PNAS
www.pnas.org

Supplementary Information for

Parallel spatial channels converge at a bottleneck in anterior word-selective cortex

Alex L. White*[1,2,3], John Palmer[3], Geoffrey M. Boynton[3], and Jason D. Yeatman[1,2]

[1]Institute for Learning & Brain Sciences
[2]Department of Speech & Hearing Sciences
[3]Department of Psychology
University of Washington, Seattle USA

*Corresponding author:
Alex L. White
alexlw@uw.edu

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1822137116

Supplementary Methods

Power analysis to determine sample size

The sample size was chosen in advance of data collection on the basis of a power analysis of a previous fMRI study (White, Runeson, Palmer, Ernst, & Boynton, 2017). That study reported a mean selective attention effect of 0.1% signal change in retinotopic cortex. We simulated resampling those data to determine the number of participants required to detect an effect half as large with the same degree of noise. Thirteen participants was the minimum required to reach 80% power. We rounded that number up to 15.

Equipment and displays

During behavioral training, we presented stimuli with an Apple Mac Mini and a linearized CRT monitor with a 120 Hz refresh rate (1024 x 640 pixels). During MRI scanning, stimuli were generated with an Apple Macbook Pro and back-projected onto a fiberglass screen with a luminance linearized Eiki LCXL100 projector (60Hz; 1280 x 1024 pixels). The display background was set to the maximum luminance (90 cd/m^2 during training and 2350 cd/m^2 in the scanner).

In behavioral training, the participant used a standard keyboard to respond in the semantic categorization task. The keys used were: "z" and "x" for the left side; and "<" and ">" for the right side. In the scanner, the participant held a small button-box in each hand, each with two buttons on it.

Eye-tracking and fixation control

We recorded the right eye's gaze position with an Eyelink 1000 eye tracker (SR Research, Ottawa, Ontario, Canada). During behavioral training, we gave immediate feedback about fixation breaks. Each trial began only if the registered gaze position was within 0.75° horizontally and 3° vertically of the fixation mark. (We allowed more vertical tolerance to account for calibration drift and pupil size changes). We then averaged the gaze position over 10 samples to determine the initial fixation position. If, during the interval between the pre-cue offset and the post-mask offset, the estimated gaze position moved more than 1° horizontally or 2° vertically from the initial fixation position, the trial was immediately aborted. Text on the screen informed the participant that they had broken fixation and required a key-press to continue to the next trial. During MRI scanning, there was no such feedback about fixation breaks and trials continued at a constant pace. For one participant, technical errors prevented the recording of eye-tracker data during scanning, but their fixation control during training was excellent and they believed that their gaze position was still being monitored during scanning. For two other participants, eye-tracking failed on 1 and 3 scans, respectively (out of 10).

We detected fixation breaks in the recorded eye traces offline. For each scan, we defined the "central gaze position" as the median of all trials' median gaze positions, each computed during the window between pre-cue and post-cue onsets. We cut out periods with blinks, ± 50 ms. We defined a fixation break as a deviation >0.8° horizontally or >2.0° vertically that lasted more than 50ms and occurred between pre-mask onset and post-cue onset.

Participants were excluded if they had fixation breaks on 5% or more of trials (applied to 2 participants after behavioral training). In behavioral training and in scanning, fixation breaks were detected on 2% and 1% of trials, respectively. Trials with fixation breaks were excluded from analysis of behavioral performance, and entire blocks with 1 or more fixation breaks were excluded from the MRI analysis.

Procedure

Each participant completed 3 or 4 one-hour training sessions before scanning. These began with the TOWRE test and task instructions, and the participant read the full list of words. Then they practiced the task, with the word-mask ISIs initially set well above threshold. The ISI was gradually shortened until accuracy in the focal cue condition settled at roughly 80% correct (averaged over left and right sides). Then the participant completed at least 14 "runs" (each run containing 21 trials of each condition). The ISI was the same in focal and distributed cue conditions, and was adjusted from day to day as necessary to maintain ~80% correct in the focal cue condition.

Each participant then completed 3 MRI sessions. The first was for retinotopic mapping (see below). In each of the 2nd and 3rd sessions, the participant completed 3 localizer scans (L) and 5 main experimental scans (M), in a fixed order: L, M, M, M, L, M, M, L.

Retinotopy

Each participant participated in a retinotopic mapping session. In each of six 4.2-minute scans, we presented one of three periodic stimulus types: a contracting ring, a rotating wedge, or alternating vertical/horizontal bow ties. All stimuli were composed of sections of radial checkerboards counter-phase flickering at 8 Hz. During each 256-s scan, the stimulus made eight ''cycles'' (rings contracting from 11.8 to 0.48 radius; wedge rotating clockwise one full circle; bow ties presented vertically then horizontally). The participant fixated a central white dot and pressed a button any time the dot briefly darkened or the checkerboard briefly dimmed in contrast. Using standard methods (Engel, Glover, & Wandell, 1997), we analyzed rings and wedge scans to identify the phase of the stimulus cycle that each voxel preferred, providing eccentricity and polar angle maps, respectively. We located the voxels representing the horizontal and vertical meridians via a general linear model (GLM) contrast of responses to the horizontal and vertical bow-tie stimuli. Using these activity patterns, we drew borders between retinotopic regions on each inflated cortical hemisphere. With these borders, we defined sets of anatomical voxels belonging to each retinotopic region. In some participants VO1 was not clearly separable from VO2, and LO1 was not always separable from VO2. Therefore, we merged each pair of sub-regions (when there were two) into LO and VO.

MRI data acquisition

Using a Philips Ingenia 3T scanner, we acquired anatomical images with a standard T1-weighted gradient echo pulse sequence (1-mm resolution). We acquired functional images with an echo planar sequence, with a 32-channel high-resolution head coil, a repetition time of 2 s, and an echo time of 25 ms. Thirty-five axial slices (80 x 80 matrix, 240 x 240 x 105-mm field of view, 0 gap) were collected per volume (voxel size: 3 x 3 x 3 mm).

MRI pre-processing

Using BrainVoyager™ software, we first pre-processed each functional scan with: trilinear slice time correction; motion correction to the first volume of the first scan (trilinear detection and sinc interpolation); phase-encoding distortion correction, based on one volume collected in the opposite direction at the start of each session; and high-pass temporal filtering (cutoff: two cycles/scan). Each functional scan was co-registered with a high-resolution anatomical scan collected in the same session, which was itself co-registered with the anatomical scan from the retinotopy session.

Statistical analyses

When using linear mixed-effects models, we included random intercepts across participants. When justified by a likelihood ratio test, we also included random slopes across participants. To assess the significance of the pairwise differences, we used bootstrapping: we built a distribution of 5000 means of N values resampled with replacement from the original sample of the N participants' differences. The two-tailed p-value is twice the proportion of bootstrapped means less than 0. At a significance cutoff of $p=0.05$, this approach is equivalent to regarding a difference as significant if the 95% confidence interval (CI) of differences excludes 0.

Table S1: Stimulus set, Non-living category

vest	belt	chalk	penny	marble
sofa	shoe	flint	plate	carbon
glue	bath	villa	sword	blouse
gown	snow	dryer	wheel	velvet
coal	roof	scarf	knife	lounge
robe	coat	torch	clock	chapel
fork	moon	shack	pants	candle
sock	mask	cloth	slacks	canyon
mill	pipe	jeans	staple	nickel
coin	flag	lodge	mosque	cellar
silk	fuel	shelf	blazer	shorts
shed	soap	ridge	faucet	fridge
boot	iron	stove	pantry	pencil
dime	hinge	spoon	blinds	pillow
tire	whisk	attic	crater	carpet
pump	stair	bench	heater	wallet
lamp	clogs	skirt	bronze	closet
whip	shawl	drill	cradle	garage
barn	latex	frame	shrine	dollar
cave	satin	towel	drapes	toilet
fort	linen	motel	canvas	mitten
tube	cloak	steel	cement	sandal
sink	slate	cabin	bunker	castle
salt	manor	cable	dagger	
sand	stool	couch	saloon	

Table S2: Stimulus set, Living category

bear	moth	heron	squid	iguana
bird	mule	horse	stork	insect
bull	newt	hound	tiger	jackal
bush	orca	human	trout	jaguar
carp	pine	hyena	tulip	lizard
clam	rose	koala	woman	maggot
crab	seal	lemur	whale	minnow
crow	slug	lilac	zebra	monkey
deer	swan	llama	amoeba	orchid
dove	toad	maple	baboon	oyster
duck	tree	moose	badger	parrot
fern	tuna	mouse	beaver	pigeon
fish	vine	otter	beetle	possum
flea	wasp	panda	cactus	python
girl	wolf	plant	canary	rabbit
frog	worm	raven	clover	salmon
goat	algae	robin	cougar	spider
hare	bison	shark	coyote	spruce
hawk	camel	sheep	donkey	turkey
kelp	cobra	shrew	falcon	turtle
lamb	daisy	shrub	ferret	walrus
lily	eagle	skunk	fungus	weasel
lion	finch	sloth	gerbil	willow
mole	goose	snail	gopher	
moss	grass	snake	hornet	

Supplementary Results

Differences between VWFA-1 and VWFA-2 do not reflect SNR

The differences we found between left VWFA-1 and VWFA-2 could be due to lower signal-to-noise of our measurements in VWFA-2. Specifically, it is plausible that increased noise obscured the presence of two channels and attention effects in VWFA-2. To test this possibility, we examined the patterns of spatial selectivity in the localizer scan data. Under the one-channel model, voxels may respond more to words in the contralateral visual field on average (**Figure 1B**), but *differences* between the spatial preferences of individual voxels are just noise. In contrast, the two-channel model assumes that voxels within a region differ meaningfully in how much they prefer the left or right side (because the two words are processed in partially separable populations of neurons).

The two models make different predictions for the across-voxel correlation between responses to single words on the left (W_L) and on the right (W_R). All else being equal, the correlation should be *weaker* in a region that contains two channels than a region with only one channel, because of the variance added by the true differences in voxel preferences. The correlation should also be weaker in an area with more measurement noise in the BOLD response. So, if VWFA-2 has two channels that are obscured by additional measurement noise, its mean correlation coefficient between W_L and W_R must be *lower* than in VWFA-1. However, the opposite was true: mean r=0.72±0.09 in left VWFA-1 and r=0.81±0.04 in left VWFA-2. Moreover, as predicted by the additional heterogeneity in voxel preferences in a two-channel area, the average standard deviation of differences between W_L and W_R was greater in left VWFA-1 (0.27±0.04) than VWFA-2 (0.13±0.01). Therefore, measurement noise alone cannot account for the different pattern of results in VWFA-2. Our interpretation is that two spatial channels in VWFA-1 merge into a single channel in VWFA-2.

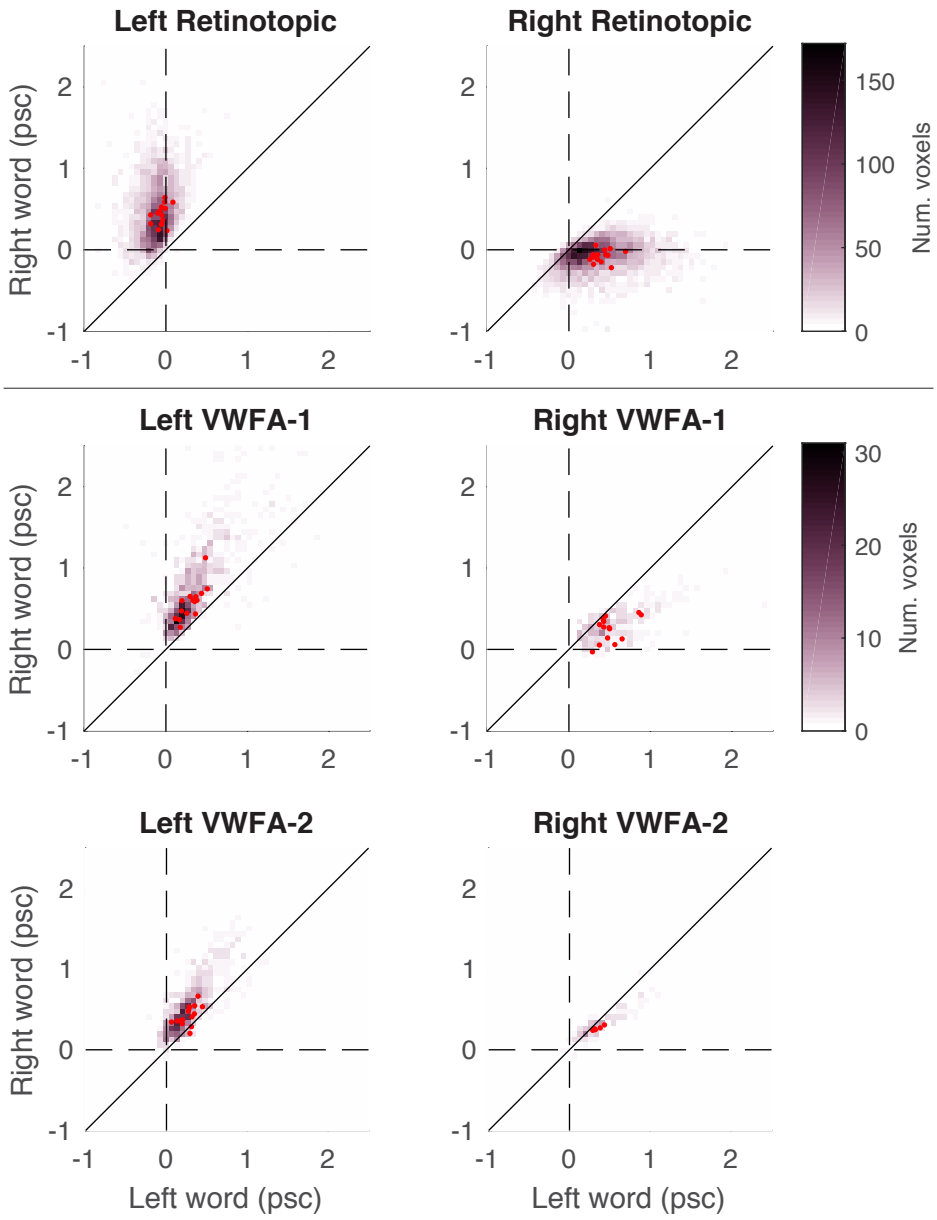

Figure S1: Voxel responses to words on the left and right of fixation in the localizer scans. The color of each point indicates the number of voxels with that combination of responses to words at the two locations. Red dots indicate the across-voxel median for individual participants. The top row contains the union of voxels from all retinotopic areas. The left column is for the left hemisphere, and right column for the right hemisphere.

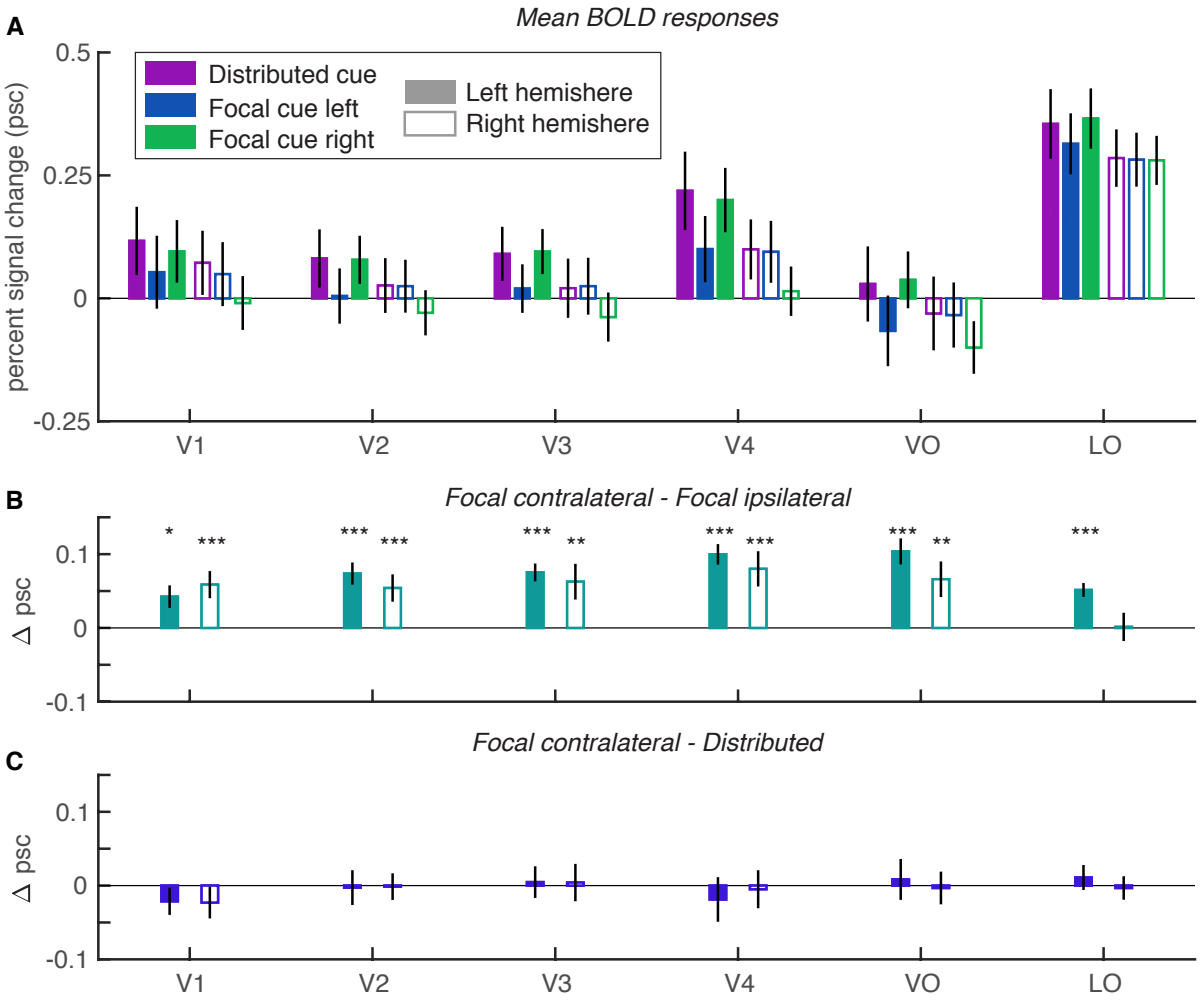

Figure S2: Mean BOLD responses and attention effects in retinotopic areas. (**A**) Mean BOLD responses and each ROI and hemisphere, divided by cue condition. (**B**) Mean selective attention effects: differences between responses when the contralateral vs. ipsilateral word was focally cued. (**C**) Mean divided attention effects: differences between responses when the contralateral word was focally cued vs. when both words were cued. All error bars = +/- 1 SEM. Asterisks indicate significant effects from bootstrapping: *** = p<0.001; ** = p<0.01; * = p<0.05.

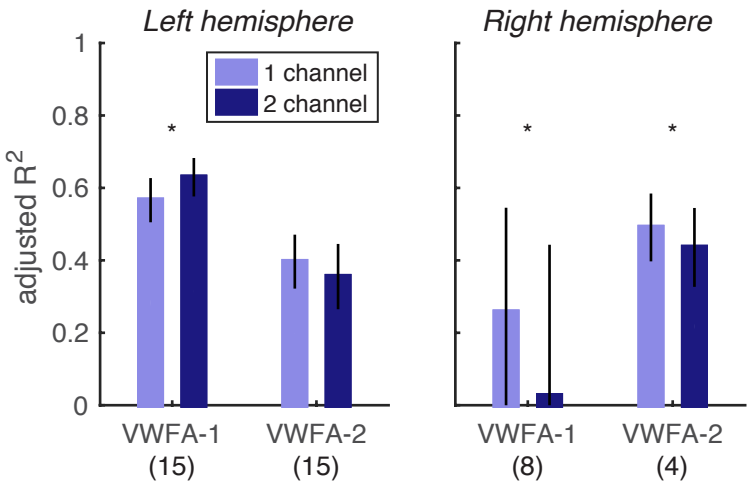

Figure S3: Adjusted r-squared values for the fit quality of the two-channel vs. the one-channel spatial encoding models. Asterisks indicate significant pairwise differences between the two models for each region ($p<0.05$ from bootstrapping). The numbers in parentheses at the bottom of the plot are the number of participants that had enough voxels to compute adjusted R^2 in each ROI.

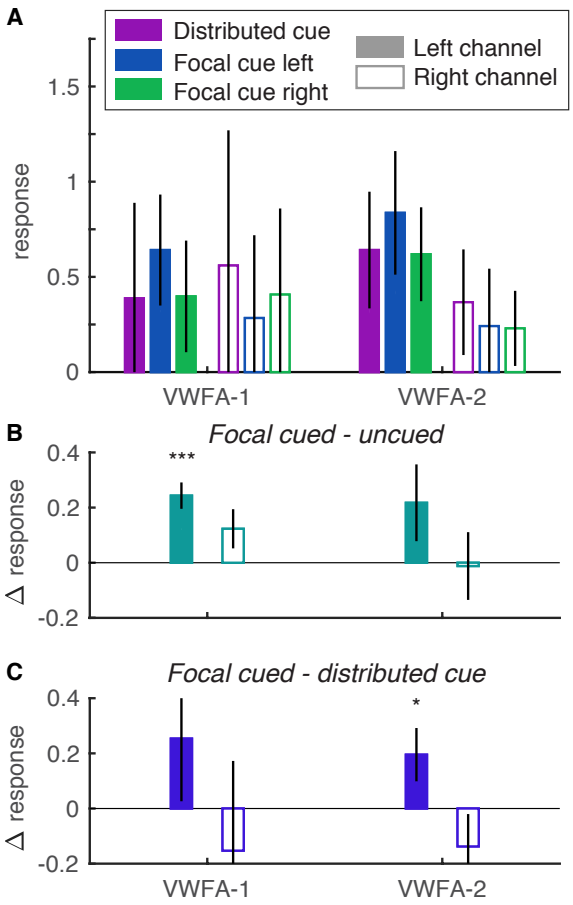

Figure S4: Estimated right hemisphere channel responses from the spatial encoding model. Error bars are ± 1 SEM. *** = p<0.001; ** = p<0.01; * = p<0.05.

(**A**) Mean responses, separately for each ROI, channel, and cue condition.

(**B**) Selective attention effects: the differences between each channel's responses when its visual field location was focally cued vs. uncued. In Right VWFA-1 (14/15 participants), there was a main effect of cue (mean = 0.18 ± 0.06; F(1,52)=12.0, p=0.001), no effect of channel (F(1,52)=0.06, p=0.80), but also an interaction (F(1,52)=5.94, p=0.018). The interaction indicates that the cue effect was bigger in the left channel (0.24 ± 0.05) than the right channel (0.12 ± 0.07), and only significant in the former. In Right VWFA-2 (5/15 participants), there was no effect of channel (F(1,16)=1.25, p=0.28), and no main effect of cue (F(1,16)=0.78, p=0.39), but an interaction (F(1,16)=44.8, $p<10^{-5}$). The selective attention effect was bigger in the left channel (0.22 ± 0.14) than the right channel (-0.01 ± 0.12), but not significant in either.

(**C**) Divided attention effects: the differences between each channel's response when its location was focally cued vs. when both sides were cued. No effects of cue (focal cued vs. distributed) or channel (left vs. right) or interactions were significant, except for an interaction in right VWFA-2 (F(1,16)=5.55, p=0.031). The left channel was reduced by dividing attention (0.2 ± 0.10) but the right channel had the opposite effect (-0.14 ± 0.12). Note that 14/15 participants had a right VWFA-1 and only 5 had a right VWFA-2, and the one-channel model fit significantly better than the two-channel model in both regions.

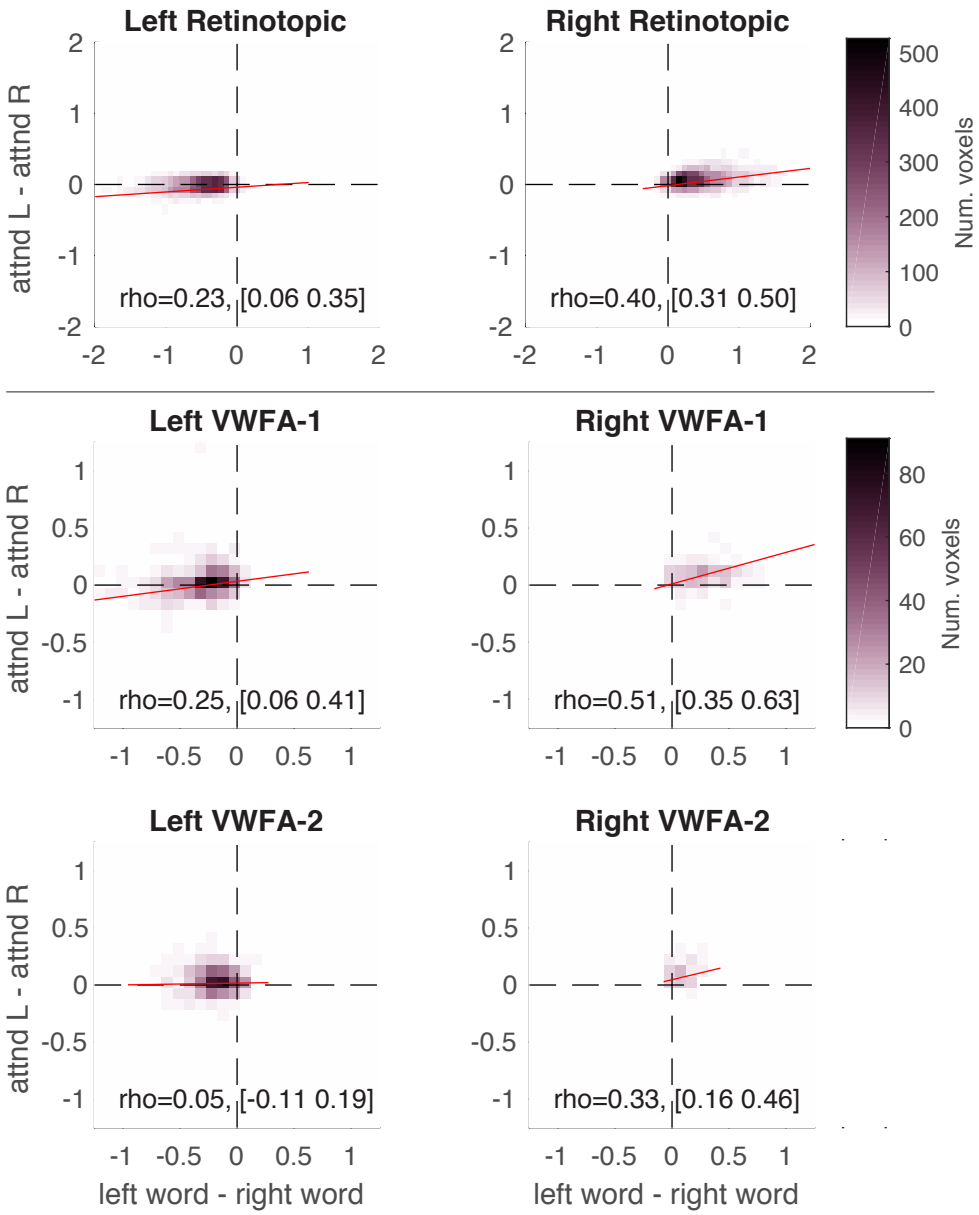

Figure S5: Voxel spatial vs. attentional selectivity. The x-axis is the difference in voxel responses between single words on the left and single words on the right in the localizer scans (spatial selectivity). The y-axis is the difference in voxel responses between the focal cue left and right conditions of the main experiment (selective attention effect). The color of each point indicates the number of voxels with each combination of those two differences. The text in each panel reports the across-participant mean correlation coefficient (rho) and it's 95% confidence interval. The red line is the best-fitting linear mixed-effects model that included random effects of slope and intercept across participants. The top row contains the union of voxels from all retinotopic areas. The left column is for the left hemisphere, and right column for the right hemisphere.

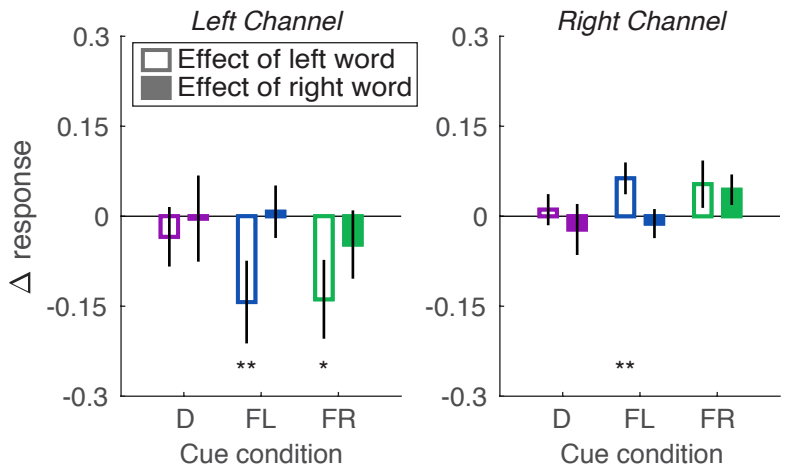

Figure S6: **Effects of lexical frequency on channel responses in left hemisphere VWFA-1.** The estimated left channel responses are plotted in the left panel, and the right channel in the right panel. In each panel, the y-axis is the mean difference between trials when the word's lexical frequency was above vs. below the median. Unfilled bars are the effect of the left word's frequency (averaging over the right word); filled bars are the effect of the right word's frequency (averaging over the left word). Bars are grouped along the x-axis by cue condition: D=distributed; FL=focal left; FR=focal right. Error bars are ± 1 SEM. Asterisks indicate two-tailed p-values computed from bootstrapping: ** = $p<0.01$; * = $p<0.05$. According to a linear mixed effects model, the left channel showed an overall effect of frequency ($F(1,84) = 6.34$, $p=0.014$) that was marginally higher for the left than right word ($F(1,84) = 3.64$, $p=0.060$) and showed no effect of cue or interaction ($F(2,84)<1$). In the right channel of left VWFA-1, the frequency effect was inverted (high>low), but not significant ($F(1,84)=3.38$, $p=0.07$), and not modulated by side or cue (all $p>0.10$).

References

Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex*, *7*, 181–192.

White, A. L., Runeson, E., Palmer, J., Ernst, Z. R., & Boynton, G. M. (2017). Evidence for unlimited capacity processing of simple features in visual cortex. *Journal of Vision*, *17(6):19*, 1–20.