Supporting Information for:

Engaging in word recognition elicits highly specific modulations in visual cortex

by Alex L. White, Kendrick Kay, Kenny Tang, and Jason D. Yeatman

Email: alwhite@barnard.edu

This PDF file includes:

Supporting text

Figures S1 to S9

Supporting Text: Methods Information

Choice of words for each trial of the main experiment:

For each of the two run types (attend-stimuli, attend-fixation), we generated 11 sets of 12 letter strings (6 real and 6 pseudo-words), one set for each of the 11 visual field positions. Within each task, the word sets were balanced in terms of the average of four different metrics: log frequency (maximum spread = 0.1), orthographic neighborhood size (maximum spread = 3.5 for real words, 0.2 for pseudowords); mean RT from the English Lexicon Project (ELP) database (1) (maximum spread = 40 ms); and mean ELP accuracy (maximum spread = 5% correct). Having created 22 such lists of letter strings (two tasks x 11 positions), we then generated 5 unique word-to-position assignments by randomly shuffling those 22 lists five times. Each subject was given a random one of those 5 assignments. The use of each stimulus list in the lexical decision or fixation task was also randomized across subjects. It was therefore rare for any two subjects to see the same word at the same location in the same task.

Localizer Experiment:

In order to localize word- and face-selective ROIs, participants completed a separate localizer experiment in their first scan session. Participants viewed sequences of images from 4 different categories: faces, objects, letter strings, and false fonts (**Figure 1C**). The letter strings were 5-letter real words, pseudowords, and consonant strings, each in two different fonts (Courier New and Sloan). For the false fonts, we used two different false characters matched in low-level visual features to the real fonts: a pseudo-Courier (2), and a pseudo-Sloan (3).

Each 4-second trial was composed of 4 frames presented in rapid sequence (700 ms each with 300 ms blanks between). Each frame contained 3 grayscale images: one small image at fixation (roughly 1.5 dva wide), and two large ones to either side (at 5.75 dva eccentricity, roughly 8 dva wide; see Fig 1A). A fixation dot 0.11 dva in diameter was superimposed on the center image. All frames on each trial contained images from the same category (e.g. 4 frames of faces).

The screen's background was set to a medium gray (63% of its maximum white). Letters and false font characters were black. The face and object images were normalized such

that pixel intensities spanned the full range [0 255], with mean equal to the background gray. The real words were all low in lexical frequency (<10 occurrences per million). The pseudowords were generated to be pronounceable with constrained bigram statistics (4).

In separate runs, the participant performed a fixation color task (press a button when the fixation dot changed color) or one-back task (press a button when the stimulus images repeat). Those "target events" (fixation color change or successive image repetition) occurred on a random 33% of trials (each set independently of the other). Each participant completed 4 localizer runs, alternating between the two tasks. To define ROIs, we analyzed localizer data from both tasks together.

Retinotopy experiment:
Each participant also completed three 5-minute runs of a standard retinotopy experiment in which the participant fixated a central spot while viewing a bar that moved across the visual field (5). The bar moved in 8 different directions, taking 32 s to cross the screen each time. The bar contained high-contrast patterns including faces and words, which changed 5 times per second. The participant's task was to maintain central fixation and press a button whenever the fixation dot changed color.

To analyze these data, we used the analyzePRF toolbox to estimate the population receptive field (pRF) of each surface vertex (6, 7). By visualizing the eccentricity and polar angle maps, we located borders between retinotopic areas (8). To analyze responses during the main experiment in these retinotopic areas, we selected subsets of voxels corresponding to each stimulus position. Each vertex was assigned to one of the positions if: the PRF model fit $R^2 > 0.2$; the horizontal coordinate of its PRF center was within 1.5 dva; and distance of its vertical coordinate from the horizontal meridian was less than 1 SD of the PRF size. The response to each stimulus was then averaged only across vertices assigned to that stimulus position. Stimuli at +/-9 deg were excluded from analysis of retinotopic areas, because the displays used for retinotopic mapping did not extend out that far.

fMRI preprocessing:

We used *fMRIPrep* 20.2.1 (9, 10), which is based on *Nipype* 1.5.1 (11, 12). T1-weighted (T1w) images were corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (13), distributed with ANTs 2.3.3 (14). The T1w-reference was then skull-stripped with a *Nipype* implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 5.0.9) (15). A T1w-reference map was computed after registration of 2 T1w images (after INU-correction) using mri_robust_template (FreeSurfer 6.0.1) (16). Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1) (17), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (18).

For each participant' BOLD runs (across all tasks and sessions), the following preprocessing was performed: First, a reference volume and its skull-stripped version were generated by aligning a single-band reference scan. A B0-nonuniformity map (or *fieldmap*) was estimated based on two (or more) echo-planar imaging (EPI) references with opposing phase-encoding directions, with 3dQwarp (AFNI 20160207) (19). Based on the estimated susceptibility distortion, a corrected EPI (echo-planar imaging) reference was calculated for a more accurate co-registration with the anatomical reference. The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration (20). Co-registration was configured with six degrees of freedom. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL 5.0.9) (21). BOLD runs were slice-time corrected using 3dTshift from AFNI (19). Then, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. The BOLD time-series (including slice-timing correction) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. The BOLD time-series were also resampled onto the Freesurfer *fsnative* and *fsaverage*

surfaces. Framewise displacement (FD) was computed for each functional run, using an implementation in *Nipype*, following Jenkinson (21) and Power (22).

Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (23). Non-gridded (surface) resamplings were performed using mri_vol2surf (FreeSurfer). Many internal operations of *fMRIPrep* use *Nilearn* 0.6.2, mostly within the functional processing workflow.

Supporting Figures

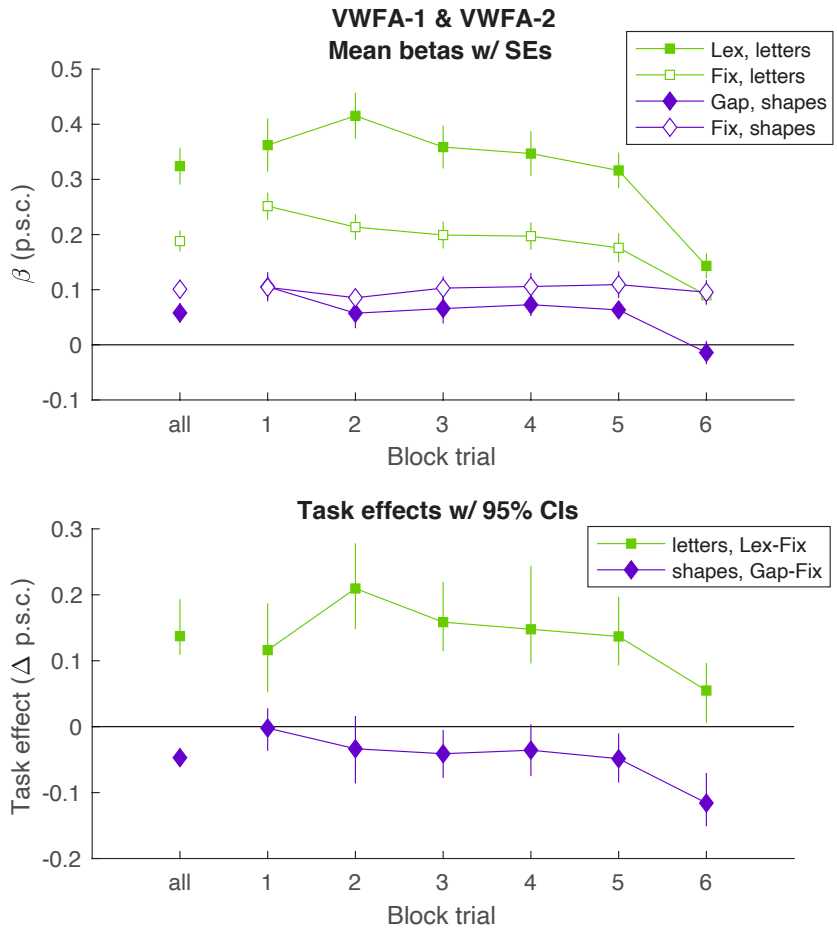

Figure S1: Mean responses as a function of trial number within the average block, in the union of left VWFA-1 and VWFA-2. During the experiment, trials came in blocks of 6 of all the same stimulus type. The stimulus type varied randomly from block to block. (The "all" condition on the left collapses across trial number within the block). The top panel shows the mean beta weights, collapsed across stimulus position, for each stimulus and task condition. Lex = lexical task, Gap=gap task, Fix=fixation task. Error bars are ± 1 SEM across subjects. The bottom panel shows the mean task effects for each stimulus type. Each point is the difference in the corresponding two points for that stimulus type shown in the top panel, with 95% bootstrapped confidence intervals to show significant differences from 0. The positive task effect for letters is significant at all trial numbers, but there was a significantly negative linear effect of trial number (p=0.02, 95% CI of slope = [-0.03 -0.002]). The task effect for shapes was essentially 0 on the 1st trial, then became more negative across trials (p=0.001, 95% CI of slope = [-0.03 -0.01]). Curiously, there was an overall trend for the absolute beta weights (response magnitudes) to decrease throughout the average block (average slope = -0.02 psc per trial, $p<10^{-14}$). That could be due to repetition suppression or adaptation to repeated stimulation.

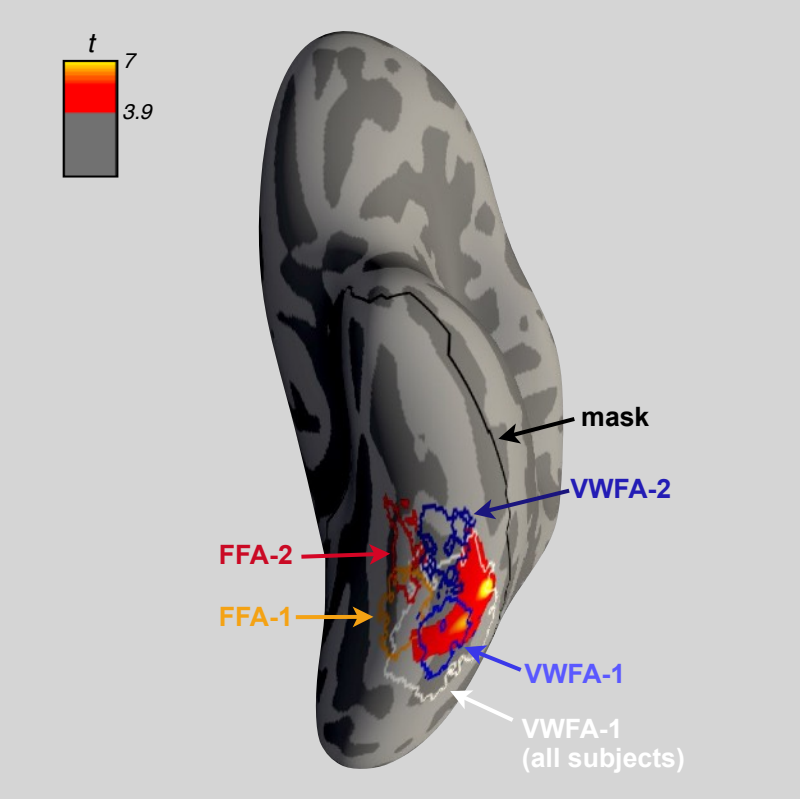

Figure S2: A map of the task-by-stimulus interaction on the left ventral cortical surface (fsaverage). The heatmap represents the t-statistic for interaction in BOLD responses: (Lexical task – Fixation task$_{letters}$) – (Gap task – Fixation task$_{shapes}$). The colormap ranges from $3.9 \leq t \leq 7$, 3.9 being the minimum t-value that was significant after correcting for multiple comparisons. No significant vertices appeared in the right hemisphere.

These data were analyzed only within a masked region (outlined in black) that included all of ventral temporal cortex, as well as early visual cortex. Specifically, masked region was the union of these Freesurfer parcellations: fusiform, inferior temporal parahippocampal, entorhinal, lateral occipital, lingual, pericalcarine, and cuneus.

The white outline labeled "VWFA-1 (all subjects)" encompasses *all* individual subject VWFA-1 ROIs after transforming to fsaverage space. The other colored ROIs are the same as those in main text Figure 1D, enclosing vertices where at least half the participants had overlap.

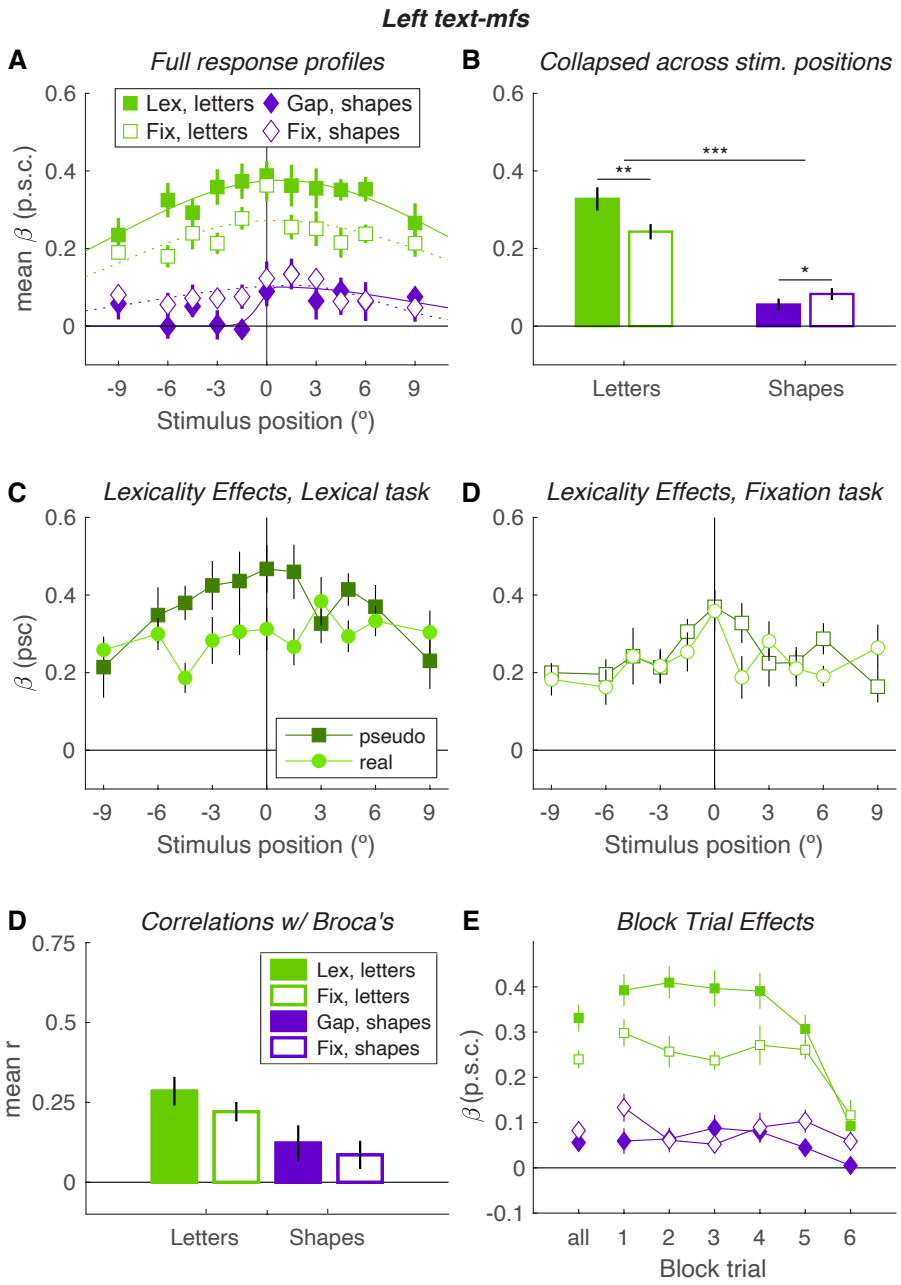

Figure S3: Data from the left "text-mfs" region, which we could localize in only 9/15 participants. Each panel corresponds to a plot of VWFA activity in the main text. **A:** According an LME model, there was a significant main effect of absolute eccentricity ($p<10^{-6}$), which was larger for letters than shapes ($p=0.007$). **B:** as indicated by the asterisks, there were task effects for each stimulus type, which went in opposite directions. **C:** Responses were overall higher for pseudo than real words ($F(1,210)=11$, $p=0.001$), and more so in the lexical task ($F(1,210)=3.9$, $p=0.049$). **D:** The trial-to-trial response variability was more strongly correlated with Broca's area when letters were presented than shapes ($p=0.008$), but was not affected by task. **E**: An LME model of response magnitudes revealed a main negative effect of trial ($F(1, 208)=46$, $p=10^{-10}$). The task effect for letters (lexical>fixation) significantly diminished across trials ($t(52)=3.63$, $p=0.001$), but the task effect for shapes was not affected ($t<1$, $p>0.5$). Unlike in the VWFAs, the gap<fixation effect was present on trial 1.

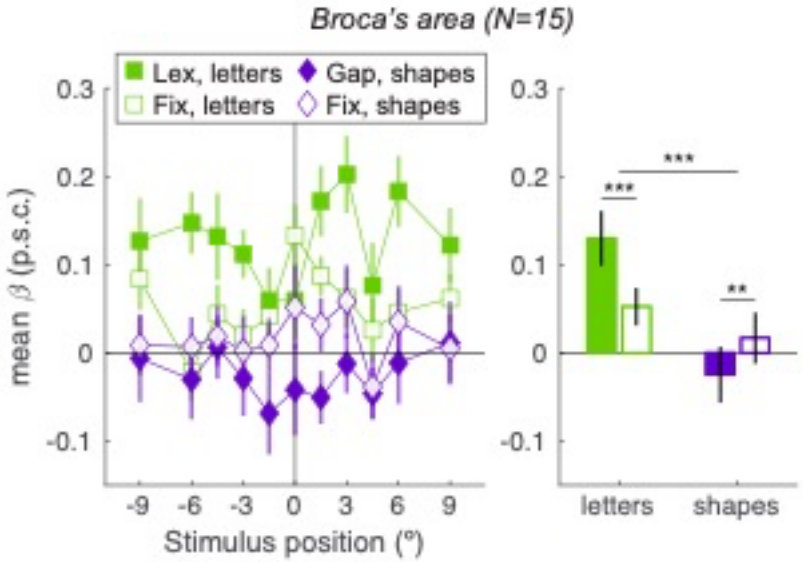

Figure S4: Mean BOLD responses in Broca's area in the left pre-central sulcus, analogous to main text Figure 2 for the VWFAs. The left panel shows the mean beta weights for each stimulus and task condition, for each visual field position. The right panel shows the same data but collapsed over stimulus position, with asterisks to show significant task effects for each stimulus type, and a significant interaction (*** indicates $p<0.001$). Note that the task effect for letters (lexical>fixation enhancement) is smallest, or even reversed, for stimuli at the fovea (0°), but larger in the periphery. In contrast, the task effect for shapes (gap<fixation suppression) is largest near the fovea, where shapes actually induce a drop in the BOLD signal below baseline.

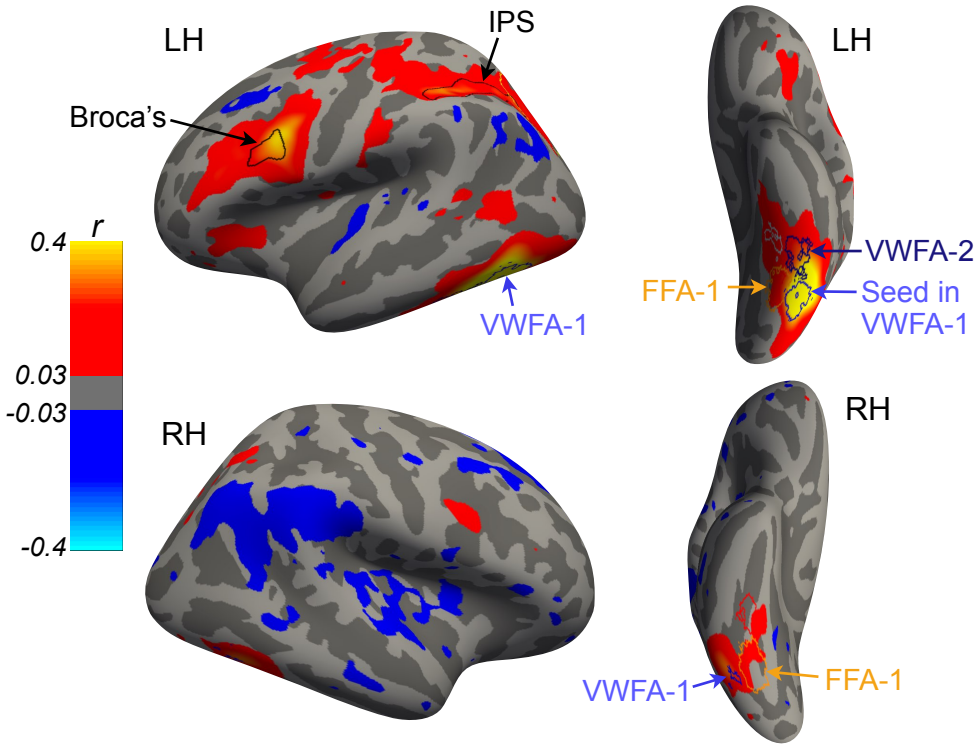

Figure S5: Across-subject mean correlation coefficients, on the fsaverage cortical surface, in trial-to-trial variability between each vertex and left VWFA-1 (the "seed" region here, unlike in Fig. 5 which had Broca's as the seed). Before averaging, each subject's data were smoothed with a 2D Guassian kernel (full-width at half-maximum = 5 mm). The data are masked to show only vertices where the correlation was significant (p<0.05, corrected for false discovery rate), peaking at r>=0.4.

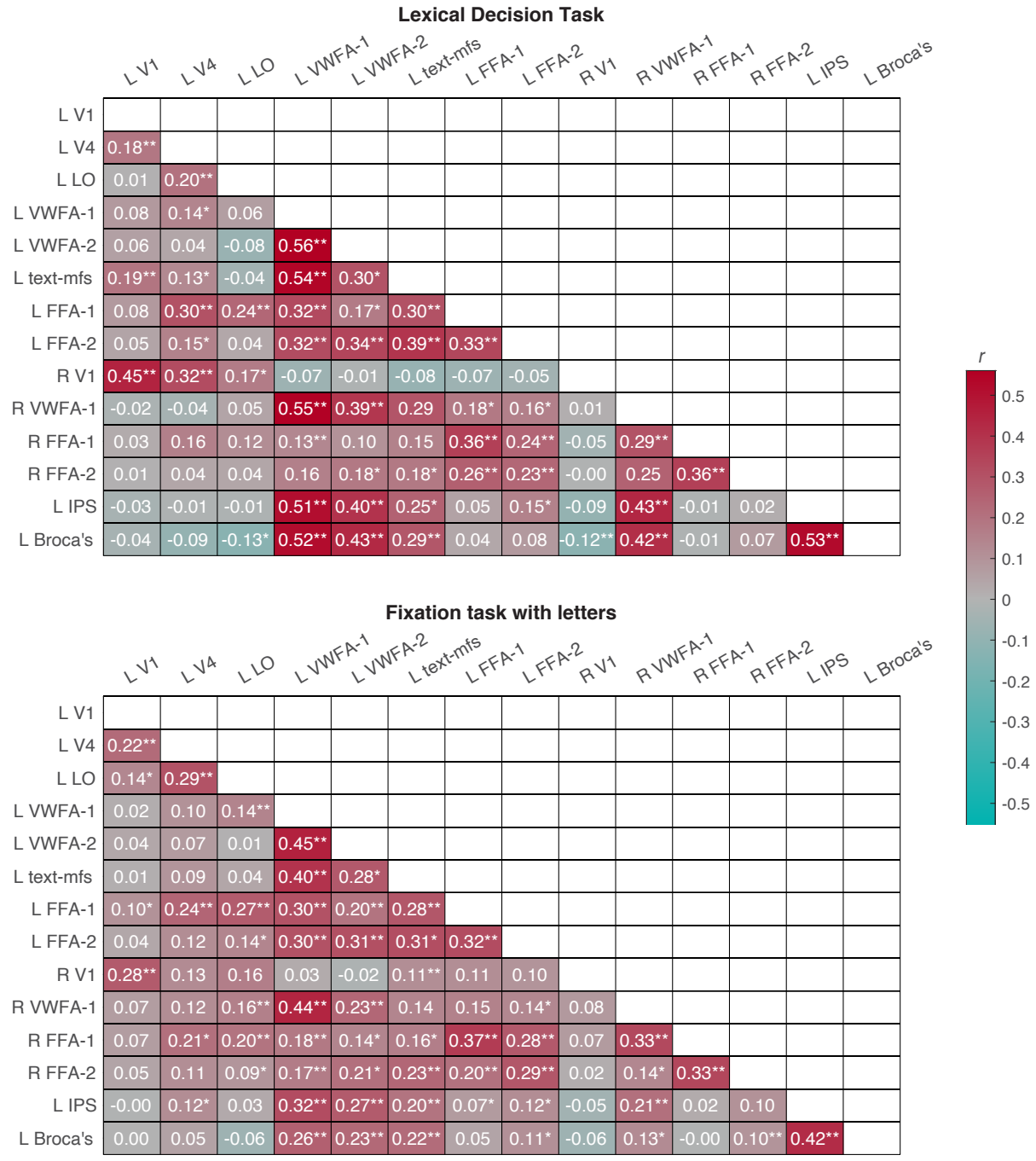

Figure S6: Correlations between trial-to-trial response variability in pairs of 14 regions, for trials when letters were presented in the lexical decision task (top) and fixation task (bottom). ** p<0.01 for a t-test comparing the mean correlation coefficient to 0. * p<0.05.

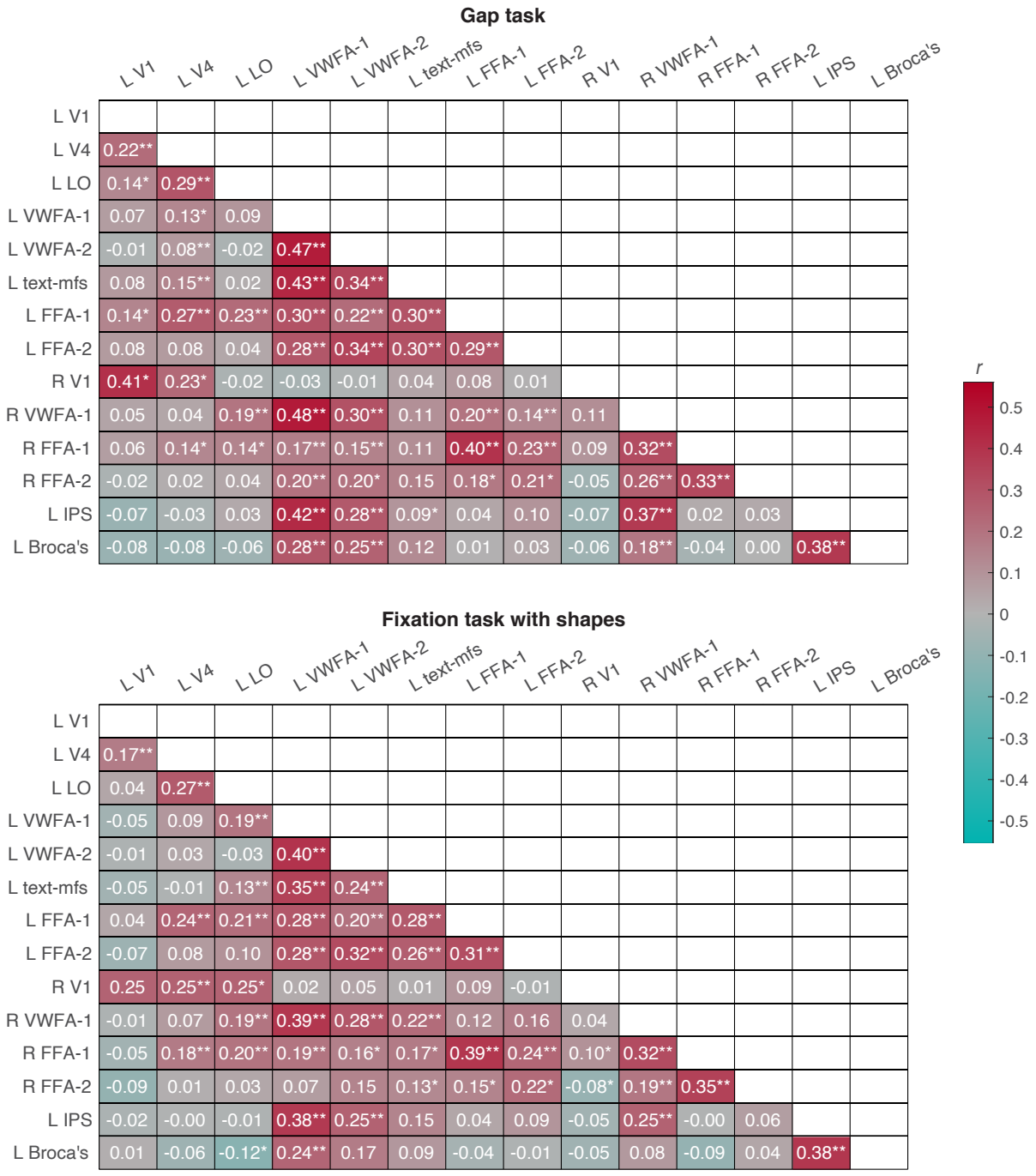

Figure S7: Correlations between trial-to-trial response variability in pairs of 13 regions, for trials when *shapes* were presented in the gap task (top) and fixation task (bottom). '**' indicates p<0.01 for a t-test comparing the mean correlation coefficient to 0. '*' indicates p<0.05.

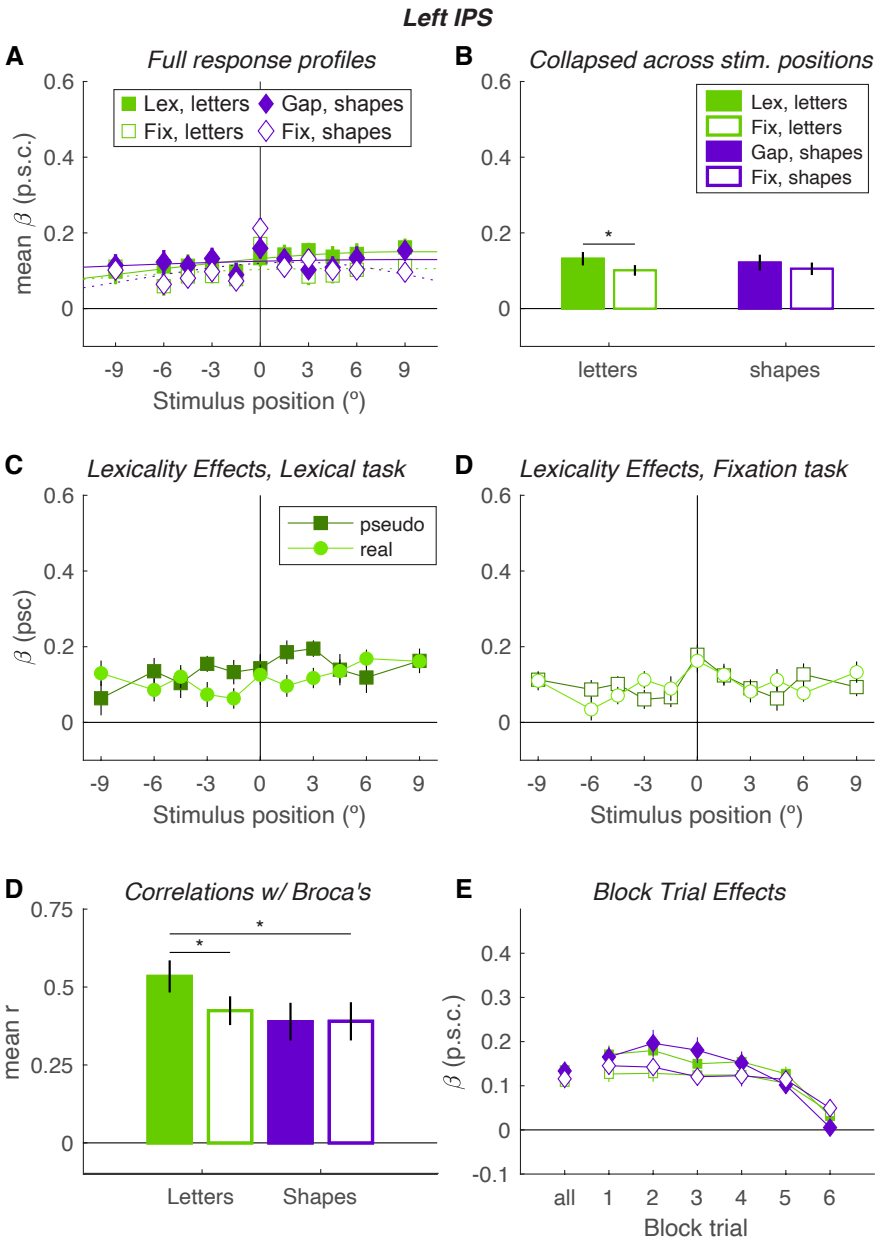

Figure S8: Data from the left IPS region, defined in fsaverage space where the correlation with Broca's area r>0.2 (see Figure 5A). Each panel corresponds to a plot of VWFA activity in the main text. **A:** Although these spatial profiles appear relatively flat, there was a significant main effect of absolute eccentricity (p=0.02), which was larger for letters than shapes (p=0.03). **B:** There was a main effect of task: "attend-stimuli" > "attend-fixation" (p=0.003), which did not interact with stimulus type (p=0.31, BF=0.42). **C:** There was no significant effect of lexicality in either task (ps>0.2). **D:** The functional connectivity with Broca's area was stronger when letters were presented than shapes (p=0.003). That effect interacted with task (p=0.04, BF=1.6), with a task effect (lexical > fixation) only for letters (p=0.01, BF=4.2). **E**: As in other areas, the overall BOLD response decreased across trials in each block ($p=3x10^{-29}$), and the task effects diminished as well (p=0.002). We also analyzed data in the a more posterior region that encompassed IPS-0, IPS-1 and IPS-2, defined from the Wang et al (2015) atlas. Its activity was very similar to what is pictured here, except with a stronger task effect (gap>fixation) in the overall BOLD response to shapes.

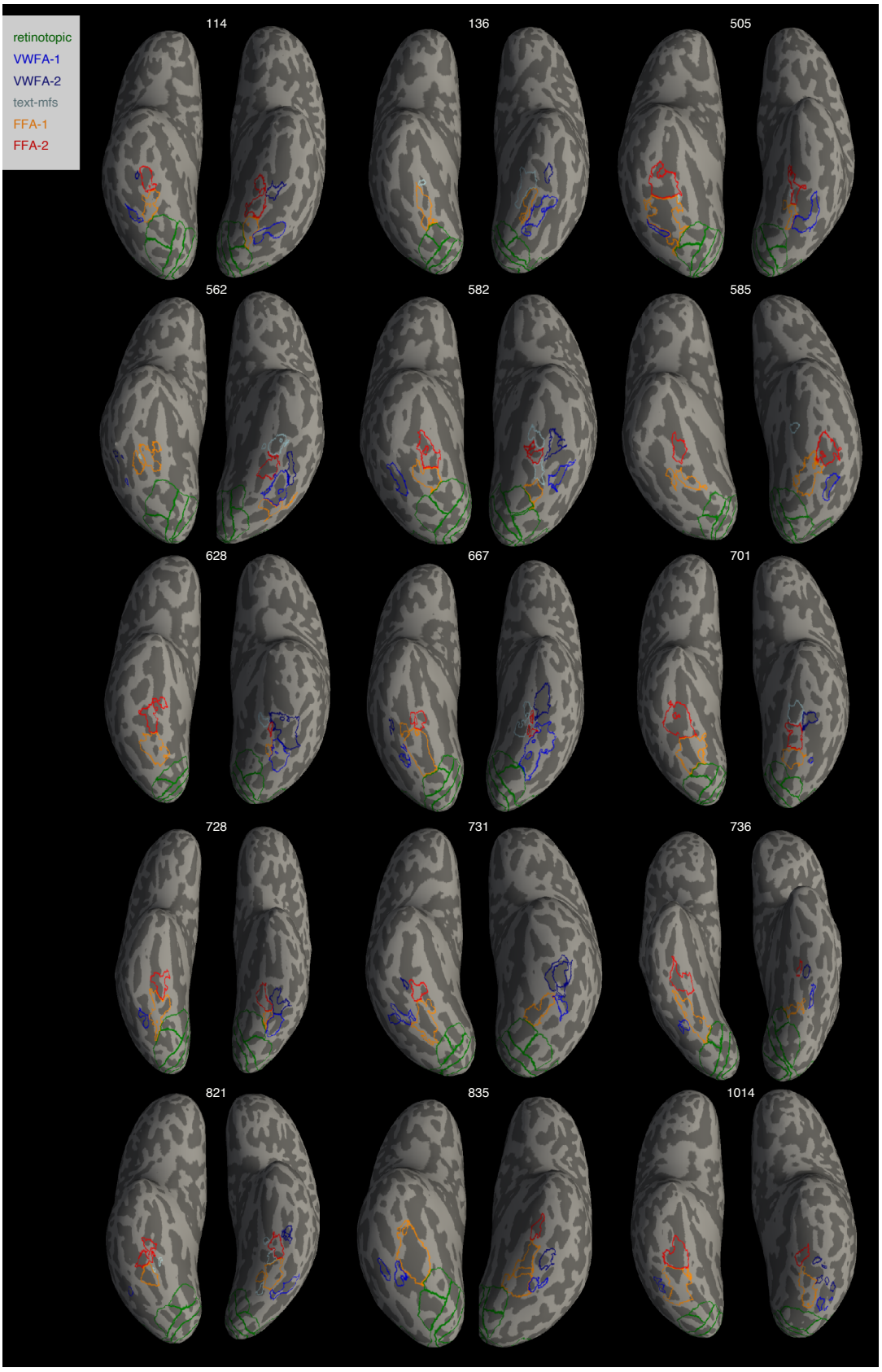

Figure S9: Visualization of ROIs on the ventral cortical surfaces for all 15 individuals.

SI References

1. D. A. Balota, *et al.*, The English Lexicon Project. *Behav. Res. Methods* **39**, 445–459 (2007).
2. C. Vidal, A. Content, F. Chetail, BACS: The Brussels Artificial Character Sets for studies in cognitive psychology and neuroscience. *Behav. Res. Methods* **49**, 2093–2112 (2017).
3. V. Vildavski, L. Lo Verde, G. Blumberg, J. Parsey, A. Norcia, PseudoSloan: A perimetric-complexity and area-controlled font for vision and reading research. *J. Vis.* **21**, 2857 (2021).
4. D. A. Medler, J. R. Binder, MCWord: An on-Line orthographic database of the English language. (2005).
5. N. C. Benson, *et al.*, The Human Connectome Project 7 Tesla retinotopy dataset: Description and population receptive field analysis. *J. Vis.* **18**, 1–22 (2018).
6. O. Esteban, *et al.*, FMRIPrep. *Software* **Zenodo**, 852659 (2018).
7. O. Esteban, *et al.*, fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).
8. K. Gorgolewski, *et al.*, Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Front. Neuroinform.* **5** (2011).
9. K. J. Gorgolewski, *et al.*, Nipype. *Software* **Zenodo** (2018).
10. N. J. Tustison, *et al.*, N4ITK: Improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**, 1310–1320 (2010).
11. B. B. Avants, C. L. Epstein, M. Grossman, J. C. Gee, Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* **12**, 26–41 (2008).
12. Y. Zhang, M. Brady, S. Smith, Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* **20**, 45–57 (2001).
13. M. Reuter, H. D. Rosas, B. Fischl, Highly accurate inverse consistent registration: A robust approach. *Neuroimage* **53**, 1181–1196 (2010).
14. A. M. Dale, B. Fischl, M. I. Sereno, Cortical Surface-Based Analysis. *Neuroimage* **9**, 179–194 (1999).
15. A. Klein, *et al.*, *Mindboggling morphometry of human brains* (2017).
16. R. W. Cox, J. S. Hyde, Software Tools for Analysis and Visualization of FMRI Data NMR in Biomedicine, in press. *NMR Biomed* **10**, 171–8 (1997).
17. D. N. Greve, B. Fischl, Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* **48**, 63–72 (2009).
18. M. Jenkinson, P. Bannister, M. Brady, S. Smith, Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *Neuroimage* **17**, 825–841 (2002).
19. J. D. Power, *et al.*, Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* **84**, 320–341 (2014).
20. C. Lanczos, Evaluation of Noisy Data. *J. Soc. Ind. Appl. Math. Ser. B Numer. Anal.* **1**, 76–85 (1964).