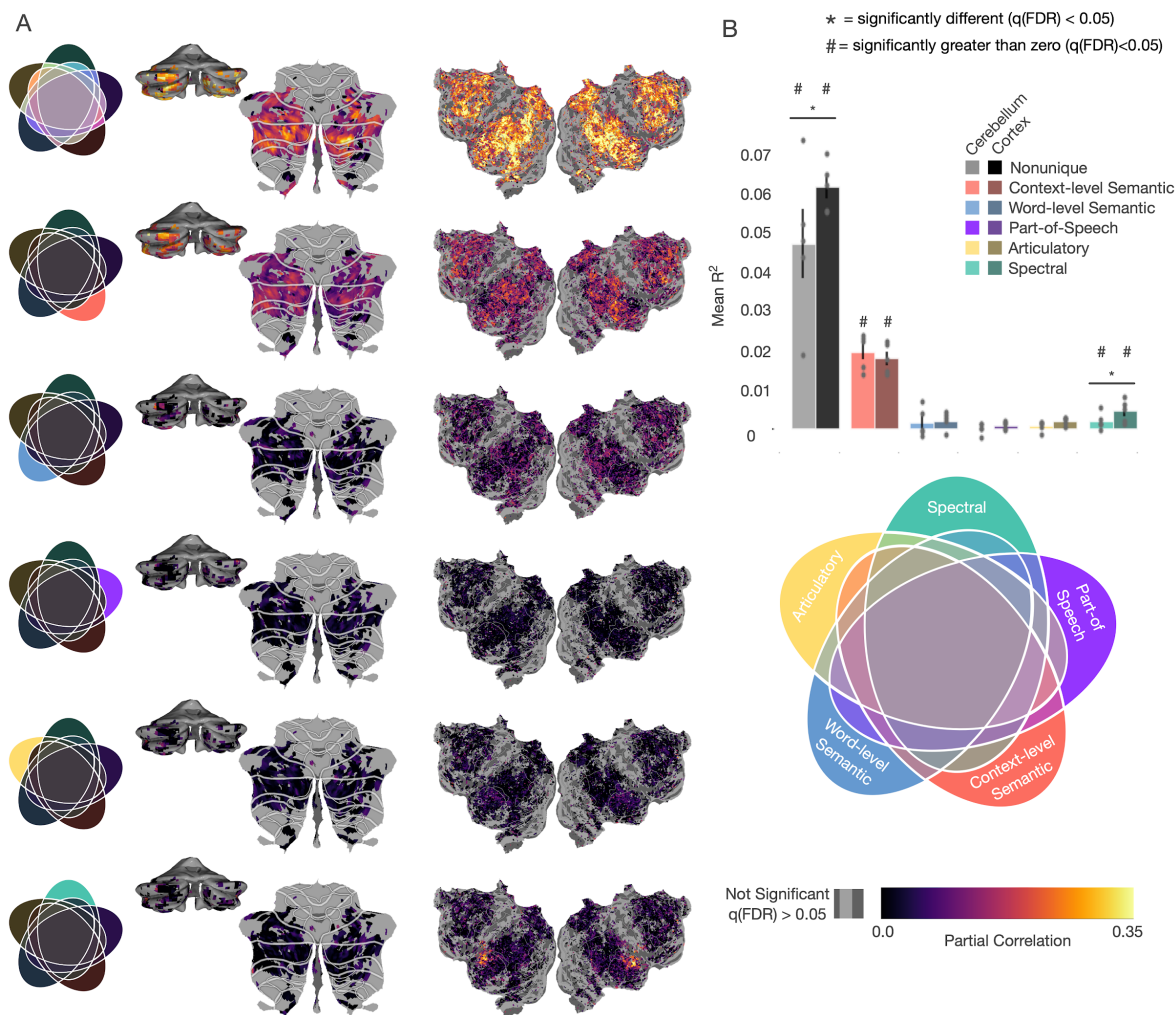

Figure 3. Unique variance explained by each feature space. To determine how much variance is uniquely explained by each feature space, six new encoding models were fit: a union model containing a concatenation of all feature spaces, and five encoding models each containing a concatenation of

four of the five feature spaces. The unique contribution of each feature space was then determined by subtracting the variance explained by the four-way concatenation model without that feature space from the union model. This shows how much variance can be explained by each feature space above and beyond the other four. Additionally, the amount of non-unique variance—i.e., any that can be explained by more than one feature space—was determined by subtracting the 5 unique variances from the union. (**A**) The voxelwise partial correlation ($\sqrt{R^2}$) for each feature space for subject UT-S-02, projected onto the cortical and cerebellar surfaces (similar maps for other subjects are in **Supplemental Figure 3)**. Only voxels that were significantly predicted (one-sided permutation test, q(FDR)<0.05) by the 5-way union model are displayed. (**B**) Mean correlations for significant voxels in the cerebellum and cortex across all subjects. The non-unique partition contains the most variance in both cortex (darker) and cerebellum (lighter), but is significantly smaller in cerebellum than cortex (two-sided permutation test, q(FDR)<0.05). The modality-specific spectral feature space explains significantly less variance in the cerebellum as compared to the cortex. Additionally, the modality-specific feature spaces do not uniquely explain any significant variance (one-sided permutation test, q(FDR)<0.05), while the context-level semantic space uniquely explains the most variance. This further supports the hypothesis that the cerebellum is largely representing language at a high, conceptual level.

478

479   *Variance Partitioning*. The previous model comparison found that many voxels in the cerebellum
480   can be significantly predicted by multiple feature spaces. These voxels might genuinely
481   represent information from multiple feature spaces. Indeed, the increased neuronal density of
482   cerebellum compared to cortex(Herculano-Houzel, 2010) raises the chance that individual
483   cerebellar voxels contain information from multiple feature spaces. However, this effect could
484   also be a consequence of correlations, or shared information, between the feature spaces. To
485   disentangle possible overlaps in information across the five feature spaces within each voxel we
486   used variance partitioning, a statistical technique for determining how much variance can be
487   uniquely explained by each set of features(de Heer et al., 2017; Lescroart et al., 2015). This
488   enables us to distinguish between overlapping but distinct representations and seemingly
489   overlapping representations that actually reflect correlations between features. For example,
490   variance partitioning would allow us to disentangle if say 50% of the voxel responds to
491   conceptual information and another 50% to auditory information, or if 100% of the voxel
492   response is to some feature that is correlated with both auditory and conceptual information.
493   Our first variance partitioning analysis shows how much variance each feature space uniquely
494   explains above all other feature spaces for each voxel, and the second shows how much
495   overlap there is between each feature space and the context-level semantic feature space.
496

497   *Unique Variance Explained.* The results in **Figure 2a** showed negligible, localized prediction
498   performance of low-level models in the cerebellum, suggesting that little low-level language
499   processing was occurring there. However, that result did not account for the possibility that
500   higher-level feature spaces could also capture some low-level information. To test for this, we
501   used variance partitioning to find the unique variance explained of each feature space in order
502   to test whether the lower level models have any unique contribution to representations in the
503   cerebellum. This was done by first fitting a five-way union encoding model with a concatenation
504   of all the feature spaces. Variance explained by any of the five feature spaces should be
505   explained by this five-way union model. Then we fit five additional encoding models, each
506   combining four of the five feature spaces. Each of these models should explain all the variance
507   captured by the five-way union model except for variance that is uniquely explained by the

508 feature space that was left out. To estimate the unique variance explained by each feature
509 space, we then subtracted the variance explained in the four-way model excluding that feature
510 space from the five-way union model (for additional details, see **Methods**). We also used these
511 models to estimate the size of the non-unique partition, which contains any variance that can be
512 explained by more than one of the five feature spaces. The size of this partition was calculated
513 by subtracting each of the unique variance partitions from the five-way union model. For this
514 analysis, we only considered voxels that were significantly predicted by the five-way union
515 model (one-sided permutation test, q(FDR)<0.05).

516

517 **Figure 3A** shows the unique variance explained by each feature space as well the non-unique
518 partition for each voxel in the cerebellum and cortex projected onto the flattened surface for one
519 subject (other subjects can be seen in **Supplemental Figure 3**). The non-unique partition is the
520 largest partition overall, suggesting that much of the variance explained by these feature spaces
521 cannot be specifically allocated to one feature space. It is important to note that this category
522 includes all possible combinations of the feature spaces and does not mean that the variance is
523 explained equally well by each of the five feature spaces. Among the unique partitions, both the
524 context-level semantic model and spectral feature space explain variance significantly greater
525 than zero (one-sided permutation test, q(FDR)<0.05). However, the spectral model explains
526 significantly (two-sided permutation test, q(FDR)<0.05) more variance in cortex than in the
527 cerebellum. **Figure 3B** shows the unique variance explained for each feature space averaged
528 across voxels for all subjects (only including voxels that were significantly predicted by the union
529 model). We compared mean partial correlations ($\sqrt{R^2}$) between cerebellum and cortex for each
530 partition using a permutation test. The result shows that the spectral feature space and non-
531 unique partitions explain significantly less variance in the cerebellum than in cortex. When
532 correcting for differences in signal-to-noise (**Supplemental Figures 4 & 5**), the context-level
533 semantic and word-level semantic feature spaces uniquely explain significantly more variance in
534 the cerebellum than cortex, and the spectral feature space uniquely explains significantly less.
535 Both of these results suggest that the cerebellum is primarily representing language at a
536 conceptual level and that these results are not simply due to neuronal pooling within voxels or
537 shared representations. However, the fact that the largest proportion of variance is in the non-
538 unique partition means that this analysis alone can not rule out the possibility for low-level
539 language representations in the cerebellum.

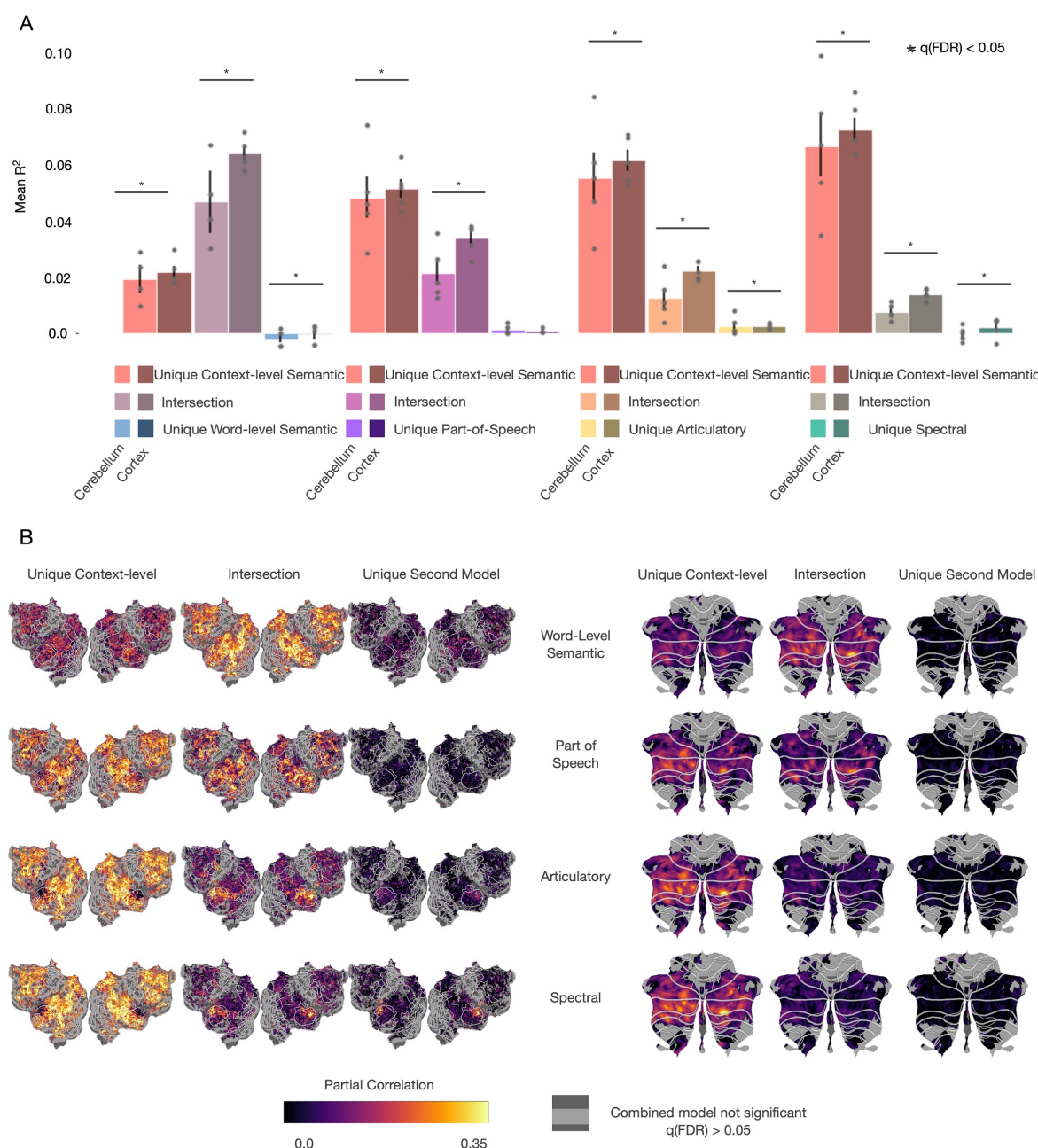*LeBel et al. Language encoding models of the cerebellum*

Figure 4. Variance partitioning between the context-level semantic feature space and each of the other feature spaces. To quantify the amount of overlap between the context-level semantic feature space and each of the four other feature spaces, three models were fit for each pair of feature spaces, including the concatenation of both feature spaces and each feature space individually. (**A**) For each pair feature spaces, the variance uniquely explained by the context-level feature space, that uniquely explained by the second feature space, and the intersection between the two is compared between the cerebellum and cortex, averaged over all subjects. The intersection—variance that could be explained by either feature space—for every pair is smaller in the cerebellum than in the cortex (two-sided permutation test, q(FDR)<0.05). Additionally, the unique partition for the spectral feature space is significantly smaller in the cerebellum than in cortex. This shows that the high prediction performance of the context-level semantic feature space in cerebellum is not merely due to correlations with

modality- and language-specific information. Instead, the context-level features uniquely explain a large amount of variance that the other features cannot. (**B**) For each pair of models, the variance in each partition in each voxel was projected onto cortical and cerebellar flatmaps (**Supplemental Figures 8, 9, & 10,** noise-ceiling corrected version to account for differences in signal-to-noise across the brain). Only voxels that were significantly predicted by each union model (one-sided permutation test, q(FDR)<0.05) are shown. There is substantially lower variance explained by the intersection between the context-level semantic feature space and the language and modality-specific feature spaces in the cerebellum than in cortex. Additionally, the unique contributions for these feature spaces in the cerebellum is near zero and is not spatially localized. This lack of spatial localization further suggests that there is no hierarchy of language processing in the cerebellum, and these results provide strong support for the hypothesis that the cerebellum only represents high level, conceptual features of language, rather than low-level features.

540   *Pairwise partitioning.* In the first variance partitioning analysis, we found that the context-level
541   semantic feature space explains the most unique variance explained and that the spectral
542   model explains significantly less variance in the cerebellum than in cortex. This suggests that
543   the cerebellum may not be representing information at modality and language-specific levels.
544   However, the largest partition in both the cerebellum and cortex was the non-unique partition,
545   which contains variance that could be explained by more than one feature space. Thus, that
546   analysis alone cannot rule out the possibility that low-level features are represented in
547   cerebellum. To test the hypothesis that the cerebellum is exclusively representing language at a
548   conceptual level, we performed a second variance partitioning analysis where each feature
549   space was separately compared to the context-level semantic feature space. We fit four union
550   models by concatenating the context-level semantic features with each one of the four other
551   feature spaces. The variance explained by each union model was then compared to models fit
552   with each feature space individually in order to determine both the unique contribution of each
553   feature space and the size of their intersection. For each pair of feature spaces, analyses were
554   restricted to voxels that were significantly predicted by the union model. If the cerebellum was
555   only representing information at the conceptual level, we would expect to find low unique
556   variance explained by the modality- and language-specific feature spaces and a high shared
557   intersection with the word-level semantic feature space.
558
559   The results of this pairwise variance partitioning analysis replicate previous results(de Heer et
560   al., 2017) showing that in cortex there is a unique contribution of both the spectral and
561   articulatory feature spaces in different cortical areas. However, this does not appear to be true
562   in the cerebellum. **Figure 4** shows the results of pairwise variance partitioning between the
563   context-level semantic feature space and each of the other four feature spaces. **Figure 4A**
564   shows the mean partial correlation for each pair of feature spaces in both the cerebellum and
565   cortex across voxels and subjects. The variance explained by the intersection of each pair of
566   models is significantly less in the cerebellum than in cortex (two-sided permutation test,
567   q(FDR)<0.05). This shows that the information present in the lower level feature spaces
568   contributes less to the explainable variance in the cerebellum and supports the hypothesis that
569   the cerebellum is primarily representing high level, conceptual information. Additionally, the
570   unique contribution from the modality- and language-specific feature spaces are negligible; the
571   spectral feature space explains significantly less variance in the cerebellum, while the
572   articulatory feature spaces explain significantly more variance in the cerebellum, although this
573   partition is small in both the cerebellum and cortex. When accounting for differences in signal-

574 to-noise (**Supplemental Figures 8, 9, & 10),** all of the unique contributions from the secondary
575 models become significantly less in the cerebellum than in cortex. Additionally, the differences
576 in the intersections between the cerebellum and cortex are no longer significant. However, the
577 unique contribution from the context-level semantic feature space is significantly larger in all
578 cases in the cerebellum and cortex. While the noise-ceiling corrected results are different due to
579 the differences in BOLD signal in the cerebellum as compared to cortex, the significantly larger
580 variance explained by the context-level semantic feature space in the cerebellum still supports
581 the hypothesis that the cerebellum is uniquely representing highly conceptual semantic
582 information. **Figure 4B** shows these results for one subject projected onto the corresponding
583 cortical and cerebellar surfaces (cortical maps for all other subjects can be found in
584 **Supplemental Figures 6** and **7**). Only voxels that were significantly predicted by the union
585 model (one-sided permutation test, q(FDR)<0.05) are displayed.
586

587 Very little variance in the cerebellum is explained uniquely by any features other than the
588 context-level semantic space. In both the cerebellum and cortex, there is a high amount of
589 variance explained by the context-level semantic feature space and in the intersection with the
590 word-level semantic feature space. This is not surprising, given that the context-level semantic
591 space has the highest predictive performance of any of the feature spaces and that the word-
592 level and context-level semantic spaces contain related semantic information. However, there is
593 very little overlap of variance explained between the context-level semantic feature space and
594 the three modality- and language-specific feature spaces. This demonstrates that the high
595 performance of the conceptual feature spaces is not merely due to this feature space being
596 correlated with low-level information. The negligible unique contribution of the modality- and
597 language-specific features in the cerebellum further supports the hypothesis that the cerebellum
598 is primarily representing conceptual representations. Finally, any variance explained by the
599 modality- and language-specific feature spaces is not anatomically localized within cerebellum,
600 which suggests that the cerebellum does not contain localized low-level language processing
601 areas. The reduced representation of language specific feature spaces in the cerebellum further
602 suggests that the cerebellum does not participate in language processing per se, but supports
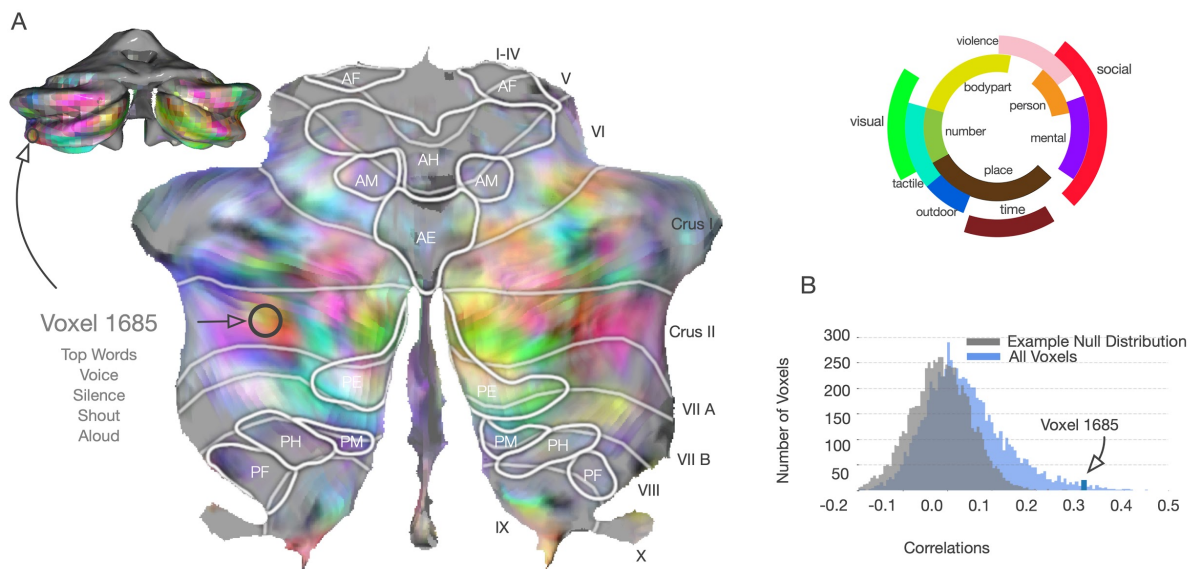603 cognition more generally.
604

Figure 5. Word-level semantic model weight interpretation. Post hoc analysis of encoding models enables us to interpret what type of semantic information is represented in each voxel. Here we used the word-level semantic feature space to interpret one individual voxel and to broadly map semantic representations across the cerebellum. (While the context-level semantic space is more predictive, we lack tools for interpreting its representations.) In the word-level space, encoding models predict the response of each voxel to each word. We used the model to find words with the largest predicted response in one voxel (voxel 1685 in subject UT-S-02), which were "voice", "silence", and "shout", suggesting that this voxel represents concepts related to social communication. To visualize representations across many voxels, we reduced the encoding model weights to three dimensions by projecting them onto a low-dimensional semantic space identified in a previous experiment(Huth et al., 2016), and then mapping these projections to RGB color channels. (**A**) The RGB values for each voxel are projected onto the SUIT cerebellar surface for subject UT-S-02. Different colors correspond to selectivity for different concepts in the semantic space (illustrated by the legend, right). This map suggests that the cerebellum contains representations of many different concepts. This histogram (**B**) shows the range of correlations for each voxel in this subject, with the example voxel marked in blue, and null distribution in gray.

605   *Semantic selectivity within the cerebellum.* Our results thus far suggest that the cerebellum is
606   not involved with language-specific processing, as there is little or no unique variance explained
607   in cerebellum by the part-of-speech, articulatory, or spectral feature spaces. Instead, language
608   representations in the cerebellum appear to be dominated by conceptual semantic features. Yet
609   all semantic representations are not alike: in cortex, earlier work revealed a patchwork tiling of
610   areas that represent different semantic categories across much of prefrontal, parietal, and
611   temporal cortex(Jeffrey R. Binder et al., 2009; Huth et al., 2016; Noppeney & Price, 2004). It is
612   possible that the cerebellum represents a different range of semantic categories than cortex,
613   and it seems likely that different categories are represented in distinct areas within the
614   cerebellum. Following the procedure detailed in Huth et al. (2016), we used the word-level
615   semantic feature space to analyze and interpret the model weights and thus reveal the semantic
616   selectivity of each voxel in the cerebellum. Due to the lack of tools currently available for
617   interpreting context-level semantic models, we chose to use the word-level model, which also
618   explains a large proportion of response variance in the cerebellum.
619

620   To demonstrate how encoding models can be analyzed per voxel, **Figure 5A** shows the word
621   level semantic regression weights projected into a three-dimensional semantic space that was
622   previously constructed from a group of subjects using principal components analysis
623   (PCA)(Huth et al., 2012). This lower-dimensional space is purely used for visualization
624   purposes. Here projections on the first, second, and third principal components (PCs) are
625   mapped into the red, green, and blue color channels respectively, for each voxel and then
626   projected onto the SUIT cerebellar surface. The color wheel shows approximately which
627   semantic category each color on the maps represents. **Figure 5A** shows the posterior view of
628   one subject's (UT-S-02) cerebellum as well as the flattened cerebellar surface in SUIT space.
629   Within the SUIT space, functional regions of interest are mapped out which include anterior foot
630   (AF), hand(AH), and mouth (AM); posterior foot(PF), hand(PH) and mouth (PM); and anterior
631   and posterior eye movement areas (AE and PE) that are active during saccades. A histogram of
632   correlations for all voxels in subject UT-S-02 is shown in **Figure 5B**. This histogram shows a
633   distribution with a long tail, with the example well-predicted voxel (voxel 1685) marked in blue.
634   Additionally, **Figure 5A** lists the five words that the word-level semantic encoding model
635   predicts will elicit the largest response in this example voxel, which are "voice", "silence",
636   "shout", and "aloud". These words were found by taking the dot product of the voxel weight
637   vector with the word-level semantic feature matrix (see **Methods** for details). This voxel seems
638   selective for concepts related to social communication and sound. Similar analysis could be
639   performed for each voxel, but would be large and difficult to interpret. However, by representing
640   semantic weights as a color we can better understand large-scale patterns of semantic
641   information. For example, Crus I and Crus II seem to be selective for many different semantic
642   categories, such as *social* and *violence* which can be found in medial Crus I.
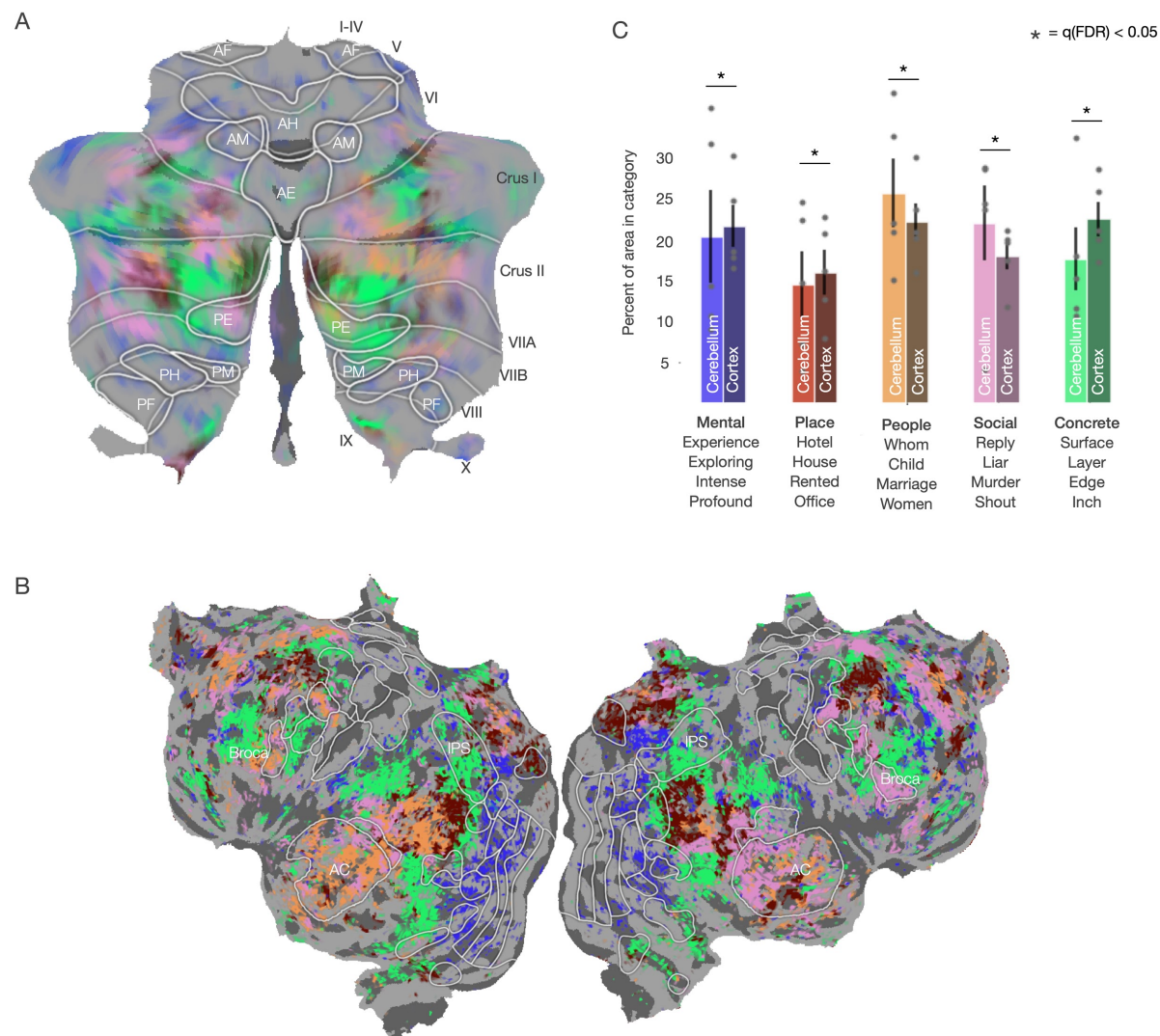
Figure 6. Differences in semantic representations between cerebellum and cortex. To check for differences in semantic representations between the cerebellum and cortex, word-level encoding model weights from both cerebellum and cortex in all subjects were concatenated, including only the top 20% best-predicted voxels. This matrix was then clustered using spherical k-means into 5 clusters, which fell at the inflection point in the inertia graph (**Supplemental Figure 11**). For visualization, the centroid for each cluster was transformed into the same RGB space used in **Figure 5**, and each voxel in that cluster was assigned that color. (**A**) shows the cluster distribution for one subject, UT-S-02, across the cerebellum and (**B**) cortex. Voxels falling into each cluster are found in both the cerebellum and cortex in every subject. (**C**) To test for differences in representation between cortex and cerebellum, the percentage of cortical and cerebellar voxels in each cluster were compared across all subjects. Each cluster was named qualitatively according to the most similar words to the cluster centroid. The four most similar words to each cluster centroid are listed below the label name. Significantly more voxels in the cerebellum were highly responsive to social categories (two-sided permutation test, q(FDR)<0.05), i.e. the "social" and "people" clusters, than in cortex. Conversely, significantly fewer voxels in the cerebellum were responsive to the "mental", "concrete", or "place" clusters than in the cortex. This shows that the cerebellum is largely representing the same semantic categories as cortex, but that there is a slight bias towards social categories.

643
644

645

646   *Comparing semantic representations between cerebellum and cortex*. The semantic map in
647   **Figure 5A** shows that different areas in the cerebellum represent different categories of words.
648   Yet it is not clear from this map whether semantic representations in the cerebellum are similar
649   to those found in cortex. To quantify the semantic categories represented in the cerebellum and
650   cortex, we compared the fraction of voxels that represent different semantic categories using a
651   cluster analysis. We concatenated model weights for the top 20% best predicted voxels in
652   cerebellum and cortex from each subject, then clustered the voxels into 5 discrete categories
653   using spherical k-means clustering (5 clusters was the elbow point of the inertia curve, see
654   **Supplemental Figure 11**; similar results are also obtained with different numbers of clusters).
655   **Figure 6A** and **B** show cerebellar and cortical flatmaps with the clustered voxels colored
656   according to their assigned cluster in one subject (similar maps for other subjects can be found
657   in **Supplemental Figure 12**). The label for each cluster was determined qualitatively from the
658   most similar words to each cluster centroid (**Figure 6C** lists the clusters, their top words, and
659   their assigned label).

660

661   Voxels belonging to every semantic cluster were found in both the cerebellum and cortex.
662   **Figure 6C** shows the percentage of cerebellar voxels in each cluster as compared to the
663   percentage of cortical voxels in each cluster, averaged across subjects. The category with the
664   highest percentage of voxels in the cerebellum is the "people" category and the category with
665   the lowest percentage is the "place" category.

666

667   Because voxels in all clusters are found in both the cerebellum and cortex, it is possible that the
668   cerebellum is receiving input from all areas of cortex. If this was true, we would expect to find an
669   equal percentage of well-predicted cerebellar voxels in each cluster as there are in cortex.
670   However, all clusters had significantly different percentages of voxels in the cerebellum as
671   compared to cortex (two-sided permutation test, $q(FDR)<0.05$). The "social" and "people"
672   clusters have a higher percentage of voxels in the cerebellum than in cortex, and the "mental",
673   "concrete", and "place" clusters have a lower percentage of voxels in cerebellum than in cortex.
674   This suggests that there is not a one-to-one mapping from cortex to the cerebellum and that the
675   cerebellum is more responsive to social semantic information.

676

677   **Discussion**
678   This study examined how language is represented in the human cerebellum. Using voxelwise
679   encoding models trained within each subject using large amounts of fMRI data, we found that
680   high-level language feature spaces–context-level and word-level semantics–were better able to
681   predict cerebellar BOLD responses than low-level language feature spaces, such as part-of-
682   speech and articulations. Additionally, the low-level feature spaces do not uniquely predict any
683   voxel in the cerebellum above the context-level semantic model, which is not true in the cortex.
684   Lastly, using the model weights from the word-level semantic model, we found that there is an
685   overrepresentation of social and people semantic categories in the cerebellum as compared to
686   cortex. These results suggest that the cerebellum is (1) representing language at a conceptual
687   level, and not at modality- or language-specific levels, (2) that there is not a homologous area to
688   auditory cortex in the cerebellum, and (3) that the cerebellum is more responsive to social

689   semantic components of language than cortex. As has been seen previously(King et al., 2019),
690   there does not appear to be any functional relevance to lobule boundaries as we do not observe
691   any pattern of language processing that corresponds to the lobule boundaries.
692
693   One complication in interpreting the results of this study is due to the use of BOLD fMRI, in
694   particular in relation to the cerebellum. The cerebellum has a significantly different metabolic
695   demand than cortex(Vaishnavi et al., 2010) due to its cellular architecture. This changes the
696   demand for oxygenated blood and thus the BOLD signal. It has previously been demonstrated
697   that only activity in granule cells and mossy fibers affects the BOLD signal(Caesar et al., 2003;
698   Mathiesen et al., 2000) in the cerebellum, but not activity in the Purkinje cells which are the sole
699   output from the cerebellum to cortex. This implies that our models do not include
700   representations of what the cerebellum is outputting back to the cortex and thus may not directly
701   address the computation the cerebellum is performing. However, the input to the cerebellum–
702   granule cells and mossy fibers– is still an important half of the equation and this work furthers
703   our understanding of what kinds of representations are being sent to the cerebellum.
704
705   One point of contention with our methodology is using a natural stimulus. While natural stimuli
706   can make interpretation of the results more difficult, it is a much richer stimulus set for analysis.
707   Additionally, it is less biased than other experimental methods that preselect a small number of
708   categories or stimuli. And while we have few subjects, we have collected a large amount of data
709   per subject. This amount of data allows for the regression models to theoretically account for
710   most stimulus correlations. Additionally, by using a prediction methodology we are able to
711   compute the variance explained by each feature space which allows us to quantify how well
712   each model does at prediction which few other methods allow for.
713
714   Many theories exist for how the cerebellum represents cognitive information based on the
715   uniformity of its cellular architecture. This architecture is believed to suggest that the cerebellum
716   is performing a similar function throughout the structure. Additionally, the cerebellum has long
717   been considered a major region in motor response and motor learning(Manto et al., 2012). Yet
718   since the 1980s, the cerebellum has been known to reliably respond during cognitive
719   tasks(Leiner et al., 1986) such as language processing(Petersen et al., 1988) and that lesions to
720   the posterior lobe of the cerebellum results in language deficits(Schmahmann & Sherman,
721   1998). The fact that both fine motor control processing and cognitive processing elicit strong
722   responses from the same architecture has long been considered a contradiction. In an effort to
723   reconcile the cerebellum as both a cognitive area and a motor area, previous reports have
724   speculated that the cerebellum is involved in some low-level component of cognitive tasks such
725   as low-level auditory processing(Petacchi et al., 2005) or motor planning in speech(Jürgens,
726   2002).
727
728   Surprisingly, our results show that low level spectral and articulatory feature spaces do not
729   uniquely predict any area of the cerebellum better than high-level feature spaces. However, the
730   spectral and articulatory models best predict areas around auditory cortex and the STG. This
731   shows that these feature spaces successfully capture auditory information and that this
732   information does not appear to be present in the cerebellum. The cerebellum is very likely

733   involved with motor components of speech production. However, we found no evidence of
734   receptive articulatory representations in the cerebellum, suggesting it must be involved in more
735   than just motor components of speech perception. Our results imply that the cerebellum lacks
736   any form of localized low-level auditory processing area and that the role of the cerebellum in
737   language processing is at a higher level than previously thought.
738

739   An important future step is to clarify the relationship between language representations in the
740   cerebellum and existing theories of cerebellar function. The universal cerebellar transform
741   (UCT) theory is the predominant theory of cerebellar computation(Diedrichsen et al., 2019),
742   positing that the cerebellum performs a single computation across all tasks, both cognitive and
743   motor. A commonly proposed computation is prediction error(Kawato & Gomi, 1992; Mariën &
744   Manto, 2018). One way to look at whether the cerebellum is involved in prediction is through
745   surprisal which is a measure of the probability of a word occurring in a sentence given the
746   previous word. Thus a word with a high surprisal is likely to have a high prediction error. Since
747   the context-level semantic model is using a neural network based language model, it inherently
748   captures some elements of surprisal(Berger et al., 1996). However, the context-level semantic
749   model best predicts both the cerebellum and cortex which suggests that the cerebellum is not
750   uniquely computing surprisal.
751

752   In language processing, many processes are specific to auditory communication such as the
753   spectral and articulatory features spaces. However, the higher order semantic features seem to
754   be more broadly used by the default mode network. Given that the cerebellum does not appear
755   to be involved in the lower level language processing, this implies that the cerebellum is not
756   participating in language processing per se, and is likely only involved in cognitive processing.
757   This theory could explain many of the language deficits seen in patients with CCAS and autism.
758   Both of these disorders are associated with cerebellar damage or morphological changes and
759   both often see deficits in language processing(Stoodley & Schmahmann, 2009). However, the
760   deficits are not specifically related to speech production or the ability to interpret sound into
761   phonemes and words, which are low-level language-specific processes. Rather, the language
762   deficits in CCAS and autism more often present as conceptual deficits, with a loss of
763   understanding of fine-tuned semantic specificity and social dynamics(Kelley et al., 2006), such
764   as understanding sarcasm and non-explicit language. Much like the cerebellum being involved
765   in the fine-tuning of motor commands over a continuous three-dimensional space, it is possible
766   that the cerebellum is similarly involved in the fine-tuning of a conceptual cognitive space.
767
768

769   References

770   Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., & Vollgraf, R. (2019). FLAIR: An

771   Easy-to-Use Framework for State-of-the-Art NLP. *Proceedings of the 2019 Conference of*

772   *the North American Chapter of the Association for Computational Linguistics*

773   *(Demonstrations)*, 54–59.

774   Allen, G., Buxton, R. B., Wong, E. C., & Courchesne, E. (1997). Attentional activation of the

775   cerebellum independent of motor involvement. *Science*, *275*(5308), 1940–1943.

776   Berger, A. L., Della Pietra, V. J., & Della Pietra, S. A. (1996). A Maximum Entropy Approach to

777   Natural Language Processing. *Computational Linguistics*, *22*(1), 39–68.

778   Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where Is the Semantic

779   System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies.

780   *Cerebral Cortex* , *19*(12), 2767–2796.

781   Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human

782   brain language areas identified by functional magnetic resonance imaging. *The Journal of*

783   *Neuroscience: The Official Journal of the Society for Neuroscience*, *17*(1), 353–362.

784   Booth, J. R., Wood, L., Lu, D., Houk, J. C., & Bitan, T. (2007). The role of the basal ganglia and

785   cerebellum in language processing. *Brain Research*, *1133*, 136–144.

786   Brissenden, J. A., Tobyne, S. M., Osher, D. E., Levin, E. J., Halko, M. A., & Somers, D. C.

787   (2018). Topographic Cortico-cerebellar Networks Revealed by Visual Attention and

788   Working Memory. *Current Biology: CB*, *28*(21), 3364–3372.e5.

789   Buckner, R. L., Krienen, F. M., Castellanos, A., Diaz, J. C., & Yeo, B. T. T. (2011). The

790   organization of the human cerebellum estimated by intrinsic functional connectivity. *Journal*

791   *of Neurophysiology*, *106*(5), 2322–2345.

792   Caesar, K., Gold, L., & Lauritzen, M. (2003). jContext sensitivity of activity-dependent increases

793   in cerebral blood flow. *Proceedings of the Natural Academy of Science*, *100*(7), 4239–4244.

794   Callan, D. E., Tsytsarev, V., Hanakawa, T., Callan, A. M., Katsuhara, M., Fukuyama, H., &

795   Turner, R. (2006). Song and speech: brain regions involved with perception and covert

796   production. *NeuroImage*, *31*(3), 1327–1342.

797   Chang, S.-E., Horwitz, B., Ostuni, J., Reynolds, R., & Ludlow, C. L. (2011). Evidence of left

798   inferior frontal-premotor structural and functional connectivity deficits in adults who stutter.

799   *Cerebral Cortex* , *21*(11), 2507–2518.

800   Cheung, C., Hamiton, L. S., Johnson, K., & Chang, E. F. (2016). The auditory representation of

801   speech sounds in human motor cortex. *eLife*, *5*. https://doi.org/10.7554/eLife.12577

802   Cook, M., Murdoch, B., Cahill, L., & Whelan, B. (2004). Higher-level language deficits resulting

803   from left primary cerebellar lesions. *Aphasiology*, *18*(9), 771–784.

804   Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation

805   and surface reconstruction. *NeuroImage*, *9*(2), 179–194.

806   de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L., & Theunissen, F. E. (2017). The

807   Hierarchical Cortical Organization of Human Speech Processing. *The Journal of*

808   *Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(27), 6539–6557.

809   Deniz, F., Nunez-Elizalde, A. O., Huth, A. G., & Gallant, J. L. (2019). The Representation of

810   Semantic Information Across Human Cerebral Cortex During Listening Versus Reading Is

811   Invariant to Stimulus Modality. *The Journal of Neuroscience: The Official Journal of the*

812   *Society for Neuroscience*, *39*(39), 7722–7736.

813   Diedrichsen, J. (2006). *A spatially unbiased atlas template of the human cerebellum*.

814   https://doi.org/10.1016/j.neuroimage.2006.05.056

815   Diedrichsen, J., King, M., Hernandez-Castillo, C., Sereno, M., & Ivry, R. B. (2019). Universal

816   Transform or Multiple Functionality? Understanding the Contribution of the Human

817   Cerebellum across Task Domains. *Neuron*, *102*(5), 918–928.

818   Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for

819   visual processing of the human body. *Science* , *293*(5539), 2470–2473.

820   Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Jr, Redfern, B. B., & Jaeger, J. J. (2004).

821   Lesion analysis of the brain areas involved in language comprehension. *Cognition*, *92*(1-2),

822   145–177.

823   Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment.

824   *Nature*, *392*(6676), 598–601.

825   Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level

826   linguistic processing in the human brain. *Proceedings of the National Academy of Sciences*

827   *of the United States of America*, *108*(39), 16428–16433.

828   Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of

829   frontal and parietal cortex. *Proceedings of the National Academy of Sciences of the United*

830   *States of America*, *110*(41), 16616–16621.

831   Fiez, J. A., Petersen, S. E., Cheney, M. K., & Raichle, M. E. (1992). Impaired non-motor

832   learning and error detection associated with cerebellar damage. A single case study. *Brain:*

833   *A Journal of Neurology*, *115 Pt 1*, 155–178.

834   FIRTH, & R, J. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in Linguistic Analysis*.

835   https://ci.nii.ac.jp/naid/10020680394/

836   Frank, B., Schoch, B., Hein-Kropp, C., Hövel, M., Gizewski, E. R., Karnath, H.-O., & Timmann,

837   D. (2008). Aphasia, neglect and extinction are no prominent clinical signs in children and

838   adolescents with acute surgical cerebellar lesions. *Experimental Brain Research.*

839   *Experimentelle Hirnforschung. Experimentation Cerebrale*, *184*(4), 511–519.

840   Gao, J. S., Huth, A. G., Lescroart, M. D., & Gallant, J. L. (2015). Pycortex: an interactive surface

841   visualizer for fMRI. *Frontiers in Neuroinformatics*, *9*, 23.

842   Herculano-Houzel, S. (2010). Coordinated scaling of cortical and cerebellar numbers of

843   neurons. *Frontiers in Neuroanatomy*, *4*(12). https://doi.org/10.3389/fnana.2010.00012

844   Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature*

845   *Reviews. Neuroscience*, *8*(5), 393–402.

846   Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural

847   speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600),

848   453–458.

849   Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space

850   describes the representation of thousands of object and action categories across the

851   human brain. *Neuron*, *76*(6), 1210–1224.

852   Jain, S., & Huth, A. G. (2018). *Incorporating context into language encoding models for fMRI*

853   (Vol. 2018-Decem, pp. 6628–6637). http://papers.nips.cc/paper/7897-incorporating-context-

854   into-language-encoding-models-for-fmri.pdf

855   Jiahong Yuan, M. L. (2008). Speaker identification on the SCOTUS corpus. *In Proceedings of*

856   *Acoustics 2008*. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.227.6546

857   Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience and Biobehavioral*

858   *Reviews*, *26*(2), 235–258.

859   Justus, T. (2004). The cerebellum and English grammatical morphology: evidence from

860   production, comprehension, and grammaticality judgments. *Journal of Cognitive*

861   *Neuroscience*, *16*(7), 1115–1130.

862   Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in

863   human extrastriate cortex specialized for face perception. *The Journal of Neuroscience:*

864   *The Official Journal of the Society for Neuroscience*, *17*(11), 4302–4311.

865   Kawato, M., & Gomi, H. (1992). A computational model of four regions of the cerebellum based

866   on feedback-error learning. *Biological Cybernetics*, *68*(2), 95–103.

867   Kelley, E., Paul, J. J., Fein, D., & Naigles, L. R. (2006). Residual Language Deficits in Optimal

868   Outcome Children with a History of Autism. *Journal of Autism and Developmental*

869   *Disorders*, *36*(6), 807–828.

870   King, M., Hernandez-Castillo, C. R., Poldrack, R. A., Ivry, R. B., & Diedrichsen, J. (2019).

871   Functional boundaries in the human cerebellum revealed by a multi-domain task battery.

872   *Nature Neuroscience*, *22*(8), 1371–1378.

873   King, M., Hernandez-castillo, C. R., Poldrack, R., & Ivry, R. B. (2018). A Multi-Domain Task

874   Battery Reveals Functional Boundaries in the Human Cerebellum. *Neuron*, 1–41.

875   Leiner, H. C., Leiner, A. L., & Dow, R. S. (1986). Does the cerebellum contribute to mental

876   skills? *Behavioral Neuroscience*, *100*(4), 443–454.

877   Lescroart, M. D., Stansbury, D. E., & Gallant, J. L. (2015). Fourier power, subjective distance,

878   and object categories all provide plausible models of BOLD responses in scene-selective

879   visual areas. *Frontiers in Computational Neuroscience*, *9*, 135.

880   Levelt, W. J. M. (1993). *Speaking: From Intention to Articulation*. MIT Press.

881   Lin, Y., Tan, Y. C., & Frank, R. (2019). Open Sesame: Getting Inside BERT's Linguistic

882   Knowledge. In *arXiv [cs.CL]*. arXiv. http://arxiv.org/abs/1906.01698

883   Liu, L., Loannides^, A. A., & Strait, M. (1999). *Single Trial Analysis of Neurophysiological*

884   *Correlates of the Recognition of Complex Objects and Facial Expressions of Emotion* (Vol.

885   11).

886   Manto, M., Bower, J. M., Conforto, A. B., Delgado-García, J. M., Suzete, &., Farias Da Guarda,

887   N., Gerwig, M., Habas, C., Hagura, N., Ivry, R. B., Mariën, P., Molinari, M., Naito, E.,

888   Nowak, D. A., Oulad, N., Taib, B., Pelisson, D., Tesche, C. D., Tilikete, C., … Naito, E.

889   (2012). Consensus Paper: Roles of the Cerebellum in Motor Control-The Diversity of Ideas

890   on Cerebellar Involvement in Movement. *Cerebellum* , *11*, 457–487.

891   Marek, S., Siegel, J. S., Gordon, E. M., Raut, R. V., Gratton, C., Newbold, D. J., Ortega, M.,

892   Laumann, T. O., Adeyemo, B., Miller, D. B., Zheng, A., Lopez, K. C., Berg, J. J., Coalson,

893   R. S., Nguyen, A. L., Dierker, D., Van, A. N., Hoyt, C. R., McDermott, K. B., … Dosenbach,

894   N. U. F. (2018). Spatial and Temporal Organization of the Individual Human Cerebellum.

895   *Neuron*, *100*(4), 977–993.e7.

896   Mariën, P., & Manto, M. (2018). Cerebellum as a Master-Piece for Linguistic Predictability.

897   *Cerebellum* , *17*(2), 101–103.

898   Mathiesen, C., Caesar, K., & Lauritzen, M. (2000). Temporal coupling between neuronal activity

899   and blood flow in rat cerebellar cortex as indicated by field potential analysis. *The Journal*

900   *of Physiology*, *523*(1), 235–246.

901 Murdoch, B. E. (2010). The cerebellum and language: Historical perspective and review. *Cortex;*

902   *a Journal Devoted to the Study of the Nervous System and Behavior*, *46*(7), 858–868.

903 Noppeney, U., & Price, C. J. (2004). Retrieval of abstract semantics. *NeuroImage*, *22*(1), 164–

904   170.

905 P Boersma, D. W. (2014). *Praat: doing phonetics by computer*.

906 Petacchi, A., Laird, A. R., Fox, P. T., & Bower, J. M. (2005). Cerebellum and auditory function:

907   an ALE meta-analysis of functional neuroimaging studies. *Human Brain Mapping*, *25*(1),

908   118–128.

909 Petersen, S. E., Fox, P. T., Posner, M. I., Minton, M., & Raichle, M. E. (1988). Positron emission

910   tomographic studies of the cortical anatomy of single-word processing. *Nature*, *331*.

911   https://www.nature.com/articles/331585a0.pdf

912 Poeppel, D., Emmorey, K., Hickok, G., & Pylkkänen, L. (2012). Towards a new neurobiology of

913   language. *The Journal of Neuroscience: The Official Journal of the Society for*

914   *Neuroscience*, *32*(41), 14125–14131.

915 Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (n.d.). *Improving Language*

916   *Understanding by Generative Pre-Training*. http://s3-us-west-2.amazonaws.com/openai-

917   assets/research-covers/language-unsupervised/language_understanding_paper.pdf

918 Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language*

919   *understanding by generative pre-training*. OpenAI.

920   https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf

921 Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models

922   are unsupervised multitask learners. *OpenAI Blog*, *1*(8), 9.

923 Schmahmann, J. D., & Sherman, J. C. (1998). The cerebellar cognitive affective syndrome.

924   *Brain: A Journal of Neurology*, *121*(4), 561–579.

925   Schoppe, O., Harper, N. S., Willmore, B. D. B., King, A. J., & Schnupp, J. W. H. (2016).

926   Measuring the Performance of Neural Models. *Frontiers in Computational Neuroscience*,

927   *10*, 10.

928   Silveri, M. C., Leggio, M. G., & Molinari, M. (1994). The cerebellum contributes to linguistic

929   production: a case of agrammatic speech following a right cerebellar lesion. *Neurology*,

930   *44*(11), 2047–2050.

931   Silveri, M. C., & Misciagna, S. (2000). Language, memory, and the cerebellum. *Journal of*

932   *Neurolinguistics*, *13*(2), 129–143.

933   Snider, R. S., & Stowell, A. (1944). RECEIVING AREAS OF THE TACTILE, AUDITORY, AND

934   VISUAL SYSTEMS IN THE CEREBELLUM. *Journal of Neurophysiology*, *7*(6), 331–357.

935   Stoodley, C. J., & Schmahmann, J. D. (2009). The cerebellum and language: Evidence from

936   patients with cerebellar degeneration. *Brain and Language*, *110*(3), 149–153.

937   Tenney, I., Xia, P., Chen, B., Wang, A., Poliak, A., Thomas McCoy, R., Kim, N., Van Durme, B.,

938   Bowman, S. R., Das, D., & Pavlick, E. (2019). What do you learn from context? Probing for

939   sentence structure in contextualized word representations. In *arXiv [cs.CL]*. arXiv.

940   http://arxiv.org/abs/1905.06316

941   Toneva, M., & Wehbe, L. (2019). Interpreting and improving natural-language processing (in

942   machines) with natural language-processing (in the brain). *Advances in Neural Information*

943   *Processing Systems*. http://papers.nips.cc/paper/9633-interpreting-and-improving-natural-

944   language-processing-in-machines-with-natural-language-processing-in-the-brain

945   Vaishnavi, S. N., Vlassenko, A. G., Rundle, M. M., Snyder, A. Z., Mintun, M. A., & Raichle, M. E.

946   (2010). Regional aerobic glycolysis in the human brain. *Proceedings of the National*

947   *Academy of Science*, *107*(41), 17757–17762.

948   Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann,

949   C., Jenkinson, M., & Smith, S. M. (2009). Bayesian analysis of neuroimaging data in FSL.

950   *NeuroImage*, *45*(1 Suppl), S173–S186.

951

952   Supplemental Figures
953
954
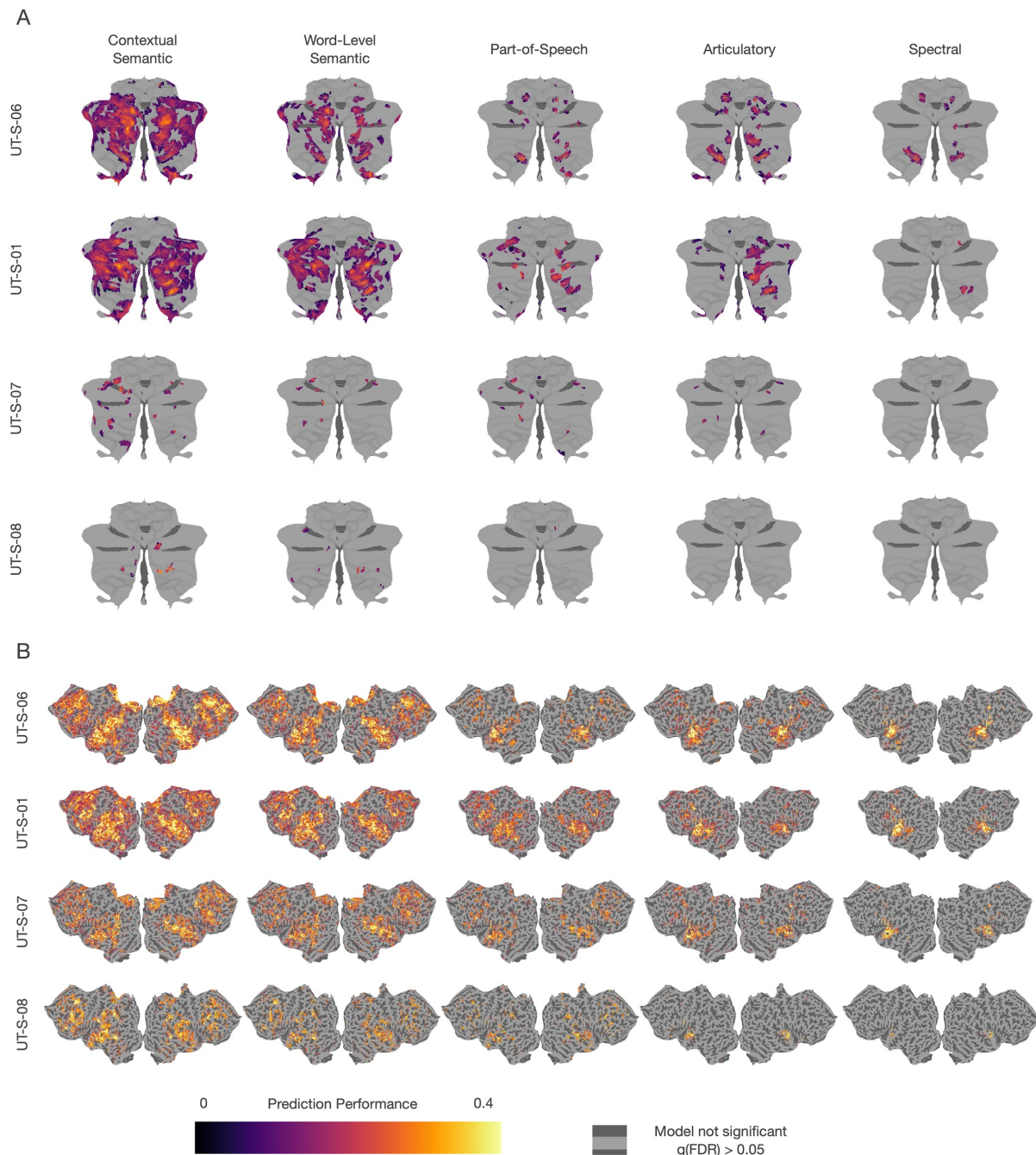
Supplemental Figure 1. Prediction performance of encoding models based on five language feature spaces in cortex and cerebellum. Encoding models fit with 5.4 hours of BOLD data were tested against a held out story (10 minutes). The correlation between predicted and actual BOLD response is plotted on flattened cerebellar (**A**) and cortical (**B**) surfaces for each subject. Significance testing for each model was done using a one-sided FDR-corrected permutation test with a threshold of

$p < 0.05$. The higher level models have better prediction performance in both cerebellum and cortex. In cortex, the areas best predicted by each of the three categories of feature spaces are spatially distinct. However, in the cerebellum the areas best predicted by each of the feature spaces is highly overlapping. This suggests that there is a hierarchy of language processing in cortex but not in the cerebellum.

955

Supplemental Figure 2. Prediction performance of encoding models based on five language

feature spaces correcting for difference in signal-to-noise in cerebellum and cortex. Encoding models fit with 5.4 hours of BOLD data were tested against a held out story (10 minutes). The correlation between predicted and actual BOLD response is plotted on flattened cerebellar (**A**) and cortical (**B**) surfaces for one subject (UT-S-02).The correlations were noise-ceiling corrected using standard techniques to account for differences in BOLD signal-to-noise ratio in the cerebellum and cortex. Significance testing for each model was done using a one-sided FDR-corrected permutation test with a threshold of $p < 0.05$. The higher level models have better prediction performance in both cerebellum and cortex. In cortex, the areas best predicted by each of the three categories of feature spaces are spatially distinct. However, in the cerebellum the areas best predicted by each of the feature spaces is highly overlapping. This suggests that there is a hierarchy of language processing in the cortex and not in the cerebellum

956

Supplemental Figure 3. Unique variance explained for each feature space. To determine the unique variance explained by each feature space, a union encoding model was fit with a concatenation of all feature spaces in addition to five other encoding models - each a concatenation of four of the feature spaces. The unique contribution of a model can be determined by the subtraction of the four way concatenation model without that feature space from the five way union model. Additionally, the amount of overlap between the feature spaces can be characterized by the non unique partition. (**A**) The unique variance explained by each feature space in each additional subject were projected onto the cortical and cerebellar surfaces. Only significant voxels (one-sided permutation test, q(FDR)<0.05) from the 5-way joint model are displayed. The partition that explains the most variance in both cortex and cerebellum is the nonunique partition, which shows that there is overlap in the information in the

feature spaces. The largest unique partition in the cerebellum and cortex is the context-level semantic feature spaces which shows that both areas likely represent conceptual information. The largest difference in unique partitions between the cerebellum and cortex is the spectral feature space, which does not uniquely explain any region of the cerebellum.
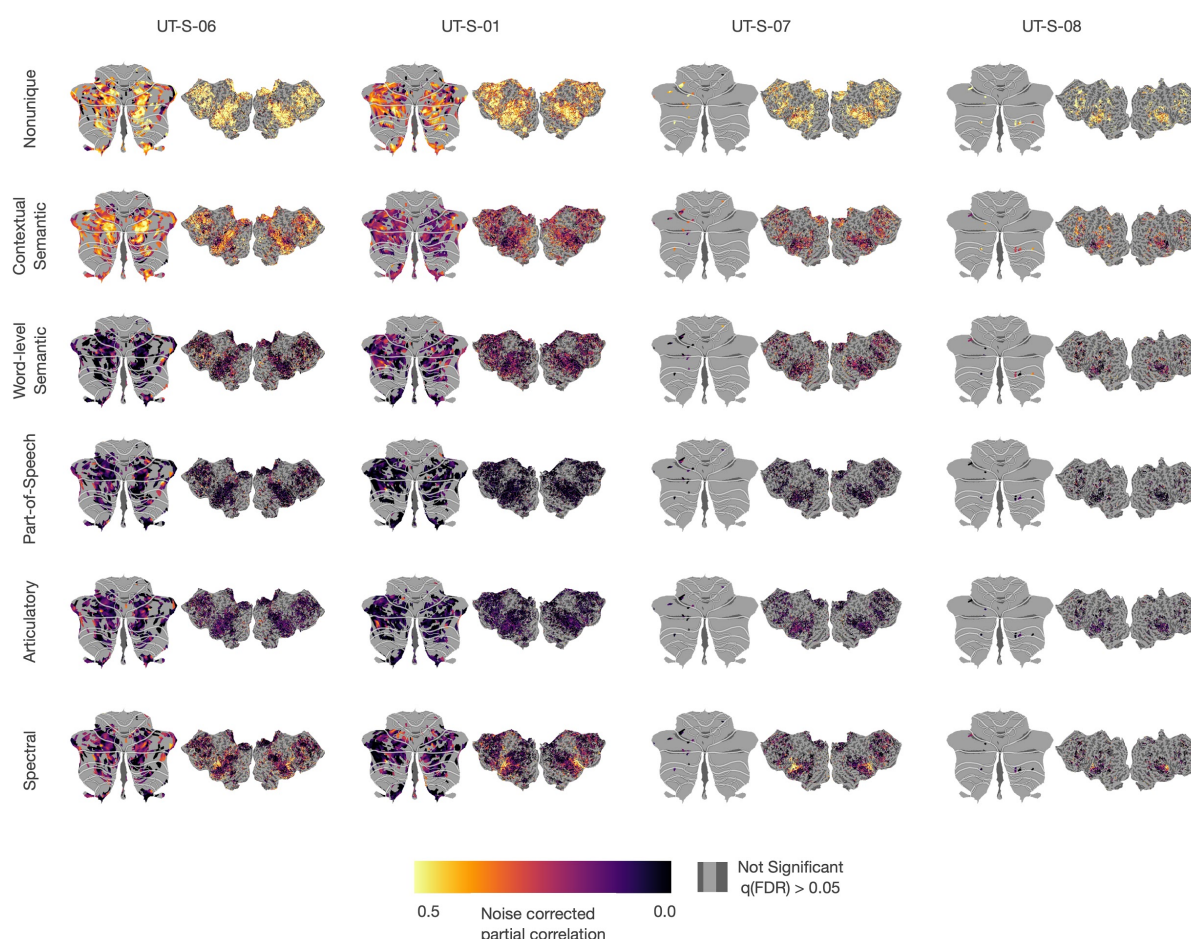
957
958
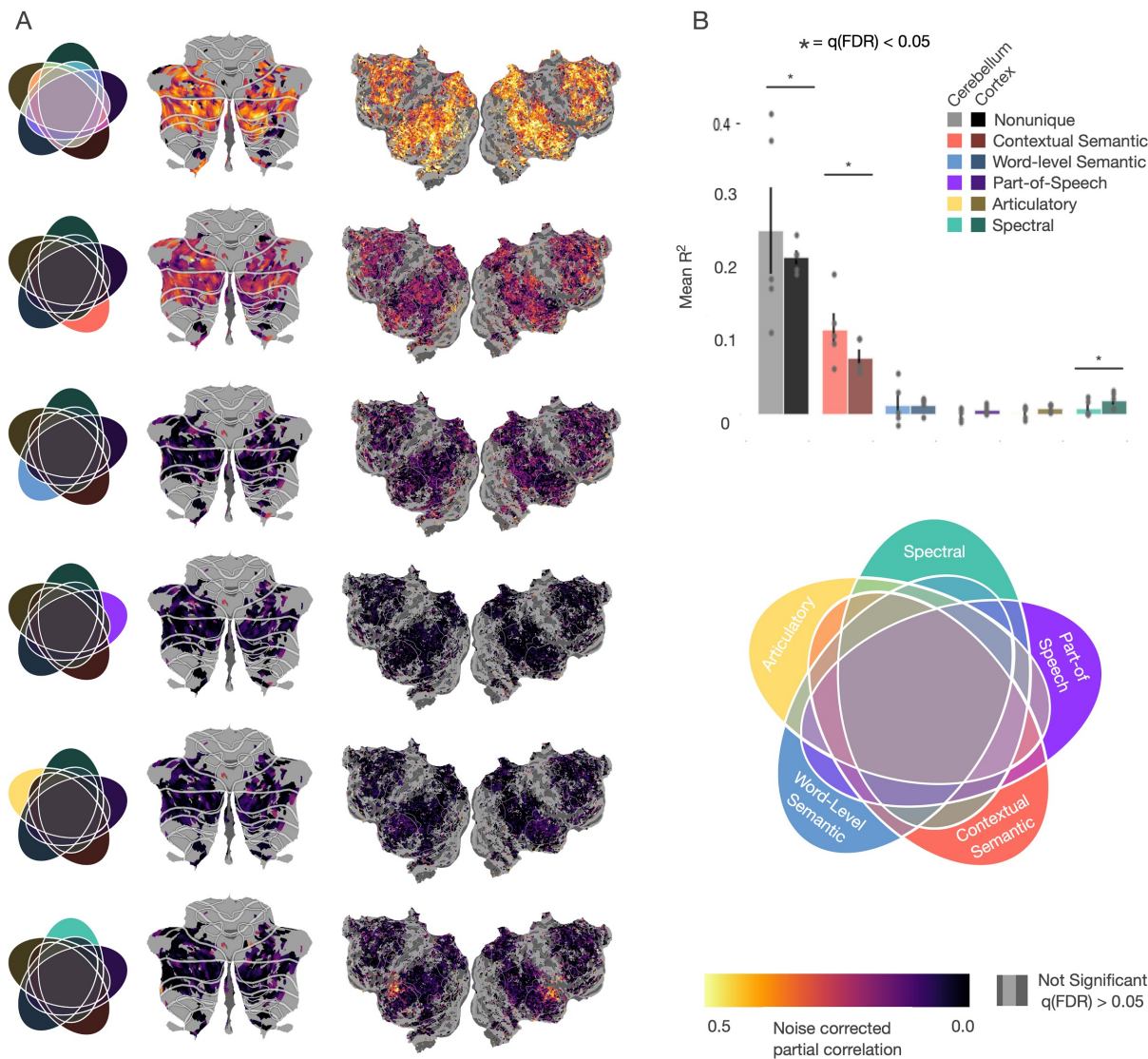
Supplemental Figure 4. Unique variance explained by each feature space adjusted for differences in signal-to-noise in cerebellum and cortex. To determine the unique variance explained by each feature space, six new encoding models were fit: a union encoding model containing a concatenation of all feature spaces, and five encoding models each containing a concatenation of four of the five feature spaces. The unique contribution of each feature space was then determined by subtracting the variance explained by the four-way concatenation model without that feature space from the union model. This shows how much variance can be explained by each feature space above and beyond the other four. Additionally, the amount of non-unique variance—i.e., that which can be explained by more than one feature space—was determined by subtracting the 5 unique variances

from the union. To account for differences in signal-to-noise between cerebellum and cortex, the correlations were corrected using standard noise-ceiling correction techniques. (**A**) The voxelwise partial correlation ($\sqrt{R^2}$) for each feature space for subject UT-S-02 was projected onto the cortical and cerebellar surfaces (Similar maps for additional subjects can be found in **Supplemental figure 5)**. Only voxels that were significantly predicted (one-sided permutation test, q(FDR)<0.05) by the 5-way union model are displayed. (**B**) Mean correlations for significant voxels in the cerebellum and cortex across all subjects. The non-unique partition explains the most variance in both cortex (darker) and cerebellum (lighter), although the variance explained by the non-unique partition is significantly (two-sided permutation test, q(FDR)<0.05) larger in cerebellum. The modality-specific spectral feature space explains significantly less variance in the cerebellum as compared to the cortex. Additionally, the modality-specific and language-specific feature spaces explain a negligible amount of variance in the cerebellum and the context-level semantic space explains the most variance among the unique partitions in cerebellum.This further supports the hypothesis that the cerebellum is largely representing language at a high, conceptual level.

959

Supplemental Figure 5. Unique variance explained for each feature space for additional subjects adjusted for differences in signal-to-noise in Cerebellum and Cortex. To determine the unique variance explained by each feature space, a union encoding model was fit with a concatenation of all feature spaces in addition to five other encoding models - each a concatenation of four of the feature spaces. The unique contribution of a model can be determined by the subtraction of the four way concatenation model without that feature space from the five way union model. Additionally, the amount

of overlap between the feature spaces can be characterized by the non unique partition. The unique variance explained by each feature space in each additional subject were projected onto the cortical and cerebellar surfaces. Only significant (one-sided permutation test, q(FDR)<0.05) voxels from the 5-way joint model are displayed.
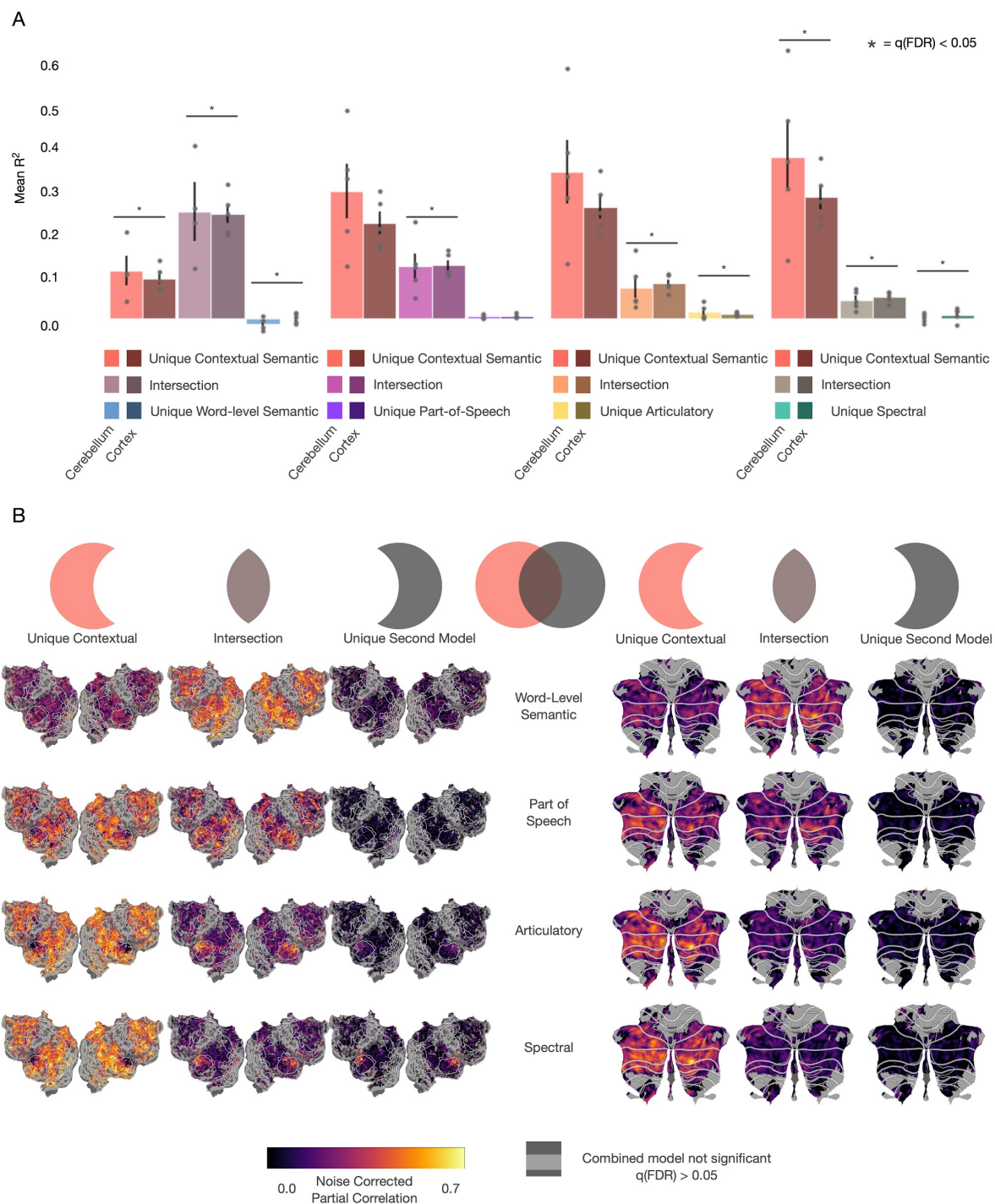
960
961

Supplemental Figure 6. Shared explained variance of the context-level semantic feature space with each of the other feature spaces for each additional subject. To quantify the amount of overlap between the context-level semantic feature space with each of the four other feature spaces, three models for each pair of feature spaces were fit which included the concatenated feature space and each feature space individually. For each pair of models, the variance explained by each partition in each voxel was projected onto the corresponding cortical flatmaps (**Supplemental Figure 9,** noise-ceiling corrected version to account for differences in signal-to-noise across the brain) Only voxels that were significantly predicted (one-sided permutation test, q(FDR)<0.05) by each union model are shown.There is substantially lower variance explained by the intersection between the context-level semantic model and the language and modality-specific feature spaces in the cerebellum than in cortex. The results are largely consistent across subjects.

962

Supplemental Figure 7. Shared explained variance of the context-level semantic feature space with each of the other feature spaces. To quantify the amount of overlap between the context-level semantic feature space with each of the four other feature spaces, three models for each pair of feature spaces were fit which included the concatenated feature space and each feature space individually. For each pair of models, the variance explained by each partition in each voxel was projected onto the corresponding cerebellar flatmaps (**Supplemental Figure 8,** noise-ceiling corrected version to account for differences in signal-to-noise across the brain) Only voxels that were significantly predicted (one-sided permutation test, q(FDR),0.05) by each union model are shown.There is substantially lower variance explained by the intersection between the context-level semantic model and the language and modality-specific feature spaces in the cerebellum than in cortex. Additionally the unique contributions for these models in the cerebellum is approaching zero and is not spatially localized. This lack of spatial localization further supports that there is no hierarchy of language processing in the cerebellum and these results provide strong support for the hypothesis that the cerebellum only represents high level, conceptual features of language, rather than low level features. This pattern of results appears consistent across subjects

963

Supplemental Figure 8. Shared explained variance of the context-level semantic feature space with each of the other feature spaces after correcting for signal-to-noise differences. To quantify the amount of overlap between the context-level semantic feature space with each of the four other feature spaces, three models for each pair of feature spaces were fit which included the concatenated feature space and each feature space individually. (**A**) For each pair models, the variance uniquely explained by the context-level feature space, uniquely explained by the second feature space, and the

intersection between the two for all subjects is compared between the cerebellum and cortex. The unique context-level partition is larger in the cerebellum than in the cortex for all feature spaces (two-sided permutation test, q(FDR)<0.05). Additionally, the unique partition for each second feature space is significantly smaller in cerebellum than in cortex for every space. This shows that the lower level feature spaces predict less unique variance in the cerebellum than cortex and further supports the hypothesis that the cerebellum is not representing modality-specific or language-specific information. (**B**) For each pair of models, the variance explained by each partition in each voxel was projected onto the corresponding cortical and cerebellar flatmaps. These results are noise-ceiling corrected to account for differences in signal-to-noise ratios across the brain. Only voxels that were significantly predicted (one-sided permutation test, q(FDR)<0.05) by each union model are shown.There is substantially lower variance explained by the intersection between the context-level semantic model and the language and modality-specific feature spaces in the cerebellum than in cortex. Additionally the unique contributions for these models in the cerebellum is approaching zero and is not spatially localized. This lack of spatial localization further supports that there is no hierarchy of language processing in the cerebellum and these results provide strong support for the hypothesis that the cerebellum only represents high level, conceptual features of language, rather than low level features.
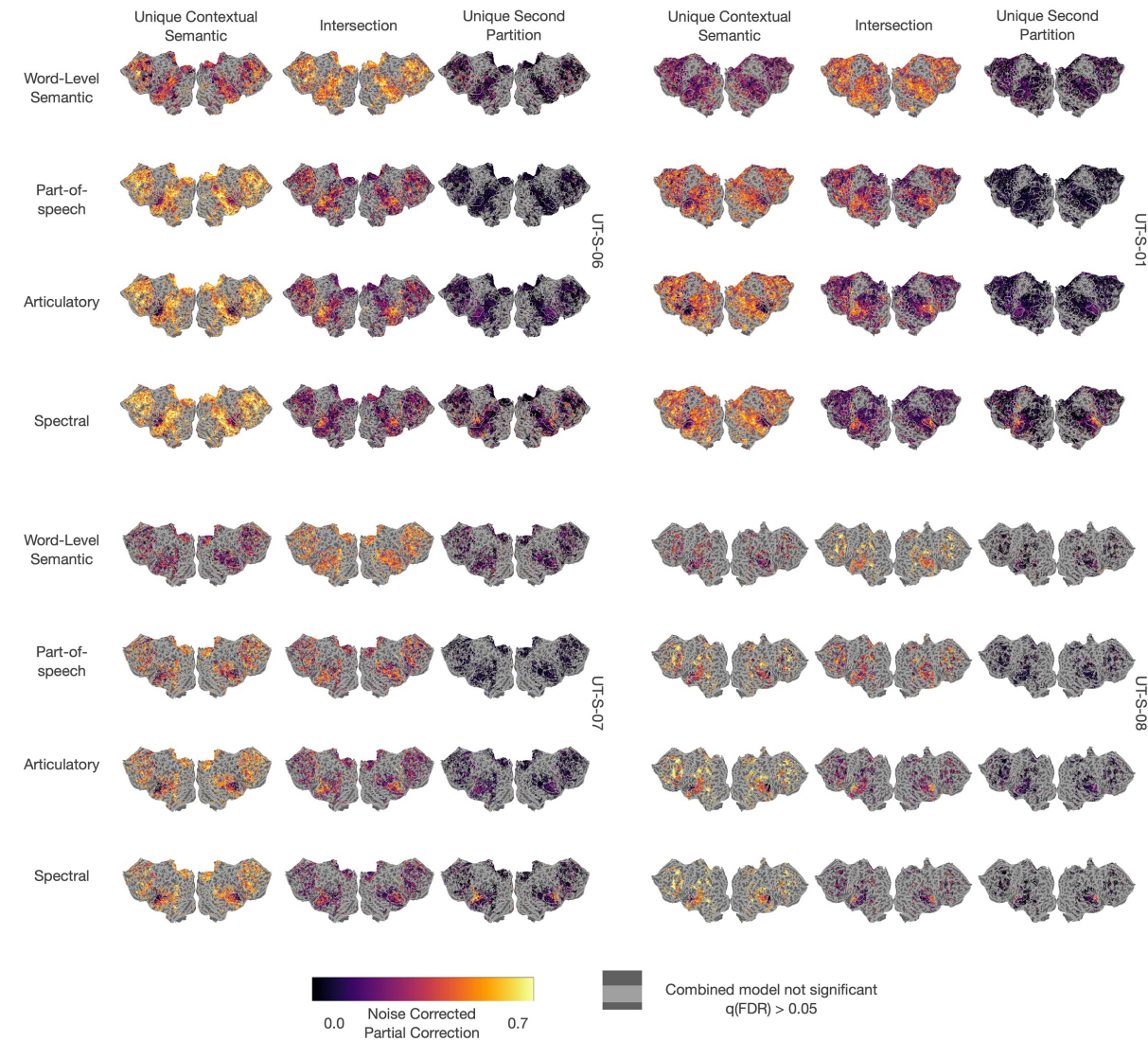
964
965

Supplemental Figure 9. Shared explained variance of the context-level semantic feature space with each of the other feature spaces corrected for differences in signal-to-noise in cerebellum. To quantify the amount of overlap between the context-level semantic feature space with each of the four other feature spaces, three models for each pair of feature spaces were fit which included the concatenated feature space and each feature space individually. Correlations were corrected using standard noise-ceiling correction techniques to account for differences in signal-to-noise in cerebellum and cortex. For each pair of models, the variance explained by each partition in each voxel was projected onto the corresponding cerebellar flatmaps. Only voxels that were significantly predicted (one-sided permutation test, q(FDR)<0.05) by each union model are shown.There is substantially lower variance explained by the intersection between the context-level semantic model and the language and modality-specific feature spaces in the cerebellum than in cortex. Additionally the unique contributions for these models in the cerebellum is approaching zero and is not spatially localized. This lack of spatial localization further supports that there is no hierarchy of language processing in the cerebellum and these results provide strong support for the hypothesis that the cerebellum only represents high level, conceptual features of language, rather than low level features.

966
967
968

Supplemental Figure 10. Shared explained variance of the context-level semantic feature space with each of the other feature spaces corrected for differences in signal-to-noise in cortex. To quantify the amount of overlap between the context-level semantic feature space with each of the four other feature spaces, three models for each pair of feature spaces were fit which included the concatenated feature space and each feature space individually. For each pair of models, the variance explained by each partition in each voxel was projected onto the corresponding cortical flatmaps. Correlations were corrected using standard noise-ceiling correction techniques to account for variance in the signal-to-noise ration. Only voxels that were significantly predicted (one-sided permutation test, q(FDR)<0.05) by each union model are shown.

969
970
971
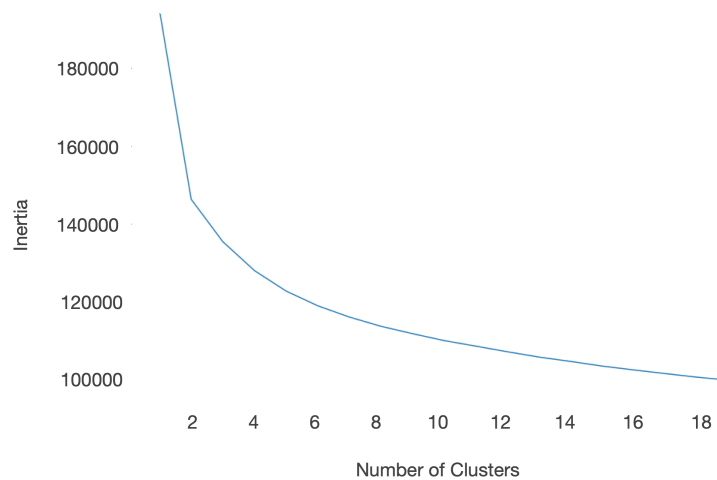
Supplemental Figure 11. Inertia of spherical K-means clustering with increasing number of clusters. To determine the ideal number of clusters to use when clustering the model weights of word-level semantic models, we calculated the inertia at each cluster amount from 1 to 20. There is a clear "elbow point" where the slope of the inertia changes from exponential to linear at 5 clusters, thus that is what we chose to use. However, we did not find any differences in the results using a different number of clusters.

972
973
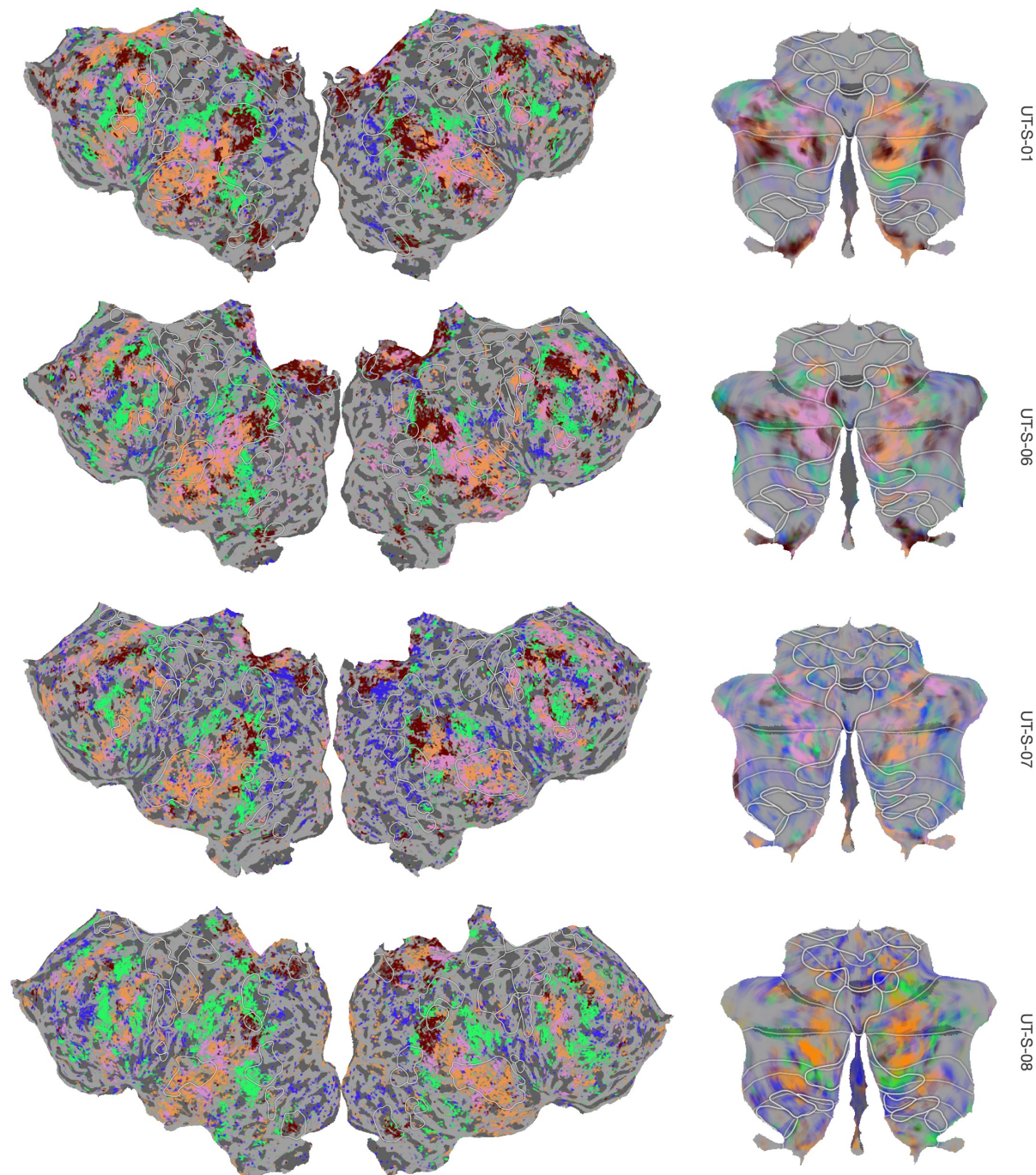
Supplemental Figure 12. Semantic Clustering of Model Weights. To check for differences in semantic representations between the cerebellum and cortex, word-level encoding model weights from both cerebellum and cortex in all subjects were concatenated, including only the top 20% best-predicted voxels. This matrix was then clustered using spherical k-means into 5 clusters, which fell at the inflection point in the inertia graph (**Supplemental Figure 11**). For visualization, the centroid for each cluster was transformed into the same RGB space used in **Figure 5**, and each voxel in that cluster was assigned that color. The cluster distribution for each subject across the cerebellum and cortex are shown.

974