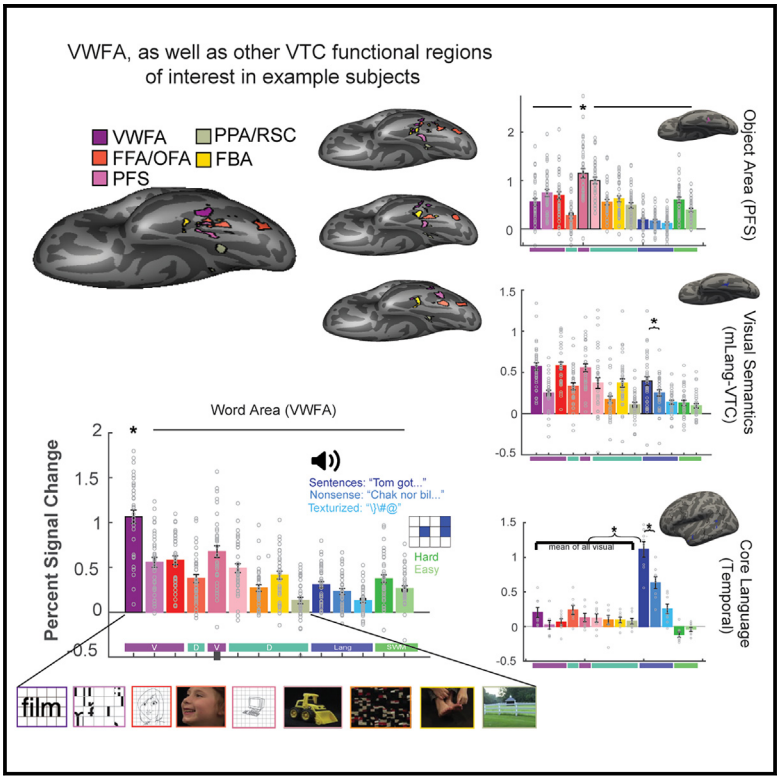
iScience

Article

Demystifying visual word form area visual and nonvisual response properties with precision fMRI

Graphical abstract



Highlights

- VWFA robustly shows the highest activation to visual words, surpassing all other categories
- VWFA has a distinct and reliable response profile, with a secondary preference for objects
- All high-level visual areas respond to attentional demand but only the VWFA to language
- Two additional VTC clusters show language responses, anterior to the VWFA

Authors

Jin Li, Kelly J. Hiersche, Zeynep M. Saygin

Correspondence

jli3618@gatech.edu (J.L.),
saygin.3@osu.edu (Z.M.S.)

In brief

Neuroscience; Sensory neuroscience; Cognitive neuroscience

Li et al., 2024, iScience *27*, 111481
December 20, 2024
https://doi.org/10.1016/j.isci.2024.111481

CellPress

iScience

CellPress
OPEN ACCESS

Article

Demystifying visual word form area visual and nonvisual response properties with precision fMRI

Jin Li,[1,2,3,*] Kelly J. Hiersche,[1,2] and Zeynep M. Saygin[1,2,4,*]
[1]Department of Psychology, The Ohio State University, Columbus, OH 43210, USA
[2]Center for Cognitive and Behavioral Brain Imaging, The Ohio State University, Columbus, OH 43210, USA
[3]School of Psychology, Georgia Institute of Technology, Atlanta, GA 30332, USA
[4]Lead contact
*Correspondence: jli3618@gatech.edu (J.L.), saygin.3@osu.edu (Z.M.S.)
https://doi.org/10.1016/j.isci.2024.111481

SUMMARY

The visual word form area (VWFA) is a region in the left ventrotemporal cortex (VTC) whose specificity remains contentious. Using precision fMRI, we examine the VWFA's responses to numerous visual and nonvisual stimuli, comparing them to adjacent category-selective visual regions and regions involved in language and attentional demand. We find that VWFA responds moderately to non-word visual stimuli, but is unique within VTC in its pronounced selectivity for visual words. Interestingly, the VWFA is also the only category-selective visual region engaged in auditory language, unlike the ubiquitous attentional demand effect throughout the VTC. However, this language selectivity is dwarfed by its visual responses even to nonpreferred categories, indicating the VWFA is not a core (amodal) language region. We also observed two additional auditory language VTC clusters, but these had no specificity for visual words. Our detailed investigation clarifies longstanding controversies about the landscape of visual and auditory language functionality within VTC.

INTRODUCTION

The ventral temporal cortex (VTC) consists of numerous regions each specializing in perceiving abstract visual stimulus categories(e.g., faces, objects, bodies, and places).[1–4] The visual word form area (VWFA) is perhaps one of the most fascinating of these VTC regions because it is specialized for processing a recent human invention: reading.[5,6] This functional specialization, as well as the experience-dependent nature of the VWFA,[7,8] make it a prime example for understanding the functional organization of the human brain. However, there is still debate over whether the VWFA is specialized *specifically* for visual words, which precludes researchers from digging deeper into the functional characteristics of the VWFA and how the human brain has the capacity to dedicate cortical tissue for new symbolic representations.

The key argument against the idea of a region that is dedicated to visual words is that the VWFA is also activated for other meaningful, non-word stimuli.[9,10] Proponents of this view argued that given the relatively recent invention of written script, the response to visual words is likely repurposed from other functionally specialized regions,[11] and still maintains other functions.[12] Studies that support this view have shown that while the response to words was quantitively less disrupted by noise, this effect was not qualitatively higher than that to line drawings of objects and false fonts.[13] Similarly, Xue and Poldrack[14] reported the lack of significant differences between known and unfamiliar scripts in the traditional VWFA. Therefore, some argued that the anatomical location of the VWFA, the posterior fusiform gyrus, is involved in complex shape processing[13,15] more generally.[16,17] However, we argue that the VWFA's responses to non-word stimuli in isolation (i.e., not in comparison to its responses to word stimuli) should not be taken as evidence against the VWFA's word selectivity. In fact, even the fusiform face area responds to non-face stimuli.[4] Instead, to better probe the function of the VWFA, one should ask (1) whether the VWFA shows similar functional characteristics as other category-selective VTC regions, responding robustly (e.g., ~twice as much, as proposed in Kanwisher et al.[18]) and significantly higher to the words than non-word categories, and (2) what is the functional profile of the VWFA in terms of its preferences to non-word categories.

Another argument against the VWFA's specialization for visual words is its involvement in auditory language processing.[9,19,20] In congenitally blind individuals, the site of the VWFA responds to both Braille words and auditory words but not tactile patterns or backward speech.[21] Similarly, activation to auditory words was also found in sighted individuals,[19] and when participants were asked to selectively attend to speech via a rhyme judgment task, both frontotemporal language regions *and* the VWFA showed increased activation as compared to when melody was presented.[22] However, the FFA is also activated during imagery[23] and by haptic stimuli of the faces.[24] But this activation

does not imply that the FFA is not specialized for faces. Instead, we should ask about the functional nature of these activations to non-orthographic stimuli. Specifically, how do these auditory language responses compare to those for written language? And how is the VWFA uniquely involved in processing auditory language, as compared to adjacent VTC regions or the entirety of the VTC? Is the VWFA another node of the core language network that responds selectively to high-level linguistics like semantics and syntax, regardless of the input modality? For example, in addition to Braille and auditory words, the VWFA was also sensitive to grammatical complexity manipulation of auditory sentences.[21] Alternatively, perhaps the VWFA mainly serves as a visual lookup dictionary for orthographic stimuli, which further passes visual inputs to frontotemporal language regions via its privileged connectivity with the language cortex.

Finally, the exact location and definition of the VWFA are not consistent in previous studies, making it difficult to reach any consensus among studies regarding the function of the VWFA. Although located in approximately the same location across individuals, the VWFA is a small region, and the exact location varies from person to person.[25] However, many previous studies examining the function of the VWFA relied on group activation maps or used anatomical coordinates (on a template brain)[26–28]. These methods may not capture word-selective voxels because they do not account for individual variability. Further, previous studies typically only defined the VWFA, or the VWFA and one other VTC comparison region. However, the mosaic-like organization of the VTC encompasses multiple category-selective regions that are located closely to each other and to the VWFA, therefore, a critical review of all VTC regions is needed in the literature. Finally, a last point of inconsistency across previous studies was the control conditions used: the VWFA was initially defined using fixation/rest or checkboard stimuli,[5,29] or otherwise poorly controlled for visual complexity and general semantic processing. However, despite these limitations, subsequent studies continued referencing this anatomical location as the VWFA. Consequently, this lack of functional specificity in its initial definition could be a contributing factor to studies reporting activations in the VWFA during non-word processing tasks.

In the present study, rather than offering simple yes or no answers to the lingering debates about the VWFA, our goal is to systematically examine the VTC's functional characteristics using a wide range of visual and nonvisual stimuli. Specifically, we utilized precision fMRI to measure the subject-specific VWFA's (along with six other high-level visual regions) functional response profile across four distinct tasks spanning multiple sessions and encompassing a total of 14 experimental conditions. This allowed us to thoroughly probe the function of the VWFA in comparison with functionally related or spatially proximate regions. We assessed the VWFA's activation in response to both high-level visual conditions and auditory language. The results not only demonstrate the robustness of the VWFA's word selectivity but also provide insight into its activation during non-word processing, revealing its distinctive involvement in auditory language processing. By transcending the binary question of whether the VWFA exclusively processes words, our findings shed light on a more detailed picture of the VWFA's responses. This understanding could potentially yield fresh insight into the development of the human brain's functional organization.

RESULTS

VWFA is visually selective for written words, showing a distinct neural response signature from adjacent VTC regions

We first examined the functional response profile of the VWFA to a wide range of visual stimuli, auditory language conditions, and conditions of a spatial working memory task (Figure 1), and compared those responses to other high-level category-selective regions (functional regions of interest (fROIs)) in neurotypical adult subjects who have completed two runs of the static and dynamic localizer ($N = 37$) and at least one run of the language and spatial working memory tasks. Using reference parcels (created from the independent group of adults in previous studies) as search spaces (see STAR Methods), these VTC fROIs were defined by contrasting the condition of interests with the remaining conditions in a localizer task (Table 1). Functional responses were extracted from left-out data that was independent of that used to define the fROIs, as well as the conditions across the other fMRI experiments. The main results focus on the left VTC given the left-lateralized nature of the VWFA (right VTC results in Figures S3 and S4; Table S2).

As expected, when defining all fROIs individually to avoid blurring the boundaries between cortically adjacent regions and account for variability across subjects (Figure 2A shows the fROIs in one example subject, all subjects Figures S1 and S2), the VWFA responded significantly higher to visual words than all other conditions (paired samples t-tests, all $p < 0.001$, Table 2; Figure 2B). When calculating the average time-course across the experimental block, response to words was higher than all other static and dynamic visual conditions (Figure 2C), Additionally, the VWFA showed no preference for conditions other than words for the duration of the block.

To what extent is the VWFA unique in its functional response pattern? First, no other region's highest response was to visual words; instead, as expected, face-, object- and scene-selective regions showed significantly highest responses for their preferred condition, while the FBA did not show a clear categorical preference (Figure 3; Table S1). Using selectivity indices (see STAR Methods) which allowed for comparisons across regions, we find the VWFA had the greatest selectivity to words compared other adjacent fROIs (Figure 4, VWFA vs. all other fROIs, $t(31)>5.13$, $p < 1.47\text{x}10^{-5}$). This is also true of the other fROIs, the selectivity to their preferred category is greater than all other fROIs selectivity to that category (see Table S3). Next, we compared the overall response profile of the VWFA vs. other VTC regions (see STAR Methods). Specifically, we found that the VWFA showed a consistent response profile across individuals to the 14 functional conditions: the VWFA's response profile between subjects was significantly more correlated than the VWFA's response profile to any other VTC region of the same subject (between-subjects VWFA-VWFA correlation vs. within-subjects VWFA-FFA: $t(36) = 3.06$, $p = 0.028$; vs. VWFA-OFA: $t(36) = 4.97$, $p = 0.002$; VWFA-FBA, $t(36) = 1.92$, $p = 0.062$; VWFA-PFS, $t(36) = 3.48$, $p = 0.004$; VWFA-RSC, $t(36) = 8.64$,

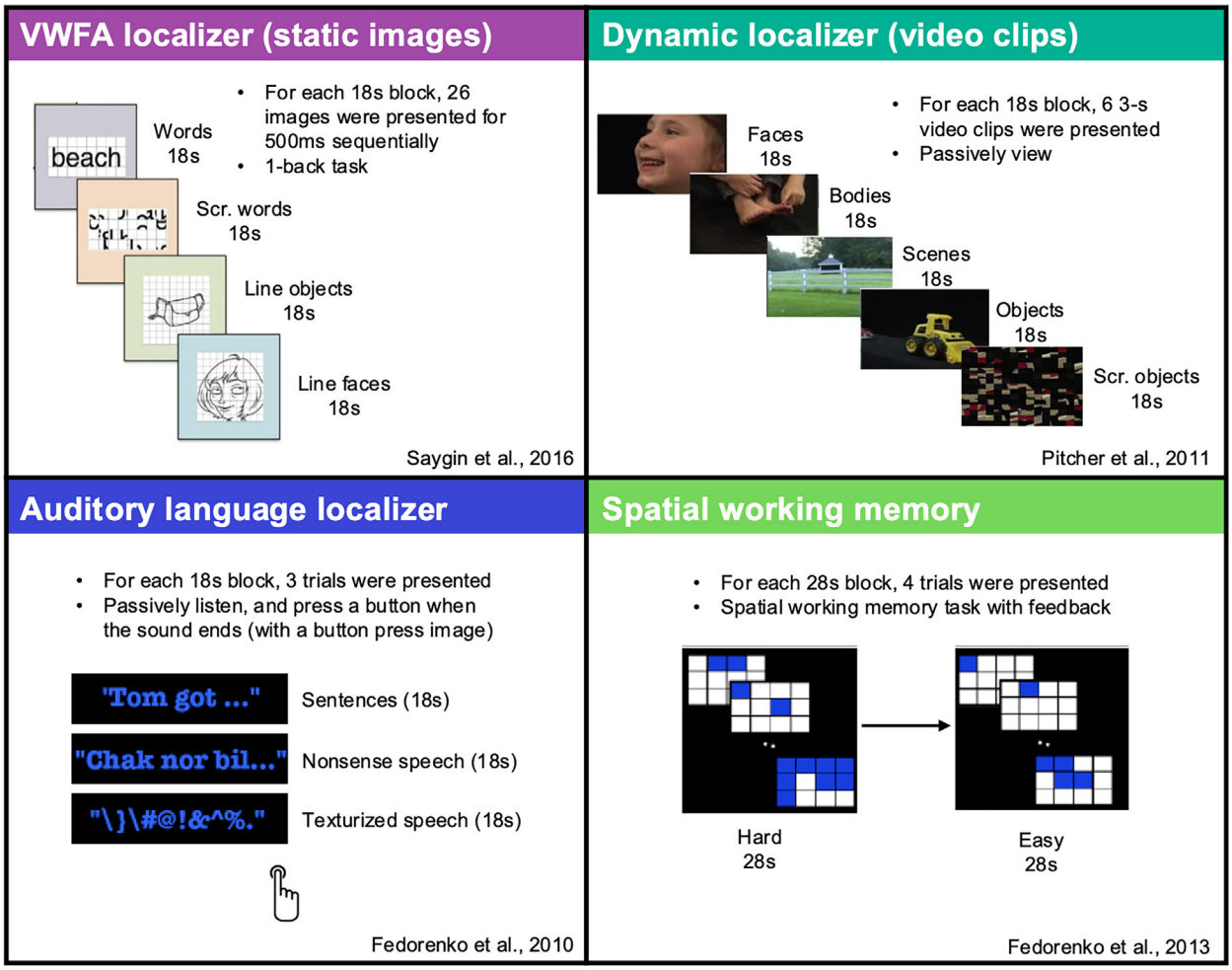


Figure 1. Functional tasks and example stimuli used in the current study

We used previously well-established tasks[30–33] (stimuli were adapted here with the permission from authors of the original studies) to localize fROIs and probe functional responses (see STAR Methods for more details).

$p < 0.001$; VWFA-PPA, $t(36) = 8.54$, $p < 0.001$; corrected). Therefore, the VWFA is a unique VTC region that not only shows the strongest activation to written words but also has a distinctive functional fingerprint across a wide range of stimuli.

Does the VWFA also show some preference for nonword stimuli, as shown in some prior work? Interestingly, when we examined the VWFA's selectivity to non-preferred conditions, we found that in addition to the VWFA's absolute preference for visual words, it also showed higher activity to objects (average of line-object and dynamic object vs. average others excluding the response to words: $t(36) = 5.89$, $p = 9.82 \times 10^{-7}$). Had we only used the dynamic localizer and examined the VWFA's selectivity using meaningless scrambled objects as the control condition, commonly done in prior work, the VWFA would appear not only object, ($t(35) = 5.19$, $p = 9.12 \times 10^{-6}$), but also face ($t(35) = 1.99$, $p = 0.054$) and even body ($t(35) = 3.60$, $p = 9.87 \times 10^{-4}$) selective. However, these response patterns should not be taken as evidence against VWFA's word selectivity. Instead, our results suggested that despite a distribution of responses to other visual categories, the VWFA shows an absolute highest word preference, highlighting the importance of comprehensively comparing the VWFA's activation to a wide range of stimuli (Figures 2 and 4; Tables S1 and S3).

Note that the static VWFA localizer had slightly different scanning parameters (see STAR Methods for details) from the dynamic localizer. Therefore, one potential confound is the observed results mainly reflect potential differences in scan parameters, motion, or SNR between experiments. Therefore, we rescanned a subset of subjects on two additional runs of the static visual localizer with matching scan parameters to the dynamic localizer. We replicate the main results even after matching these potential confounding factors (Figure S6; Tables S5 and S6). Therefore, any activation differences observed for the conditions of the static and dynamic localizer do not impact the categorical selectivity of the fROIs (see Discussion).

The functional landscape of the VTC

In this section, we examine possible overlap between the category-selective fROIs and examine the responses of the VTC to

Table 1. Tasks and contrasts to define the functional regions of interest

	fROIs[a]	TASKS AND CONTRASTS TO DEFINE
high-level visual fROIs		**The static VWFA localizer**
		Words, Scrambled Words, Line Faces, Line Objects
	VWFA	Words >average of other conditions
		The dynamic localizer
		Faces, Objects, Bodies, Scenes, Scrambled Objects
	FFA	Faces >average of others
	OFA	Faces >average of others
	PFS	Objects >average of others
	RSC	Scenes >average of others
	PPA	Scenes >average of others
	FBA	Bodies >average of others
		The language task (Auditory)
		Sentences (Sn), Nonsense Sounds (Ns), Texturized Sounds (Tx)
	language fROIs[b]	Sentences > Texturized
		The spatial working memory task
		Hard, Easy
	MD fROIs[c]	Hard > Easy

[a]VWFA, Visual Word Form Area; FFA, Fusiform Face Area; OFA, Occipital Face Area; PFS, Posterior Fusiform Sulcus; RSC, Retrosplenial Cortex; PPA, Parahippocampal Place Area; FBA, Fusiform Body Area
[b]6 language fROIs (3 frontal and 3 temporal)
[c]10 multiple-demand (MD) fROIs (7 frontal and 3 parietal)

all conditions, without using predefined search spaces (parcels), to control for potential biases in the above presented results. First, it is possible that the above results were biased by 1) the specific method (i.e., top 150 vertices) used to define fROIs; or 2) the predefined search spaces we used when defining the fROIs. Using the top 150 method results controls for the size of the fROIs in comparison, and also allowed us to identify a relatively small set of the most responsive voxels, which further avoided the overlap between regions to ensure spatial specificity. Indeed, when we explicitly quantified any overlap among the VWFA and other VTC fROIs using the top 150 vertices (see STAR Methods), we found little, if any, overlap within an individual (mean overlap of VWFA with FFA:13, VWFA with FBA: 9, VWFA with PFS: 3 vertices). Additionally, we found similar results (Table S1) when applying different criteria to identify the fROIs: selecting the top 10% of vertices within the search space and using a hard significance threshold ($p < 0.005$) (see Table S7 for descriptive information (e.g., number of subjects has the significant fROIs, size and the overlap between regions) for the fROIs defined with these other two methods).

Further, to complement the fROI analyses, which might miss selective responses outside the predefined search spaces and reveal little about the spatial spread of potential subregions of the word-selective areas, we examined voxel-wise selectivity to all visual stimuli, from posterior to anterior VTC in both fusiform and inferior temporal cortex (see STAR Methods). Interestingly, we found selective responses for words, faces, and objects within the fusiform cortex, from the mid-fusiform and extending posteriorly (Figure 5). This suggested that selective voxels for different high-level conditions were close, but distinct, to each other within a relatively small swarth of the fusiform cortex. Moreover, only word-selective responses were found in the inferior temporal cortex, consistent with the notion that word selectivity is often found to be more lateral. Again, no reliable body-selective response was found even within the entire VTC; thus, the left FBA was excluded from further analyses.

The VWFA, compared to adjacent VTC regions, selectively responds to auditory language, but is not a core part of amodal high-level language network

Next, we asked the extent to which the VWFA is multimodal, also responding to auditory language. We first investigated if the VWFA shows language-selective response by comparing its responses to English sentences (Sn) with nonword sequences (nonsense), presented auditorily (see STAR Methods). Note that the nonword condition shares speech features like prosody and phonological processing with sentences, thus, the difference indicates selective responses to high-level linguistic features (i.e., semantics and syntax). We found that only the VWFA showed significant language selective responses (Sn > Ns: $t(35) = 2.85$, $p = 2.20\text{x}10^{-2}$; corrected) and that none of the other category-selective lVTC regions (FFA, PFS, OFA, RSC, PPA) differentiated between sentences and nonword sequences (all $p > 0.05$; Table S4). This language preference in the VWFA is also clear when examining the time-course of responses during the language task (Figure 6A).

How does the VWFA respond to auditory language compared to the canonical language network[30]? Unsurprisingly, the fronto-temporal language fROIs (2 temporal and 3 frontal regions, see STAR Methods) showed language selectivity (Sn > Ns, paired samples t-tests: collapsed across temporal fROIs: $t(33) = 7.43$, $p = 1.54\text{x}10^{-8}$; and frontal fROIs: $t(33) = 4.59$, $p = 6.21\text{x}10^{-5}$). Critically, as shown in Figure 6C, compared to canonical language fROIs, VWFA showed significantly lower activation to auditory language, regardless of conditions (Sn: temporal vs. VWFA: $t(33) = 11.26$, $p = 7.58\text{x}10^{-13}$; frontal vs. VWFA: $t(33) = 5.21$, $p = 9.90\text{x}10^{-6}$; Ns: temporal vs. VWFA: $t(33) = 8.71$, $p = 4.53\text{x}10^{-10}$; frontal vs. VWFA: $t(33) = 4.51$, $p = 7.80\text{x}10^{-5}$). There was a significant region × conditions (Sn, Ns) interaction between VWFA and temporal language ($F(1,33) = 39.22$, $p = 4.47\text{x}10^{-7}$), and a trending interaction between VWFA and frontal language regions ($F(1,33) = 3.96$, $p = 0.055$), indicating that the condition effect (Sn>Ns) observed in the language regions was different (larger) than that in the VWFA. Moreover, selectivity indices calculated across all task conditions (to normalize task differences) showed that VWFA's word selectivity was significantly higher than its language selectivity ($t(34) = 7.43$, $p = 1.27\text{x}10^{-8}$). Additionally, not only does the VWFA respond lower to auditory stimuli than the language network, its response profile to all visual and nonvisual conditions was different from

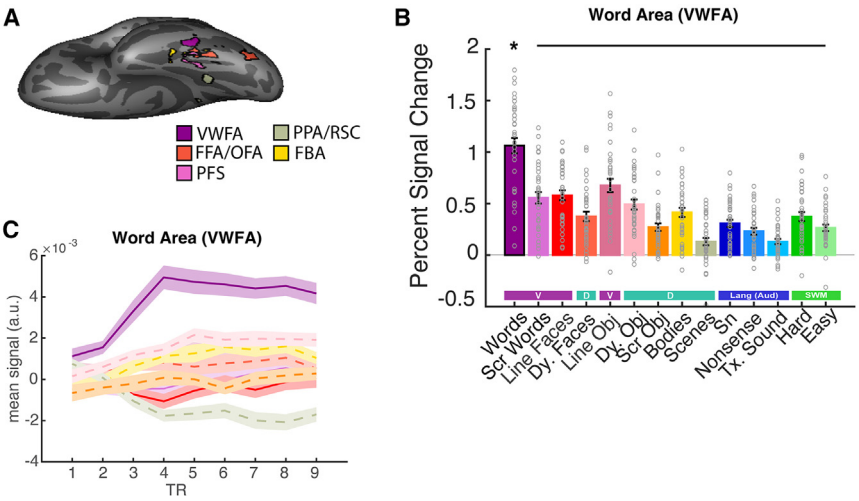


Figure 2. Example VTC fROIs and response profile of the VWFA

(A) VTC fROIs in the left hemisphere for an example subject.

(B) Functional profile of the left VWFA. The VWFA's percent signal change is significantly higher for Words compared to all other visual and non-visual conditions. The mean percent signal change (data are represented as mean ± SEM) to various visual and non-visual conditions are plotted. Colored boxes at the bottom note the task each condition belongs to: static VWFA localizer (purple), dynamic visual localizer (aquamarine), auditory language (blue), and spatial working memory (SWM) (green). The preferred category, words, has a thick black outline. Individual subject PSCs are shown with gray hollow circles. Significance is noted (*$p < 0.05$, corrected for 13 total pairwise t-test comparisons with Bonferroni-Holm method) for words (denoted by an asterisk) only, with a black line showing all conditions significantly lower than words (see Table 2 for all statistical results).

(C) Average time-course of VWFA's responses to blocks of different experimental conditions. Responses were plotted by TR (TR = 2s), starting from the onset of each block. Throughout a block, the VWFA's greatest response is for its preferred category: words. Solid line for mean across all subjects for conditions of the static localizer, dashed line for mean across all subjects for dynamic localizer conditions, and shading for standard error. VWFA, visual word form area; FFA, fusiform face area; OFA, occipital face area; PFS, posterior fusiform sulcus; PPA, parahippocampal place area; RSC, retrosplenial cortex (RSC); FBA, fusiform body area.

that observed in the canonical frontotemporal language regions (Figure 7). The language network responds much higher to auditory sentences than all other visual conditions (including words) (Sentences vs. average of all visual conditions: frontal language regions, W = 35, $p = 0.015$; temporal language regions, W = 36, $p = 0.008$; Wilcoxon signed rank test for the small sample size ($N = 8$)) (Figure 7), whereas Figure 2 shows that the VWFA responds significantly more to any visual category (even non-preferred ones) than auditorily presented linguistic stimuli (paired t-test between mean response to all non-preferred visual categories and auditory sentences: $t(35) = 2.83$, $p = 0.0076$), suggesting that the VWFA is primarily a visual region. Moreover, frontal and temporal language regions did not show preferential responses to words vs. other high-level visual categories (words vs. other conditions, all $p > 0.07$), except that the temporal language regions showed more activation to words than to line-drawing faces (W = 32, $p = 0.055$; uncorrected). If we had only explored language activation within the VTC, we may have concluded that the VWFA was in fact selective to amodal linguistic processing, but these analyses show a clearer picture, in which the VWFA is responds to auditory language significantly, but to a lesser degree than the language network and less than its typical response to visual stimuli. These results further highlight the VWFA's function as a high-level visual region specifically for processing orthographic stimuli.

Table 2. Comparing the VWFA's response to words with all other visual conditions

VWFA Words vs.	t (df)	Corrected p
Scrambled Words	7.03 (35)	7.10×10^{-8}[a]
Line Faces	7.93 (35)	1.25×10^{-8}[a]
Dynamic Faces	8.73 (35)	1.56×10^{-9}[a]
Line Objects	5.57 (35)	2.87×10^{-6}[a]
Dynamic Objects	7.19 (35)	6.45×10^{-8}[a]
Scrambled Objects	9.29 (35)	3.95×10^{-10}[a]
Dynamic Bodies	7.91 (35)	1.25×10^{-8}[a]
Dynamic Scenes	10.27 (35)	3.36×10^{-11}[a]

[a]Bonferroni-Holm $p < 0.05$.

As a comparison, we also examined the effect of domain-general attentional demands on the VTC, as frontoparietal multiple demand (MD) regions (see STAR Methods) are in close vicinity of the frontal language regions and previous studies also reported connectivity between the VWFA and dorsal parietal attention region. We confirmed that these MD regions demonstrated a significant attentional demand effect, measured by comparing the response to Hard vs. Easy conditions in a spatial working memory task: frontal MD: $t(33) = 10.24$, $p = 9.02 \times 10^{-12}$; parietal MD: $t(33) = 10.60$, $p = 3.64 \times 10^{-12}$. Critically, in contrast with the unique effect of language on the VWFA, almost all VTC fROIs (except for the RSC) were significantly modulated by attention (Figure 6C, right) and the effect of attentional demand in the VWFA was similar to other fROIs (e.g., FFA) or even lower than other fROIs (e.g., PFS and OFA; see full pairwise comparisons in Table S4). Time course analyses of VTC fROIs during the spatial working memory task also confirmed this ubiquitous attentional effect (Figure 6B). Moreover, even though the spatial working memory task requires visual processing, we found significant a region × condition (Hard, Easy) interaction between VWFA and frontal MD ($F(1,33) = 89,17$, $p = 6.67 \times 10^{-11}$) and parietal MD ($F(1,33) = 130.88$, $p = 5.02 \times 10^{-13}$), suggesting that the magnitude of the attentional load effect in the VWFA was significantly lower than the effect observed in the MD regions. These results suggested that, in contrast to the linguistic effect, the modulation of attention is general within the VTC, highlighting the unique involvement of the VWFA during auditory language processing.

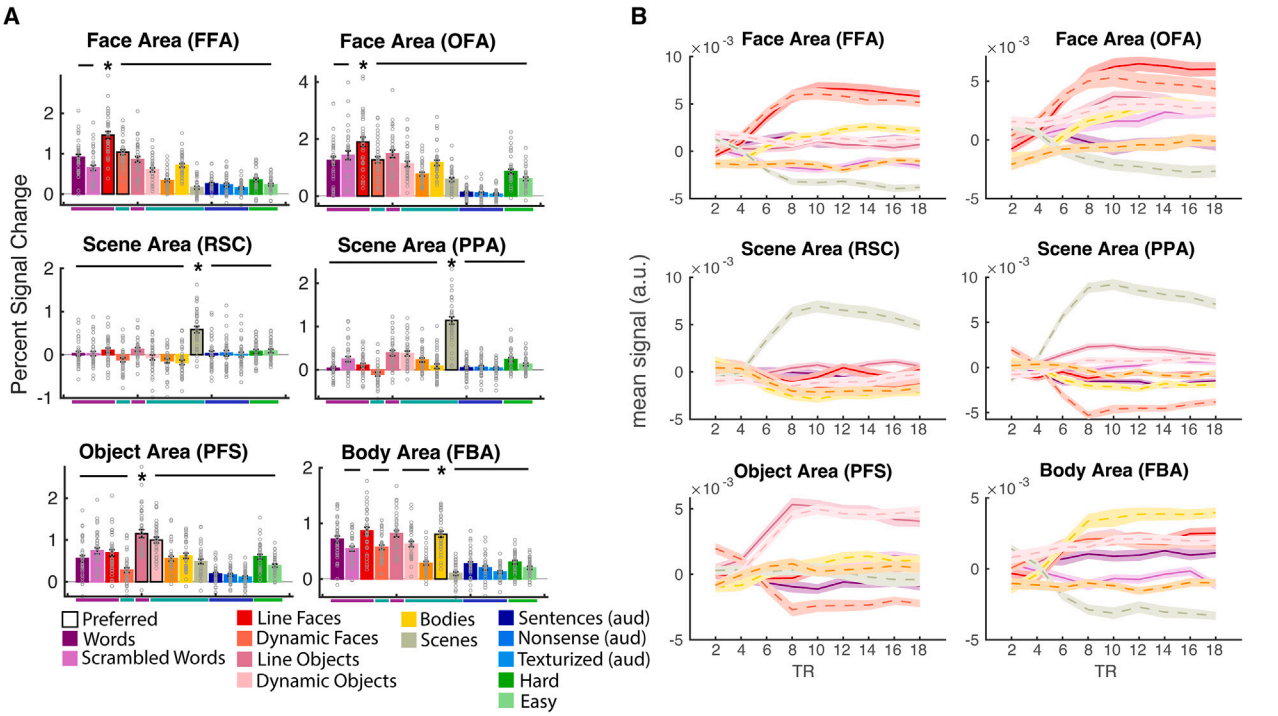


Figure 3. Functional responses in category-selective regions of the left VTC

(A) Functional profiles for left VTC high-level visual regions. All regions show significantly highest percent signal change to their expected, preferred category compared to all other visual and non-visual categories (except FBA). The preferred condition(s) for each fROI is outlined in black. The mean percent signal change (data are represented as mean ± SEM) to various visual and non-visual conditions are plotted. Colored boxes at the bottom note the task each condition belongs to: static localizer (purple), dynamic visual localizer (aquamarine), language (blue), and spatial working memory (SWM) (green). Individual subject PSCs are shown with gray hollow circles. Significance is noted (*$p < 0.05$, Bonferroni-Holm) for the preferred category (denoted by an asterisk) only, with a black lining showing all conditions significantly lower than the preferred category (pairwise t-test). See Table S1 for a full list of pairwise comparisons.

(B) Average time-course of each fROI's responses to blocks of different experimental conditions. Responses were averaged every two TRs (TR = 1s), and plotted from the onset of each block. Throughout a block, each fROI showed the greatest response is for its preferred condition. The solid line for the mean across all subjects for conditions of the static localizer, dashed line for mean across all subjects for dynamic localizer conditions, and shading for standard error. FFA: fusiform face area, OFA: occipital face area, RSC: retrosplenial cortex, PPA: parahippocampal place area, PFS: posterior fusiform sulcus, FBA: fusiform body area.

The VWFA is distinct from the basal temporal language area

As we have shown above, while VWFA uniquely showed some auditory language sensitivity, this response was lower than that in the core amodal language network. Lastly, previous studies have proposed a "basal temporal language area (BTLA)" located between the left temporal pole and the VWFA.[34] Here, we asked if there exist such language clusters within the VTC that are distinct from the VWFA. We first examined the probabilistic map of the auditory language activation (see STAR Methods). We found two clusters located in the left VTC that showed language selectivity (Sn>Tx): anterior language VTC (aLang-VTC, Figure 8A) and medial language VTC (mLang-VTC) (Figure 8B; similar clusters were observed for Sn vs. Ns; Figure S8; see also Figure S9 for RH hotspots with similar effect for attentional demand but weaker language activation). Using these two clusters as the search spaces, we defined subject-specific fROIs for mLang-VTC and aLang-VTC and examined their functional profile (see Figure 8C). In the following section, we characterized the functionality of these two "language" regions quantitively at individual level in comparison to the VWFA and left amodal frontotemporal language network.

As expected, both the aLang-VTC and mLang-VTC were language selective (paired-samples t-tests (Sent>Ns): aLang-VTC, $t(33) = 3.84$, $p = 5.32 \times 10^{-4}$; mLang-VTC, $t(32) = 3.35$, $p = 0.002$). Critically, aLang-VTC's response to sentences was significantly higher than its average response to all visual categories (paired-samples, two-tailed, t-tests: $t(33) = 3.45$, $p = 0.0015$), and it did not display a clear preference among these high-level visual categories, responding equally as high to multiple visually categories (one-way rmANOVA of responses to the static localizer: $F(2,64) = 0.04$, $p = 0.959$; one-way rmANOVA of response to dynamic localizer: $F(3,96) = 7.02$, $p = 2.55 \times 10^{-4}$; post-hoc t-tests show significantly lower response to scenes, but bodies, faces, and objects are not distinguishable). In contrast, the mLang-VTC, likely to be the BTLA, showed comparable responses to visual conditions as to auditory language (sentences vs. average response to visual stimuli: $t(33) = 0.45$, $p = 0.66$)). Importantly, however, just like aLang-VTC, mLang-VTC responded to different high-level visual stimuli equally and did not show a clear category preference; it

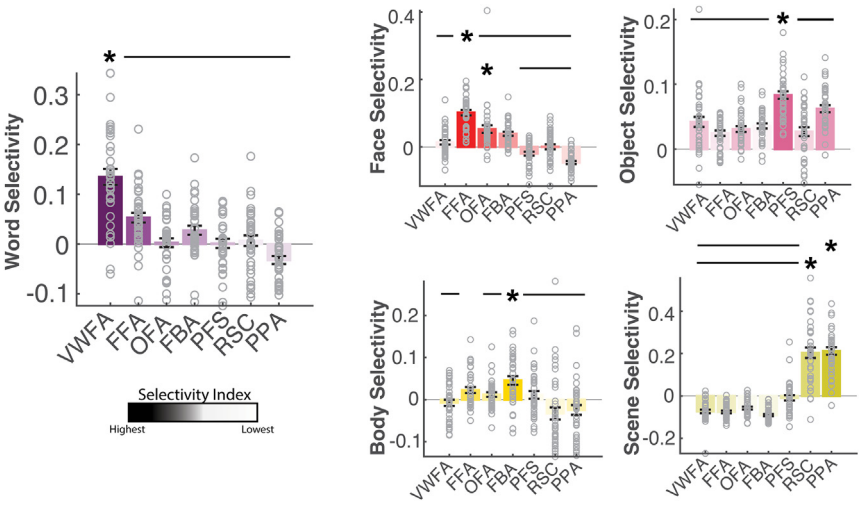


Figure 4. Category selectivity indices in the category-selective regions

All regions show highest selectivity for their preferred category, and greater selectivity to that category than all other fROIs (except OFA). For each visual category, we computed the selectivity indices (see STAR Methods). Data are represented as mean ± SEM. Asterisks denote the specific category-selective regions associated with each category selectivity. Horizontal lines indicate the values are significantly (pairwise t-test, Bonferroni-Holm corrected) lower than that of the corresponding region. See Table S3 for all pairwise comparison results.

responded equally as strong to multiple visual categories (one-way rmANOVA of responses to the static localizer: $F(2,66) = 0.25$, $p = 0.78$; one-way rmANOVA of response to dynamic localizer: $F(3,93) = 11,67$, $p = 1.47x10^{-6}$; post-hoc t-tests show significantly lower response to scenes, but bodies, faces, and objects are not distinguishable). These results, along with the observation that there is spatial overlap between the mLang cluster and the VWFA parcel, possibly explains why previous studies may have conflated the VWFA with these more anterior language regions. Here we show that there are two anterior language clusters that are functionally distinct from the VWFA (i.e., respond equally to all high-level visual categories) and may engage in abstract semantic processing generally.

DISCUSSION

Studies have continued to debate the existence and functional characteristics of the VWFA.[9,10,35] Our study investigated the VWFA's comprehensive response profile to a wide range of visual and non-visual stimuli and compared its neural signature to those of other spatially adjacent VTC regions. We found that while responding moderately to objects, the VWFA's response to visual words towers above its responses to all other high-level visual categories. Moreover, we found that while the VWFA is the only VTC region that showed sensitivity to auditory language, the VWFA is modality-dependent. The VWFA is primarily visual: its responses to even non-word visual stimuli surpass its response to auditory language and it has a distinct functional profile from language regions, suggesting that it is not part of the core language network. In the following sections, we discuss defining a category-selective region (conceptually and methodologically), methodological discrepancies that might have contributed to the inconsistency regarding the VWFA's function, the implications of non-word responses in the VWFA, and the hierarchical organization of the VTC: from posterior regions that respond to visual forms of the words (i.e., VWFA) to anterior areas associated with abstract semantics.

The definition of the VWFA

Our results suggest that the muddled picture of the VWFA was at least partially driven by the failure to take into account the individual variability of the VWFA as well as the intertwined nature of category-selective VTC regions (evident in the fROI map for each individual; Figures S1 and S2). Here, we defined the VWFA with rigorous methodological considerations: the subject-specific approach[30] to account for individual differences, different thresholds to select the candidate fROI voxels to examine the robustness of observed results (see Tables S1 and S2), multiple control conditions to match either visual complexity or conceptual semantics for functional specificity, and simultaneously defining adjacent VTC regions to ensure high spatial specificity. While previous studies utilizing whole-brain group analysis observed no word-selective responses,[9,10,28,36] here we were able to localize a VWFA fROI in each individual that responded significantly higher to visual words than to all other visual and non-visual stimuli (Table S7). Our results highlight the importance of defining the VWFA in each individual and echo recent emphasis on anatomical precision when defining VWFA[25,37–42]: when lumping all subjects together, either by implementing group analysis or drawing arbitrary spheres around predefined coordinates, we lose the precision and spatial resolution to identify word-selective voxels, as illustrated by the extensive overlap between the group-level probabilistic maps of different category-selective activations (Figures S10 and S11). This might be one reason why previous studies have reported that the VWFA responds to non-word stimuli like faces, objects, or symbols.[16,17,28,43]

Critically, we found that the VWFA was functionally different from other VTC fROIs: it showed minimal overlap with other VTC fROIs at the individual level, and also showed the highest responses to visual words versus other non-word conditions (see Figure 2, e.g., nearly twice as much to words (average PSC 1.06 ± 0.44) as to the second highest category (i.e., objects, average PSC of 0.58 ± 0.35)). This aligns with the definition of a category-selective region that is domain-specific.[18] For example, the FFA shows higher activation to objects versus non-face conditions, but these responses are much lower than its responses to faces (usually twice as low[4]; see[18] for a discussion). Finally, the VWFA shows a more similar response profile across subjects than it does to other fROIs within a subject.

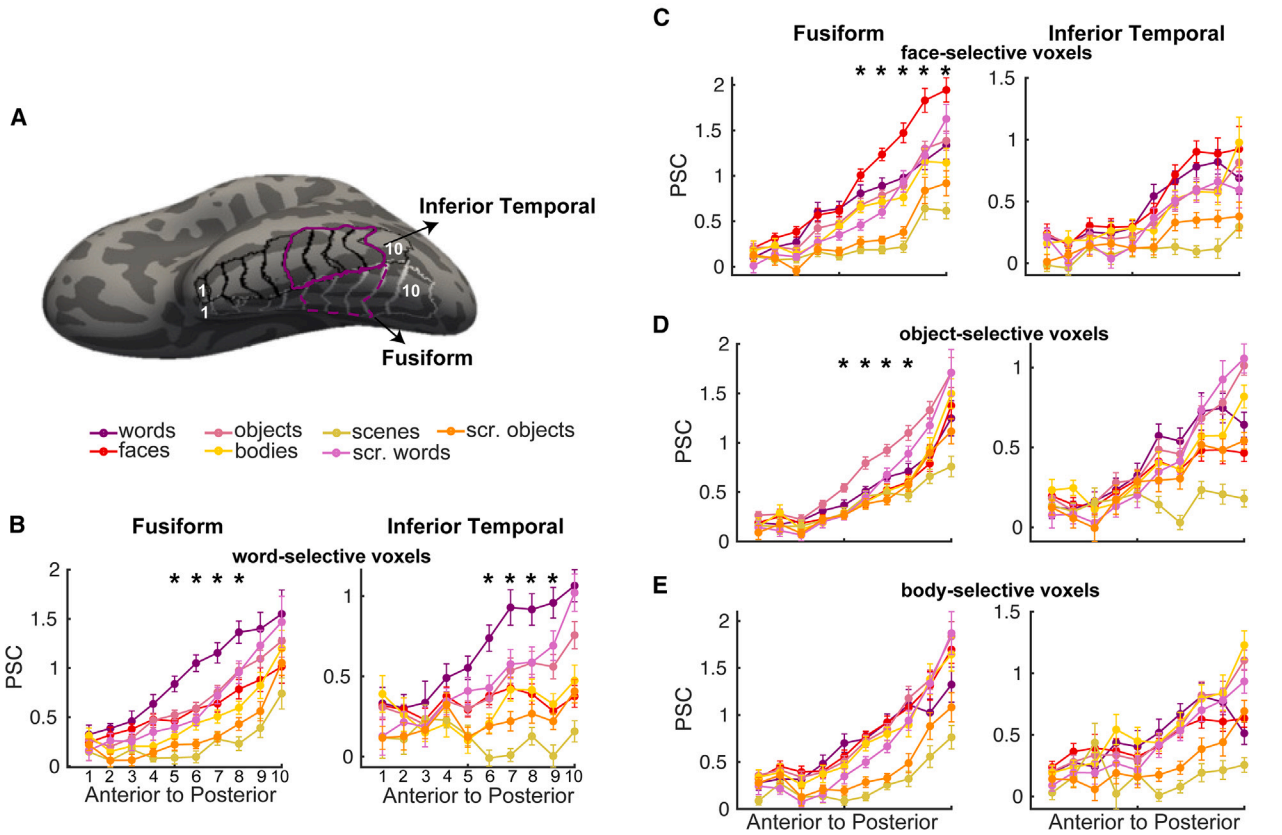


Figure 5. Categorical responses from the posterior to anterior left VTC

(A) Inferior temporal (black outline) and fusiform (white outline) parcel that comprise the VTC from the Desikan-Killiany parcellation. We divided each anatomical parcel into 10 equal sections from posterior to anterior. Purple lines indicate segments where we found significant word-selective responses.

(B–E) PSC to each of the visual conditions at each section along the posterior-to-anterior axis. Data are represented as mean ± SEM. The fusiform gyrus contains sections that responded highest to words, faces, and objects. The inferior temporal gyrus only contains sections with word-selective voxels. The asterisk denotes that the PSC to the condition of interests is significantly higher than all the other conditions at a given location (*$p < 0.05$, Bonferroni-Holm corrected for 10 pairwise t-tests across anatomical segments; see Table S9 for all statistical results). Note that for faces and objects, we averaged PSC from the static and dynamic localizers.

Functional profile of the VWFA: Activation for non-word stimuli

To what extent does the VWFA show preferential activity to other high-level visual categories? Answering this question may give us clues as to how this piece of cortex is able to process words. Previous work investigating the neuronal tuning of the ventral visual stream showed that the VTC is organized by underlying neuronal preferences for different visual and semantic features (e.g., fovea/peripheral bias[44]; simple geometrical features[45,46]; rectilinearity[47]; spatial frequency[48]; spikiness[49]; animacy and real-world size[50]). Therefore, some researchers proposed that the VWFA may emerge or be repurposed from part of another high-level visual region[11] that shares similar visual features with visual words. In this section, we discuss insights we gain from these non-word responses: that the cortical tissue later becomes word-selective also shows some sensitivity to local visual features like line segments and junctions, stimuli in the center visual field, and stimuli that encode abstract semantic information.

First, we found that while the VWFA responds more to words than other visual categories, its response to the scrambled words condition was surprisingly high (about as high to faces and higher than scenes and bodies). This result suggests that the VWFA's preference for visual words may emerge from existing preferences for geometrical visual features such as line segments, junctions and contours[51–53]

Second, the foveal hypothesis of the VTC proposes a medial-to-lateral dissociation in the ventral visual stream for processing peripheral and fovea stimuli, respectively. Consequentially, the lateral portion of the VTC houses both the FFA and VWFA, as visual words and faces are processed foveally.[44] We see that in our results as well, and we also find that the VWFA shows the least responses to scenes, fitting the lateral-to-medial functional division. We might then expect to see strong face responses in the VWFA,[54] as compared to other visual stimuli; however, we do not find that the VWFA responds more to faces than other high-level visual conditions in either static or dynamic localizer, except dynamic scrambled objects which only controls for low-level visual features such as color and edges, resembling the checkerboard condition of early VWFA functional localization.[5] This aligns with prior work showing overlap between

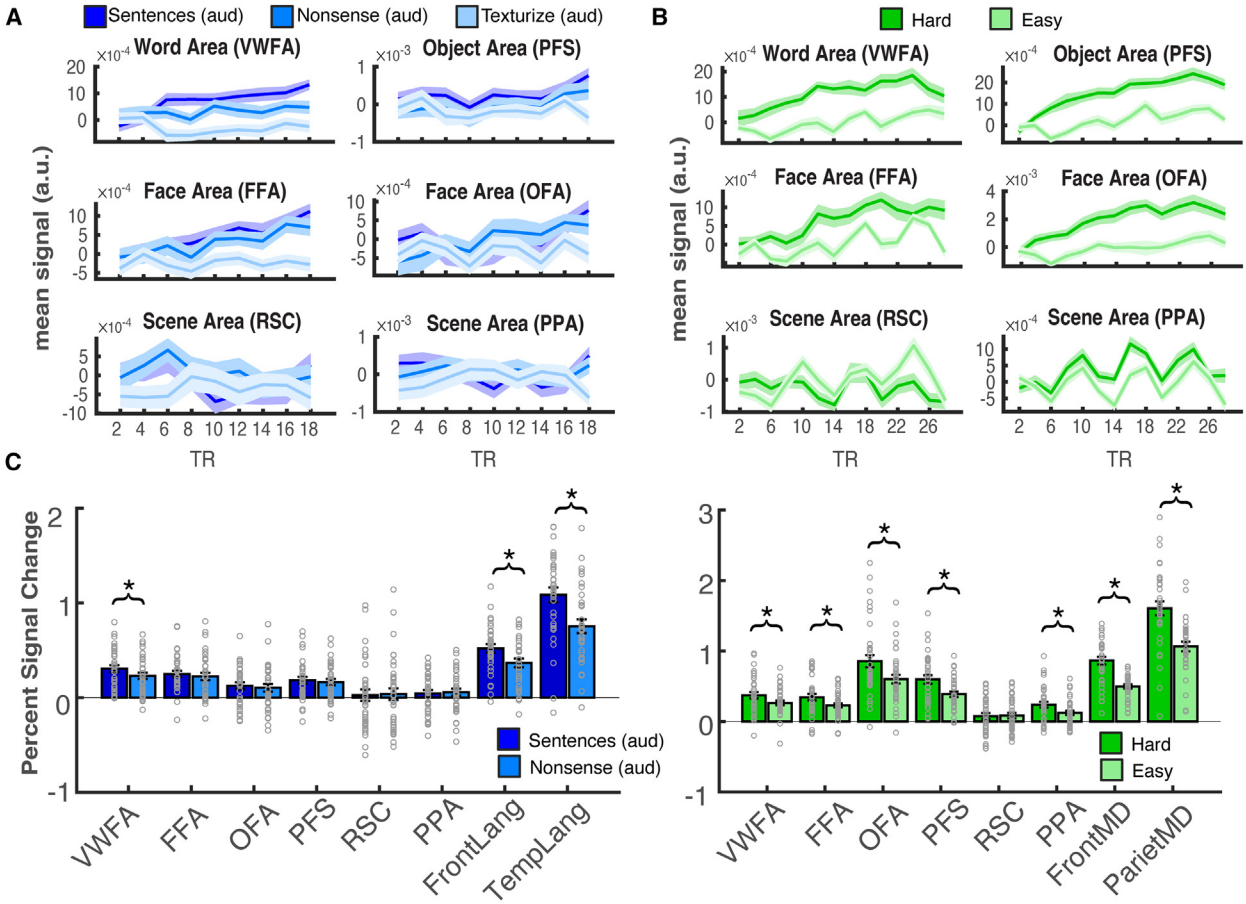


Figure 6. Functional responses in auditory language and spatial working memory tasks

(A) The time-course of each VTC fROI during the language task (averaged across all blocks). Only the VWFA differentiates the sentences and nonsense conditions. Data are represented as mean ± SEM.

(B) The time-course of each VTC fROI during the spatial working memory task (averaged across blocks). For A and B, responses were averaged every two TRs (TR = 1s), and plotted from the onset of each block. All fROIs (except RSC) differentiate the hard and easy conditions, suggesting the effect of attentional demand for all regions. Dark line for mean across all subjects and shading for standard error.

(C) Mean percent signal change (data are represented as mean ± SEM) for each VTC fROI and the language fROIs to the language task (left) and the MD fROIs to the spatial working memory task (right). The VWFA, along with frontal and temporal language respond significantly more to sentences than nonsense. All VTC fROIs (except RSC), as well as the frontal and parietal multiple demand (MD) regions respond more to the hard than easy condition of a spatial working memory task. Asterisks denote significantly (*$p < 0.05$, Bonferroni-Holm corrected for 8 pairwise t-tests across fROIs) higher responses to auditory sentence vs. nonsense speech (language task) or higher for Hard vs. Easy (spatial working memory task). See Table S4 for all statistical results. Individual subject PSCs are shown with gray hollow circles.

word and face responses in the VWFA only when using fixation as a control condition to define VWFA (thus presumably including a large portion of lateral VTC rather than just the VWFA) Nestor et al.[55]

Finally, among all non-word visual conditions, the VWFA responds highest to objects (Figure 2B): it responds the second-highest to the line-drawing of objects and responses to objects (average of static and dynamic) are significantly higher than the average of all other non-word stimuli. This relatively high activation to objects was also observed in a previous study, where Ben-Shachar et al.[13] reported that the VWFA responded second-highest to line-drawing objects, followed by false fonts. This may further explain the difficulties of differentiating word-selective responses from objects, as noted in previous studies.[56,57] Perhaps the representation of high-level visual objects is one of the other functions carried out by this piece of cortex. One possible explanation for the VWFA's responses to nameable objects could be attributed to the top-down effects on the VWFA.[10,58] Interestingly, Song et al. found that compared to nonsymbolic scenes, the VWFA responded higher to both nameable objects (e.g., chairs) and symbolic scenes (e.g., the Eiffel Tower). This representation of abstract semantic information is likely driven by top-down feedback from language regions via their connectivity[31,59–61] (see more discussion below). On the other hand, this secondary preference to objects may be in line with the neuronal recycling hypothesis,[11] which proposes that the VWFA is repurposed from other preexisting functions. Specifically, recent longitudinal studies found that the cortical tissue

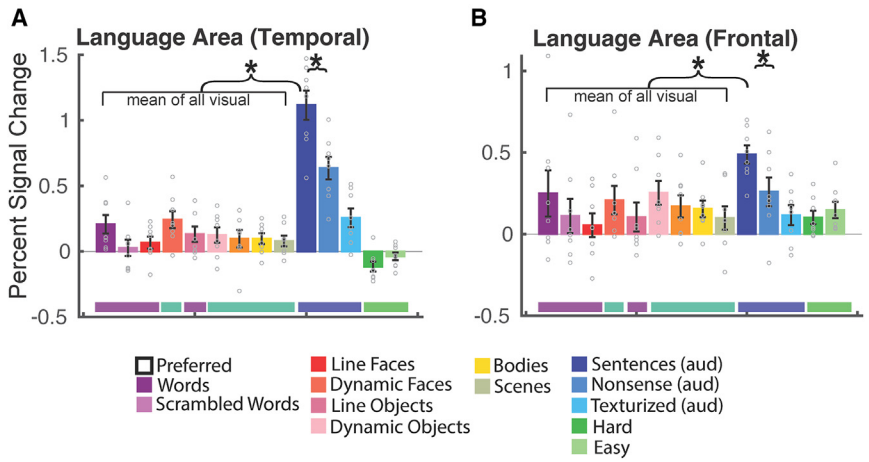


Figure 7. Response patterns of canonical language regions in participants with identical scan parameters across tasks

(A) Percent signal changes are shown for all experimental conditions for the temporal language regions.

(B) Percent signal changes were shown for all experimental conditions for the frontal language regions. Data are represented as mean ± SEM. Both the temporal and frontal language regions show language selectivity, responding significantly more to sentences than nonsense, despite temporal language regions having the lowest tSNR among all tasks (while frontal language regions showed comparable tSNR across tasks). Additionally, the sentences response is greater than the mean of all visual conditions. Wilcoxon signed rank was used due to the small sample size. Asterisks only denote significantly higher responses to auditory sentences vs. nonsense speech and sentences vs. average of all visual conditions (*$p < 0.05$). All statistical details are in the related Results section. Note that we were not able to do a version of this comparison with the full sample used in the main analysis because of the partial coverage of the original VWFA task. Also note that tasks are motion-matched across subjects. Colored boxes at the bottom note the task each condition belongs to: static VWFA localizer (purple), dynamic visual localizer (aquamarine), language (blue), and spatial working memory (SWM) (green).

later developed as the VWFA showed some initial preference for objects (e.g., tools like flashlights, similar to objects used in our study) and interestingly, also bodies/limbs.[62,63] However, even in the subset of subjects scanned with the same parameters for the static and dynamic localizers, we failed to observe robust body-selective responses throughout. Consistent with Pitcher et al.,[64] we found this lack of selectivity is mainly driven by comparable responses to faces and bodies, which might be due to the change in "visual diet" (changes in the preference for looking at hands or faces) in development.[63]

Crucially however, these secondary preferences within the VWFA were much lower than the VWFA's responses to words and also much lower than the selectivity of face and object-preferring regions like the FFA and PFS, respectively, suggesting that while this piece of tissue may be able to represent other high-level visual categories, it's response are primarily driven by word stimuli.

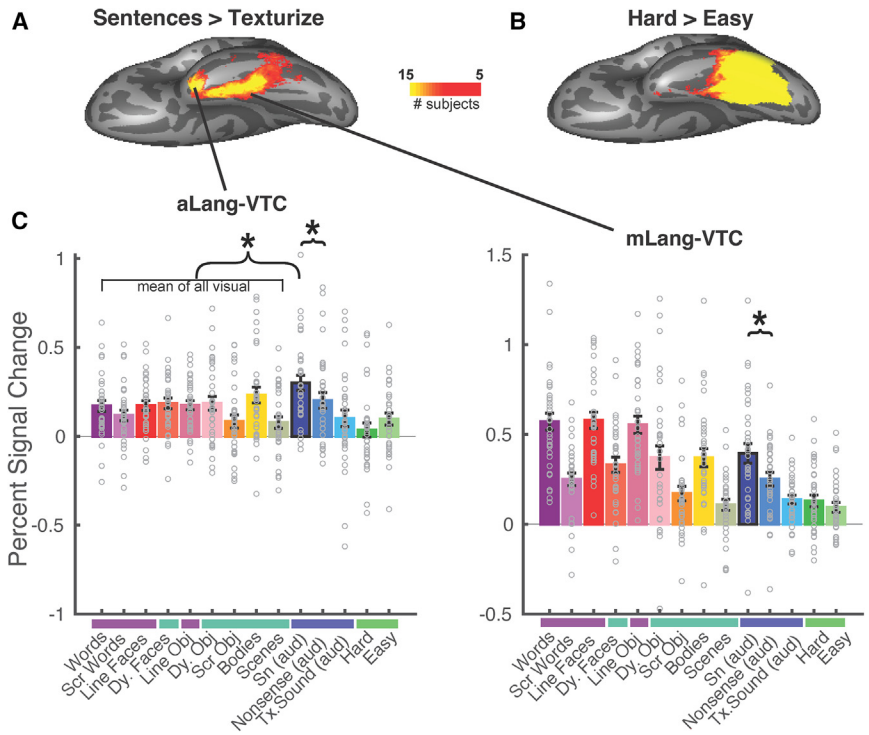


Figure 8. High-level linguistic and attentional demand effects within the left VTC

(A) Probabilistic map for Sn>Tx showing subjects with overlapping activation during the language localizer within the VTC, with two spatial clusters (mLang-VTC and aLang-VTC).

(B) Probabilistic map for Hard > Easy effect within the VTC, showing subjects with overlapping activation during the spatial working memory task across the entire VTC. For A and B, each subject's statistical map was thresholded at $p < 0.01$ and the resulting binarized maps were added together (minimum overlap = 5 subjects) (see STAR Methods for details).

(C) Functional profile of the left aLang-VTC and mLang VTC fROIs. aLang shows a preference for auditory language (with greater sentences than nonsense response), and higher percent signal change to auditory sentences than the mean of all visual conditions. mLang shows a preference for auditory language as well, and does not significantly differ between auditory sentences and the mean visual condition response. The mean percent signal change (data are represented as mean ± SEM) to various visual and non-visual conditions are plotted. Colored boxes at the bottom depict the task each condition belongs to: static VWFA localizer (purple), dynamic visual localizer (aquamarine), language (blue), and spatial working memory (SWM) (green). The preferred category (i.e., auditory sentences) is outlined in black. Individual subject PSCs are shown with gray hollow circles. Asterisks denote a significant difference between sentences and nonsense (*$p < 0.05$, pairwise test) and a significantly higher response to auditory sentences than visual categories on average for aLang (*$p < 0.05$, pairwise test). All statistical details are in the related Results section.

Word-selective responses along the posterior-to-anterior VTC

Studies have shown that distinct areas along the mid-fusiform and occipital temporal sulcus may be involved in processing different aspects of visual words.[5,39,42,65–67] For example, some work suggests a hierarchical organization of orthographic representation, becoming more abstract as one progresses more anteriorly.[68] Specifically, the posterior region at the tail of the OTS, known as the pOTS (or VWFA-1)[69] holds parallel spatial channels for two words[39] and responds to the visual features of words. However, this region at the end of the OTS has traditionally been defined by less stringent contrasts (e.g., checkerboards, phase scrambled stimuli),[39,61] that do not control for simple visual features like line segments, and is absent when using more controlled contrasts.[38] For example, this posterior region may correspond to a character-selective region,[70] which was not specific to orthography but defined together with numbers that share low-level visual features with words. Therefore, we did not include this posterior OTS region in our main analysis. These groups have also identified a more anterior region at the mid-OTS, known as the mOTS (or VWFA-2),[69] which straddles the fusiform gyrus and inferior temporal gyrus). This region, filtered by the single-word bottleneck, responds to word form and language units (Lerma-Usabiaga et al., 2018; White et al., 2019). Other studies[71] also report more anterior word-selective regions but it remains unclear whether these regions represent orthography/script or whether they represent general visual semantics/abstract concepts (e.g., another object area) because previously used fMRI contrasts were limited to words versus meaningless letter-like stimuli.

In the current study, the resulting VWFA fROIs are comparable to the previously reported VWFA locations (mOTS or VWFA-2) in studies that use similarly well-controlled contrasts.[40,57] Moreover, we found that for most of our participants, multiple word-selective patches were identified in middle and anterior OTS, aligning with observations in recent studies.[38,42] Interestingly, White et al.[40] found that more than half of their participants also had a region in the more anterior and ventral part of VTC (so-called text-mfs). This is in line with the results of our gradient analysis along the VTC (Figure 5), where we found that word-selective voxels extended from mOTS to a more anterior mid-fusiform region. Critically, the contrast used in the current study (and other studies that observe this activation) controlled for not only simple visual features but also abstract concepts (by contrasting with objects), suggesting that these mid- and anterior regions are specialized for orthographic lexicon.[72] In line with the idea of this orthographic selectivity that relies on the recognition of letter sequences of recurring word parts (rather than the meaning of words), previous studies have shown that the VWFA is sensitive to orthographic regularity (e.g., frequency of letter bigrams or trigrams) rather than lexical status (distinguishing pseudowords from real words).[68,73] General lexical or visual semantics, interestingly, might be associated with an even more anterior cluster,[65] which we will discuss further below.

Amodal linguistic activation in the left VTC

Another goal of the current study was to determine to what extent the VWFA responds to auditory language. We observed higher activation of auditory sentences vs. nonsense speech within the VWFA; but perhaps more importantly, the high-level linguistic response was significantly lower than the VWFA's response to written language (i.e., visual words) and even to non-preferred visual categories. Further, responses to auditory language within the VWFA were dwarfed by the language responses of the frontotemporal language regions. Moreover, the differences between visual and auditory stimuli were unlikely to be attributed to task design discrepancies, because 1) selectivity indices calculated across all conditions that normalized task differences showed that the VWFA's word selectivity was significantly higher than its language selectivity and 2) the two language clusters (Figure 8) in the VTC and the canonical frontotemporal language regions (Figure 7) exhibited higher or at least comparable response levels for auditory conditions as compared to visual ones. Altogether, our results suggested that the VWFA is dominated by visual stimuli, rather than a modality-independent language-related region as claimed in a recent review[35]; or at least, our result suggests that VWFA might function differently than canonical language regions as it is in fact more tuned for visual aspects of language.

Interestingly, while not part of the core language network, the VWFA is the only a-priori-defined VTC region that shows high-level linguistic sensitivity. This tuning is likely due to coactivation between the VWFA and frontotemporal language regions via privileged connectivity between them (i.e., connectivity hypothesis[74]; with empirical evidence provided by[31,60,75]). Similarly, Buckner et al.[76] observed a repetition priming effect for auditory words on the inferior temporal cortex and they further proposed that the top-down effect from frontal regions might account for this auditory activation, likely via connectivity between the VWFA and frontal regions.[31,77,78] Therefore, the connectivity between the VWFA and language regions may prepare that piece of the cortex for language-related stimuli, and with the visual experience of written language (i.e., orthographic stimuli), it further tunes for and becomes functionally selective for visual words as shown in our results here. This aligns with the idea that both connectivity and experience further shape and constrain its functional specialization.[79] Conversely, when no visual input is available, the VWFA may function as a language region that demonstrates sensitivity to grammatical complexity.[21] Surprisingly, a visual inspection of Figure 6A shows a potential speech effect in the lFFA. Further statistical analysis showed the lFFA is sensitive to speech in general (Sn/Ns>Tx, $p < 0.05$). We speculate that this result could be due to top-down influences, for example via the connectivity between FFA and speech sensitive regions within the superior temporal sulcus).[80–82] The FFA could be activated for both Sentences and Nonsense conditions due to interactions with speech areas. Previous work supports this idea, showing selective activation increases in the FFA during tasks related to recognizing identity through voices.[83,84]

In addition to privileged connectivity with the language network, the VWFA also connects with the frontoparietal MD regions[26,27,85]). This provides one possible explanation for the VWFA's activation in e.g., non-orthographic tasks,[86,87] which might be due to top-down feedback through VWFA's connectivity,[20] either by explicit task manipulations and demands[88–91] or long presentation times (e.g., 1.5s[28]). However, in the current

study, by implementing a 1-back task with a relatively fast presentation of visual stimuli (500ms), we demonstrated the VWFA's dominant role in rapid and efficient visual word perception.[92,93] And further, we show that most VTC regions are engaged more for hard versus easy conditions during the spatial working memory task, suggesting that the VWFA is not unique in this regard. Taken together, we suggest that the VWFA's responses to auditory language and cognitive effort are the result of top-down influence and connectivity from other cortices, rather than robust neural preferences to these stimuli.

A language cluster in VTC that is distinct from VWFA

Interestingly, our exploratory analysis showed that anterior to the VWFA (and anterior to VWFA-1/p-OTS and VWFA-2/mOTS), there are two language clusters within the left VTC that show linguistic selectivity, i.e., higher responses to auditory English sentences than to nonsense speech. Importantly, however, by directly comparing the functional response profile of these two regions to the VWFA as well as other VTC regions, we found these two language regions do not distinguish between different visual categories (including words, unlike the VWFA). In fact, the anterior cluster (aLang-VTC) is seated at the tip of the inferior temporal and fusiform gyrus, and likely corresponds to the temporal pole region that was previously associated with language comprehension and semantic processing[34,94] regardless of modality. Consistent with this, our results showed that aLang-VTC prefers auditory sentences more than visual stimuli. This was the only region within VTC that showed higher preferences for auditory sentences than other stimuli.

The more medial and posterior region, the mLang-VTC, however, showed comparable language activation to visual activation. While showing responses to visual categories in general, this region is likely not the domain-general visual imagery node (fusiform imagery node, FIN),[95,96] which is located in the "left posterior OTS".[97] Instead, this cluster might be the "basal temporal language area (BTLA)" that is situated between the left temporal pole and the VWFA according to Purcell et al.[34] Note though, the role and even the anatomical location of the BTLA remains unclear and the term "basal temporal language area" is often used to refer to any or all language areas in the basal temporal lobe. Nevertheless, our results provide some insight into the role of this region: instead of specifically serving as the interface between semantics and orthography per se,[34] this multimodal region may play a role in the semantic processing of both words and other visual categories.[71,98] This notion is supported by observations that the resection of the BTLA shows the strongest association with deficits in object naming compared to other VTC sites.[99]

Limitations of the study

The present study systematically examined the role of the precisely defined VWFA and provided a clear characterization of the nature of its orthographic selectivity by looking at its activity in response to visual words, other non-word visual stimuli, and spoken language.

However, limitations and open questions remain. First, as an effort to estimate a more comprehensive response profile of the VWFA, we scanned participants with multiple tasks. Ideally, we would want to test the function of the VWFA in a single task that includes as many conditions as possible to better compare between conditions. While we matched the scanning parameters (and most of the tasks were scanned within the same session), future studies should test the functionality of the VWFA with rich stimuli in the same task setting to further verify our results. Relatedly, while some studies have shown that dynamic stimuli elicited more robust responses compared to static stimuli,[100,101] others suggested this effect was mainly in the dorsal pathway.[64,102] When matching parameters, we see comparable responses between tasks in the VTC fROIs. Future studies could directly design their experiments to test whether the VWFA responds differently to static and moving stimuli. Second, we complemented our fROI analysis with a gradient analysis to further probe the anatomical location of the word-selective voxels. However, it remains unclear whether the voxels we found in the mid-fusiform and inferior temporal cortex belong to one single cluster or if they are two distinct/separate clusters. Additionally, unsmoothed data can be used in future studies since our supplementary analysis showed that smoothing might not be necessary as it did not change the functional profile of the fROIs (although unsmoothed data might yield a slightly smaller effect size). Third, while we performed a gradient analysis of category-selectivity along the VTC, our experiment was not set up to explore the progression of abstract word-form representations along the VTC. And so it remains unclear whether the VWFA processes words vs. consonant strings/pseudowords differently[40,68,103] (although see Baker et al.[7] for a comparably defined VWFA responding similarly to words and meaningless consonant strings) or whether the VWFA differentiates stimuli with different levels of orthographic regularities. Moreover, in addition to this spatial hierarchy, recent studies using intracranial recording also found evidence for the temporal dissociation for processing orthographic stimuli.[67,104] Fourth, does the VWFA respond differently to visual words compared to other human-invented visual signs (e.g., numbers, traffic signs)? For example, Changizi et al.[105] noted that there are common line configurations in human-invented visual signs which are not only presented in orthography, but also in other visual symbols; therefore, the possible sensitivity of the VWFA or other VTC regions should be explored with respect to these symbols. Finally, what does the VWFA do prior to literacy, and what computations is this neural tissue capable of? Our hope is that this paper will clarify the role of the VWFA as a high-level visual, category-specific region for words, allowing research to move forward with understanding how this region develops its specialization as a fascinating example of uniquely human neural cognition.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Zeynep Saygin (saygin.3@osu.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The data reported in this paper have been deposited at Mendeley Data (STAR Methods) and are publicly available as of the date of publication. The link to access the data repository was provided in the key resources table.
- All original codes used in the current study are available from the corresponding author upon request.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We appreciated the participation of our subjects. We would like to thank the Saygin Developmental Cognitive Neuroscience Lab members for helping with data collection and providing suggestions and feedback. We would like to acknowledge the support from the Center for Cognitive and Behavioral Brain Imaging (CCBBI) and The Ohio Supercomputer Center. This research is supported by the Alfred P. Sloan Fellowship (to Z.M.S.) and NSF Graduate Research Fellowship Program (DGE-1343012) to K.H.

AUTHOR CONTRIBUTIONS

J.L.: Conceptualization, formal analysis, methodology, writing–original full draft and editing; K.H.: Conceptualization, formal analysis, writing–original draft (sections) and editing; Z.M.S.: Conceptualization, supervision, writing–review and editing.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the author(s) (J.L.) used ChatGPT and Grammarly only to check grammar. After using these tools/services, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - fMRI tasks
  - Data acquisition and preprocessing
  - Functional regions of interest (fROIs)
  - Response time-course
  - Functional profile comparison
  - Probabilistic map
  - Gradient analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.111481.

Received: January 15, 2024
Revised: June 5, 2024
Accepted: November 22, 2024
Published: November 26, 2024

REFERENCES

1. Downing, P.E., Jiang, Y., Shuman, M., and Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. Science *293*, 2470–2473.
2. Epstein, R., and Kanwisher, N. (1998). A cortical representation of the local visual environment. Nature *392*, 598–601.
3. Grill-Spector, K., Kushnir, T., Edelman, S., Itzchak, Y., and Malach, R. (1998). Cue-Invariant Activation in Object-Related Areas of the Human Occipital Lobe. Neuron *21*, 191–202.
4. Kanwisher, N., McDermott, J., and Chun, M.M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. J. Neurosci. *17*, 4302–4311.
5. Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., and Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. Brain *125*, 1054–1069.
6. McCandliss, B.D., Cohen, L., and Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. Trends Cognit. Sci. *7*, 293–299.
7. Baker, C.I., Liu, J., Wald, L.L., Kwong, K.K., Benner, T., and Kanwisher, N. (2007). Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. Proc. Natl. Acad. Sci. USA *104*, 9087–9092.
8. Dehaene, S., Cohen, L., Morais, J., and Kolinsky, R. (2015). Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. Nat. Rev. Neurosci. *16*, 234–244.
9. Price, C.J., and Devlin, J.T. (2003). The myth of the visual word form area. Neuroimage *19*, 473–481.
10. Price, C.J., and Devlin, J.T. (2011). The Interactive Account of ventral occipitotemporal contributions to reading. Trends Cognit. Sci. *15*, 246–253.
11. Dehaene, S., and Cohen, L. (2007). Cultural Recycling of Cortical Maps. Neuron *56*, 384–398.
12. Vogel, A.C., Petersen, S.E., and Schlaggar, B.L. (2014). The VWFA: it's not just for words anymore. Front. Hum. Neurosci. *8*, 88.
13. Ben-Shachar, M., Dougherty, R.F., Deutsch, G.K., and Wandell, B.A. (2007). Differential Sensitivity to Words and Shapes in Ventral Occipito-Temporal Cortex. Cerebr. Cortex *17*, 1604–1611.
14. Xue, G., and Poldrack, R.A. (2007). The Neural Substrates of Visual Perceptual Learning of Words: Implications for the Visual Word Form Area Hypothesis. J. Cognit. Neurosci. *19*, 1643–1655.
15. Roberts, D.J., Woollams, A.M., Kim, E., Beeson, P.M., Rapcsak, S.Z., and Lambon Ralph, M.A. (2013). Efficient Visual Object and Word Recognition Relies on High Spatial Frequency Coding in the Left Posterior Fusiform Gyrus: Evidence from a Case-Series of Patients with Ventral Occipito-Temporal Cortex Damage. Cerebr. Cortex *23*, 2568–2580.
16. Mei, L., Xue, G., Chen, C., Xue, F., Zhang, M., and Dong, Q. (2010). The "visual word form area" is involved in successful memory encoding of both words and faces. Neuroimage *52*, 371–378.
17. Neudorf, J., Gould, L., Mickleborough, M.J.S., Ekstrand, C., and Borowsky, R. (2022). Unique, Shared, and Dominant Brain Activation in Visual Word Form Area and Lateral Occipital Complex during Reading and Picture Naming. Neuroscience *481*, 178–196.
18. Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. Proc. Natl. Acad. Sci. USA *107*, 11163–11170.
19. Ludersdorfer, P., Wimmer, H., Richlan, F., Schurz, M., Hutzler, F., and Kronbichler, M. (2016). Left ventral occipitotemporal activation during orthographic and semantic processing of auditory words. Neuroimage *124*, 834–842.
20. Planton, S., Chanoine, V., Sein, J., Anton, J.L., Nazarian, B., Pallier, C., and Pattamadilok, C. (2019). Top-down activation of the visuo-orthographic system during spoken sentence processing. Neuroimage *202*, 116135.

21. Kim, J.S., Kanjlia, S., Merabet, L.B., and Bedny, M. (2017). Development of the Visual Word Form Area Requires Visual Experience: Evidence from Blind Braille Readers. J. Neurosci. *37*, 11495–11504.

22. Yoncheva, Y.N., Zevin, J.D., Maurer, U., and McCandliss, B.D. (2010). Auditory Selective Attention to Speech Modulates Activity in the Visual Word Form Area. Cerebr. Cortex *20*, 622–632.

23. O'Craven, K.M., and Kanwisher, N. (2000). Mental Imagery of Faces and Places Activates Corresponding Stimulus-Specific Brain Regions. J. Cognit. Neurosci. *12*, 1013–1023.

24. Kitada, R., Johnsrude, I.S., Kochiyama, T., and Lederman, S.J. (2009). Functional Specialization and Convergence in the Occipito-temporal Cortex Supporting Haptic and Visual Identification of Human Faces and Body Parts: An fMRI Study. J. Cognit. Neurosci. *21*, 2027–2045.

25. Glezer, L.S., and Riesenhuber, M. (2013). Individual Variability in Location Impacts Orthographic Selectivity in the 'Visual Word Form Area. J. Neurosci. *33*, 11221–11226.

26. Chen, L., Wassermann, D., Abrams, D.A., Kochalka, J., Gallardo-Diez, G., and Menon, V. (2019). The visual word form area (VWFA) is part of both language and attention circuitry. Nat. Commun. *10*, 5601.

27. Vogel, A.C., Miezin, F.M., Petersen, S.E., and Schlaggar, B.L. (2012). The Putative Visual Word Form Area Is Functionally Connected to the Dorsal Attention Network. Cerebr. Cortex *22*, 537–549.

28. Vogel, A.C., Petersen, S.E., and Schlaggar, B.L. (2012). The Left Occipitotemporal Cortex Does Not Show Preferential Activity for Words. Cerebr. Cortex *22*, 2715–2732.

29. Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.A., and Michel, F. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. Brain *123*, 291–307.

30. Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S., and Kanwisher, N. (2010). New Method for fMRI Investigations of Language: Defining ROIs Functionally in Individual Subjects. J. Neurophysiol. *104*, 1177–1194.

31. Saygin, Z.M., Osher, D.E., Norton, E.S., Youssoufian, D.A., Beach, S.D., Feather, J., Gaab, N., Gabrieli, J.D.E., and Kanwisher, N. (2016). Connectivity precedes function in the development of the visual word form area. Nat. Neurosci. *19*, 1250–1255.

32. Fedorenko, E., Duncan, J., and Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. Proc. Natl. Acad. Sci. USA *110*, 16616–16621.

33. Pitcher, D., Dilks, D.D., Saxe, R.R., Triantafyllou, C., and Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. Neuroimage *56*, 2356–2363.

34. Purcell, J.J., Shea, J., and Rapp, B. (2014). Beyond the visual word form area: The orthography–semantics interface in spelling and reading. Cogn. Neuropsychol. *31*, 482–510.

35. Dębska, A., Wójcik, M., Chyl, K., Dzięgiel-Fivet, G., and Jednoróg, K. (2023). Beyond the Visual Word Form Area – a cognitive characterization of the left ventral occipitotemporal cortex. Front. Hum. Neurosci. *17*, 1199366.

36. Kherif, F., Josse, G., and Price, C.J. (2011). Automatic Top-Down Processing Explains Common Left Occipito-Temporal Responses to Visual Words and Objects. Cerebr. Cortex *21*, 103–114.

37. Caffarra, S., Karipidis, I.I., Yablonski, M., and Yeatman, J.D. (2021). Anatomy and physiology of word-selective visual cortex: from visual features to lexical processing. Brain Struct. Funct. *226*, 3051–3065.

38. Pillet, I., Cerrahoğlu, B., Philips, R.V., Dumoulin, S., and de Beeck, H.O. (2024). The position of visual word forms in the anatomical and representational space of visual categories in occipitotemporal cortex. Imaging Neurosci. *2*, 1–28. https://doi.org/10.1162/imag_a_00196.

39. White, A.L., Palmer, J., Boynton, G.M., and Yeatman, J.D. (2019). Parallel spatial channels converge at a bottleneck in anterior word-selective cortex. Proc. Natl. Acad. Sci. USA *116*, 10087–10096.

40. White, A.L., Kay, K.N., Tang, K.A., and Yeatman, J.D. (2023). Engaging in word recognition elicits highly specific modulations in visual cortex. Curr. Biol. *33*, 1308–1320. https://doi.org/10.1016/j.cub.2023.02.042.

41. Yeatman, J.D., and White, A.L. (2021). Reading: The Confluence of Vision and Language. Annu. Rev. Vis. Sci. *7*, 487–517.

42. Zhan, M., Pallier, C., Agrawal, A., Dehaene, S., and Cohen, L. (2023). Does the visual word form area split in bilingual readers? A millimeter-scale 7-T fMRI study. Sci. Adv. *9*, eadf6140.

43. Planton, S., Longcamp, M., Péran, P., Démonet, J.-F., and Jucla, M. (2017). How specialized are writing-specific brain regions? An fMRI study of writing, drawing and oral spelling. Cortex *88*, 66–80.

44. Hasson, U., Levy, I., Behrmann, M., Hendler, T., and Malach, R. (2002). Eccentricity Bias as an Organizing Principle for Human High-Order Object Areas. Neuron *34*, 479–490.

45. Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. Psychol. Rev. *94*, 115–147.

46. Sablé-Meyer, M., Fagot, J., Caparos, S., van Kerkoerle, T., Amalric, M., and Dehaene, S. (2021). Sensitivity to geometric shape regularity in humans and baboons: A putative signature of human singularity. Proc. Natl. Acad. Sci. USA *118*, e2023123118.

47. Nasr, S., Echavarria, C.E., and Tootell, R.B.H. (2014). Thinking Outside the Box: Rectilinear Shapes Selectively Activate Scene-Selective Cortex. J. Neurosci. *34*, 6721–6735.

48. Woodhead, Z.V.J., Wise, R.J.S., Sereno, M., and Leech, R. (2011). Dissociation of Sensitivity to Spatial Frequency in Word and Face Preferential Areas of the Fusiform Gyrus. Cerebr. Cortex *21*, 2307–2312.

49. Bao, P., She, L., McGill, M., and Tsao, D.Y. (2020). A map of object space in primate inferotemporal cortex. Nature *583*, 103–108.

50. Konkle, T., and Caramazza, A. (2013). Tripartite Organization of the Ventral Stream by Animacy and Object Size. J. Neurosci. *33*, 10235–10242.

51. Lanthier, S.N., Risko, E.F., Stolz, J.A., and Besner, D. (2009). Not all visual features are created equal: early processing in letter and word recognition. Psychon. Bull. Rev. *16*, 67–73.

52. Szwed, M., Cohen, L., Qiao, E., and Dehaene, S. (2009). The role of invariant line junctions in object and visual word recognition. Vis. Res. *49*, 718–725.

53. Szwed, M., Dehaene, S., Kleinschmidt, A., Eger, E., Valabrègue, R., Amadon, A., and Cohen, L. (2011). Specialization for written words over objects in the visual cortex. Neuroimage *56*, 330–344.

54. Plaut, D.C., and Behrmann, M. (2011). Complementary neural representations for faces and words: A computational exploration. Cogn. Neuropsychol. *28*, 251–275.

55. Nestor, A., Behrmann, M., and Plaut, D.C. (2013). The Neural Basis of Visual Word Form Processing: A Multivariate Investigation. Cerebr. Cortex *23*, 1673–1684.

56. Augustine, E., Jones, S.S., Smith, L.B., and Longfield, E. (2015). Relations Among Early Object Recognition Skills: Objects and Letters. J. Cognit. Dev. *16*, 221–235.

57. Kubota, E.C., Joo, S.J., Huber, E., and Yeatman, J.D. (2019). Word selectivity in high-level visual cortex and reading skill. Dev. Cogn. Neurosci. *36*, 100593.

58. Song, Y., Tian, M., and Liu, J. (2012). Top-Down Processing of Symbolic Meanings Modulates the Visual Word Form Area. J. Neurosci. *32*, 12277–12283.

59. Li, J., Kean, H., Fedorenko, E., and Saygin, Z. (2022). Intact reading ability despite lacking a canonical visual word form area in an individual born without the left superior temporal lobe. Cogn. Neuropsychol. *39*, 249–275.

60. Stevens, W.D., Kravitz, D.J., Peng, C.S., Tessler, M.H., and Martin, A. (2017). Privileged Functional Connectivity between the Visual Word Form Area and the Language System. J. Neurosci. *37*, 5288–5297.

61. Yeatman, J.D., Rauschecker, A.M., and Wandell, B.A. (2013). Anatomy of the visual word form area: Adjacent cortical circuits and long-range white matter connections. Brain Lang. *125*, 146–155.

62. Dehaene-Lambertz, G., Monzalvo, K., and Dehaene, S. (2018). The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. PLoS Biol. *16*, e2004103.

63. Nordt, M., Gomez, J., Natu, V.S., Rezai, A.A., Finzi, D., Kular, H., and Grill-Spector, K. (2021). Cortical recycling in high-level visual cortex during childhood development. Nat. Human Behav. *5*, 1686–1697. https://doi.org/10.1038/s41562-021-01141-5.

64. Pitcher, D., Ianni, G., and Ungerleider, L.G. (2019). A functional dissociation of face-body- and scene-selective brain areas based on their response to moving and static stimuli. Sci. Rep. *9*, 8242.

65. Lerma-Usabiaga, G., Carreiras, M., and Paz-Alonso, P.M. (2018). Converging evidence for functional and structural segregation within the left ventral occipitotemporal cortex in reading. Proc. Natl. Acad. Sci. USA *115*, E9981–E9990.

66. Weiner, K.S., Golarai, G., Caspers, J., Chuapoco, M.R., Mohlberg, H., Zilles, K., Amunts, K., and Grill-Spector, K. (2014). The mid-fusiform sulcus: A landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. Neuroimage *84*, 453–465.

67. Woolnough, O., Donos, C., Rollo, P.S., Forseth, K.J., Lakretz, Y., Crone, N.E., Fischer-Baum, S., Dehaene, S., and Tandon, N. (2021). Spatiotemporal dynamics of orthographic and lexical processing in the ventral visual pathway. Nat. Human Behav. *5*, 389–398.

68. Vinckier, F., Dehaene, S., Jobert, A., Dubus, J.P., Sigman, M., and Cohen, L. (2007). Hierarchical Coding of Letter Strings in the Ventral Stream: Dissecting the Inner Organization of the Visual Word-Form System. Neuron *55*, 143–156.

69. Grill-Spector, K., and Weiner, K.S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. Nat. Rev. Neurosci. *15*, 536–548.

70. Rosenke, M., van Hoof, R., van den Hurk, J., Grill-Spector, K., and Goebel, R. (2021). A Probabilistic Functional Atlas of Human Occipito-Temporal Visual Cortex. Cerebr. Cortex *31*, 603–619.

71. Jobard, G., Crivello, F., and Tzourio-Mazoyer, N. (2003). Evaluation of the dual route theory of reading: a metanalysis of 35 neuroimaging studies. Neuroimage *20*, 693–712.

72. Wimmer, H., and Ludersdorfer, P. (2018). Searching for the Orthographic Lexicon in the Visual Word Form Area. In Reading and Dyslexia: From Basic Functions to Higher Order Cognition, T. Lachmann and T. Weis, eds. (Springer International Publishing), pp. 57–69. https://doi.org/10.1007/978-3-319-90805-2_3.

73. Binder, J.R., Medler, D.A., Westbury, C.F., Liebenthal, E., and Buchanan, L. (2006). Tuning of the human left fusiform gyrus to sublexical orthographic structure. Neuroimage *33*, 739–748.

74. Mahon, B.Z., and Caramazza, A. (2011). What drives the organization of object knowledge in the brain? Trends Cognit. Sci. *15*, 97–103.

75. Li, J., Osher, D.E., Hansen, H.A., and Saygin, Z.M. (2020). Innate connectivity patterns drive the development of the visual word form area. Sci. Rep. *10*, 18039.

76. Buckner, R.L., Koutstaal, W., Schacter, D.L., and Rosen, B.R. (2000). Functional MRI evidence for a role of frontal and inferior temporal cortex in amodal components of priming. Brain *123 Pt 3*, 620–640.

77. Bouhali, F., Thiebaut de Schotten, M., Pinel, P., Poupon, C., Mangin, J.F., Dehaene, S., and Cohen, L. (2014). Anatomical Connections of the Visual Word Form Area. J. Neurosci. *34*, 15402–15414.

78. Thiebaut de Schotten, M., Cohen, L., Amemiya, E., Braga, L.W., and Dehaene, S. (2014). Learning to Read Improves the Structure of the Arcuate Fasciculus. Cerebr. Cortex *24*, 989–995.

79. Bedny, M. (2017). Evidence from Blindness for a Cognitively Pluripotent Cortex. Trends Cognit. Sci. *21*, 637–648.

80. Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. Nature *403*, 309–312.

81. Blank, H., Anwander, A., and von Kriegstein, K. (2011). Direct Structural Connections between Voice- and Face-Recognition Areas. J. Neurosci. *31*, 12906–12915.

82. Kriegstein, K.V., and Giraud, A.-L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. Neuroimage *22*, 948–955.

83. von Kriegstein, K., Kleinschmidt, A., Sterzer, P., and Giraud, A.-L. (2005). Interaction of Face and Voice Areas during Speaker Recognition. J. Cognit. Neurosci. *17*, 367–376.

84. Maguinness, C., and von Kriegstein, K. (2021). Visual mechanisms for voice-identity recognition flexibly adjust to auditory noise level. Hum. Brain Mapp. *42*, 3963–3982.

85. Vin, R., Blauch, N.M., Plaut, D.C., and Behrmann, M. (2024). Visual word processing engages a hierarchical, distributed, and bilateral cortical network. iScience *27*, 108809. https://doi.org/10.1016/j.isci.2024.108809.

86. Mano, Q.R., Humphries, C., Desai, R.H., Seidenberg, M.S., Osmon, D.C., Stengel, B.C., and Binder, J.R. (2013). The Role of Left Occipitotemporal Cortex in Reading: Reconciling Stimulus, Task, and Lexicality Effects. Cerebr. Cortex *23*, 988–1001.

87. Starrfelt, R., and Gerlach, C. (2007). The Visual What For Area: Words and pictures in the left fusiform gyrus. Neuroimage *35*, 334–342.

88. Moore, C.J., and Price, C.J. (1999). Three Distinct Ventral Occipitotemporal Regions for Reading and Object Naming. Neuroimage *10*, 181–192.

89. Phillips, J.A., Humphreys, G.W., Noppeney, U., and Price, C.J. (2002). The neural substrates of action retrieval: An examination of semantic and visual routes to action. Vis. Cognit. *9*, 662–685.

90. Wang, X., Xu, Y., Wang, Y., Zeng, Y., Zhang, J., Ling, Z., and Bi, Y. (2018). Representational similarity analysis reveals task-dependent semantic influence of the visual word form area. Sci. Rep. *8*, 3047.

91. White, A.L., Kay, K.N., Tang, K.A., and Yeatman, J.D. (2023). Engaging in word recognition elicits highly specific modulations in visual cortex. Curr. Biol. *33*, 1308–1320.e5.

92. Grill-Spector, K., Kushnir, T., Hendler, T., and Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. Nat. Neurosci. *3*, 837–843.

93. Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. Nature *381*, 520–522.

94. Binder, J.R., Desai, R.H., Graves, W.W., and Conant, L.L. (2009). Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. Cerebr. Cortex *19*, 2767–2796.

95. Spagna, A., Hajhajate, D., Liu, J., and Bartolomeo, P. (2021). Visual mental imagery engages the left fusiform gyrus, but not the early visual cortex: A meta-analysis of neuroimaging evidence. Neurosci. Biobehav. Rev. *122*, 201–217.

96. Spagna, A., Heidenry, Z., Miselevich, M., Lambert, C., Eisenstadt, B.E., Tremblay, L., Liu, Z., Liu, J., and Bartolomeo, P. (2024). Visual mental imagery: Evidence for a heterarchical neural architecture. Phys. Life Rev. *48*, 113–131.

97. Liu, J., Zhan, M., Hajhajate, D., Spagna, A., Dehaene, S., Cohen, L., and Bartolomeo, P. (2024). Visual mental imagery in typical imagers and in aphantasia: A millimeter-scale 7-T fMRI study. Preprint at bioRxiv. https://doi.org/10.1101/2023.06.14.544909.

98. Thompson-Schill, S.L., Aguirre, G.K., Desposito, M., and Farah, M.J. (1999). A neural basis for category and modality specificity of semantic knowledge. Neuropsychologia *37*, 671–676.

99. Wilson, S.M., Lam, D., Babiak, M.C., Perry, D.W., Shih, T., Hess, C.P., Berger, M.S., and Chang, E.F. (2015). Transient aphasias after left hemisphere resective surgery. J. Neurosurg. *123*, 581–593.

100. Fox, C.J., Iaria, G., and Barton, J.J.S. (2009). Defining the face processing network: Optimization of the functional localizer in fMRI. Hum. Brain Mapp. *30*, 1637–1651.

101. Russ, B.E., and Leopold, D.A. (2015). Functional MRI mapping of dynamic visual features during natural viewing in the macaque. Neuroimage *109*, 84–94.

102. Küçük, E., Foxwell, M., Kaiser, D., and Pitcher, D. (2024). Moving and Static Faces, Bodies, Objects, and Scenes Are Differentially Represented across the Three Visual Pathways. J. Cognit. Neurosci. *36*, 2639–2651. https://doi.org/10.1162/jocn_a_02139.

103. Dehaene, S., Cohen, L., Sigman, M., and Vinckier, F. (2005). The neural code for written words: a proposal. Trends Cognit. Sci. *9*, 335–341.

104. Woolnough, O., Donos, C., Curtis, A., Rollo, P.S., Roccaforte, Z.J., Dehaene, S., Fischer-Baum, S., and Tandon, N. (2022). A Spatiotemporal Map of Reading Aloud. J. Neurosci. *42*, 5438–5450.

105. Changizi, M.A., Zhang, Q., Ye, H., and Shimojo, S. (2006). The Structures of Letters and Symbols throughout Human History Are Selected to Match Those Found in Objects in Natural Scenes. Am. Nat. *167*, E117–E139.

106. Julian, J.B., Fedorenko, E., Webster, J., and Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. Neuroimage *60*, 2357–2364.

107. Blank, I., Kanwisher, N., and Fedorenko, E. (2014). A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. J. Neurophysiol. *112*, 1105–1118.

108. Weiner, K.S., Barnett, M.A., Lorenz, S., Caspers, J., Stigliani, A., Amunts, K., Zilles, K., Fischl, B., and Grill-Spector, K. (2017). The Cytoarchitecture of Domain-specific Regions in Human High-level Visual Cortex. Cerebr. Cortex *27*, 146–161.

109. Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage *31*, 968–980.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Functional parcels and localizer tasks	VWFA parcel and the static localizer (Saygin et al., 2016)[31]	https://www.zeynepsaygin.com/ZlabResources.html
	Other VTC parcels and the dynamic localizer (Julian et al., 2012)[106]	https://web.mit.edu/bcs/nklab/GSS.shtml
	Language parcels and the auditory language localizer (Fedorenko et al., 2010)[30]	https://evlab.squarespace.com/resources-all/download-parcels
	MD parcels and the spatial working memory localizer (Fedorenko et al., 2014)[32]	https://evlab.squarespace.com/resources-all/download-parcels
FreeSurfer, for structural and functional data processing	Laboratories for Computational Neuroimaging (LCN) at the Athinoula A. Martinos Center for Biomedical Imaging	https://surfer.nmr.mgh.harvard.edu/
MATLAB, for conducting experiments and customized data analysis	MathWorks	https://www.mathworks.com/?s_tid=gn_logo
FSL, for imaging data analysis	Wellcome Center for Human Neuroimaging	https://fsl.fmrib.ox.ac.uk/fsl/docs/#/
Original code	Available from the corresponding author upon request	N/A
R and R Studio, for statistical analysis	The R Foundation; Posit	https://www.r-project.org/; https://posit.co/download/rstudio-desktop/
Deposited data		
Preprocessed fMRI data for all fROIs and all conditions	This paper	https://github.com/annaq1027/VWFA_Li

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Sixty-three healthy adults were recruited from The Ohio State University (OSU) and the local community. As part of an ongoing project exploring the relationship between functional organization of the human brain and connectivity, participants completed a battery of functional tasks in the scanner. In the current study, all 63 individuals completed at least one run of at least one of the functional tasks of interest: a VWFA localizer ($N = 56$), a dynamic visual localizer ($N = 52$), an auditory language task ($N = 48$) and a spatial working memory task ($N = 57$). The full sample of a given task was used to generate probabilistic maps for a given functional contrast. Critically, only individuals who completed two runs of the VWFA localizer and the dynamic localizer as well as at least one run of the other two tasks were included in the main analysis, to ensure independence when defining the VTC category-selective regions and investigating their functional responses. Therefore, a sample of 37 fluent English speakers were included in the main analysis: mean age: 24.70, age range: 18.01–55.66; 4 left-handed (see Table S8 for demographic details). Among them, 8 subjects speak more than one language. The number of participants was relatively balanced between sexes, with 22 females and 15 males. Ancestry and socioeconomic status were not collected, and the authors do not expect these demographic details to have potential impact on the results reported here. When exploring the activation within the language and MD network, 3 individuals with only one run of the language task were further excluded. All participants had either normal vision or corrected-to-normal vision and reported no neurological, neuropsychological, or developmental conditions. The study was approved by the Institutional Review Board (study approval number: 2017H0353) at OSU and all participants provided written consent.

METHOD DETAILS

fMRI tasks

VWFA localizer (visual)

We used a VWFA localizer task (Figure 1)[31] to define the VWFA. Briefly, static images of words, scrambled words, objects, and faces were presented in blocks. In our critical visual word condition, a single printed word in black text color was shown and all words were nouns. Each word was presented on a black grid inside a white rectangular box, and for the scrambled words condition, each block of this grid was randomly arranged, to preserve low-level visual features of words (lines, curves intersections). Line drawings of faces and objects (also presented on a grid) were implemented to control for the effects of visual similarity,

complexity, and abstract semantic meaning. All stimuli were displayed on a white rectangular box and a gray grid was superimposed on the stimuli to match edges that are present in the scrambled words condition (all stimuli are available for download at https://www.zeynepsaygin.com/ZlabResources.html). For this localizer, participants were asked to perform a one-back task. In each block (18s), 26 stimuli (including 2 repetitions) of the same category were presented one-by-one for 500ms followed with a 193ms ISI. Each run contains 4 experimental blocks for each condition and 3 fixation blocks and 2 runs of the VWFA localizer were collected for each participant.

Dynamic localizer (visual)

To define other VTC category-selective regions that are adjacent to the VWFA, we used a dynamic visual localizer task where participants were asked to passively view video clips (with natural colors) from faces, objects, bodies, natural scenes and scrambled objects (Figure 1).[33] Briefly, the face condition included faces of young children dancing and playing, the object condition included different moving objects (e.g., round block toys swinging), and the body condition showed different body parts (e.g., legs, hands; no faces) naturally moving. These clips were filmed on a black background. Scrambled objects were generated by scrambling each frame of the object movie clip into a 15-by-15 grid. Scene conditions were recorded from a car window while driving through a suburb. Six 3-s video clips from the same category were shown in each block (i.e., 18s per block) and 2 blocks of each experimental condition were presented in each run (alternating with a palindromic manner) with 3 rest blocks with full-screen colors alternating at the beginning, middle and the end of each run. The order across participants was randomized.

Language task (auditory)

A language task (Figure 1)[30] was used to investigate how the VWFA, as well as other VTC regions respond to the lexical and structural properties of auditory language. Participants listened to blocks of meaningful sentences (Sn), nonsense sentences (Ns, controlling for prosody but constructed from phonemically intact nonsense words), and texturized (degraded) speech (Tx, controlling for low-level auditory features). Each run consisted of 4 blocks of each condition and three 14-s fixation blocks. Each block (18s) contained three trials (6s each); each trial ended with a visual queue to press a button. Language selective response is usually characterized by Sn>Tx or Sn>Ns, the former targets both linguistic and speech-related processing and the latter specifically targets high-level lexico-semantic information. Note that, we also defined the frontotemporal language regions (along with frontoparietal multiple-demand regions, see below) so that we can compare the response of the VWFA to auditory languages with that of the canonical language-selective regions.

Spatial working memory task

As a comparison of the high-level linguistic effect, a spatial working memory task (Figure 1)[32] with blocked easy and hard conditions was used to examine the effect of attentional demand on the VTC regions. The hard condition elicits higher activation than the easy condition due to the attentional load, which is the neural signature for the domain-general multiple-demand (MD) network.[32,107] In each trial, participants viewed a grid of six blocks. For the easy condition, three blocks would flash blue sequentially, and participants were expected to remember which of the blocks in the grid had been colored. Two grids would then appear, one with the same pattern of blocks colored as the previous sequence (the match) and the other with a different pattern of blocks highlighted (no-match), and participants used a button press to indicate the match. For the hard condition, 3 pairs of 2 blocks would flash sequentially, increasing the amount of information the participant needed to attend to and remember. Each run consisted of 3 experimental sets of 2 blocks per condition (28s each), separated by 16-s rest blocks with a fixation on the screen. The order of conditions within each block was randomized across participants.

Data acquisition and preprocessing

All structural and functional MRI images were acquired on a Siemens Prisma 3T scanner (at the Center for Cognitive and Behavioral Brain Imaging (CCBBI) at the OSU) with a 32-channel phase array receiver head coil. All participants completed a whole-head, high resolution T1-weighted magnetization-prepared rapid acquisition with gradient echo (MPRAGE) scan (repetition time (TR) = 2300 ms, echo time (TE) = 2.9ms, voxel resolution = 1.00 mm^3). A semi-automated processing stream (recon-all from FreeSurfer) was used for structural MRI data processing. Major preprocessing steps include intensity correction, skull strip, surface co-registration, spatial smoothing, white matter and subcortical segmentation, and cortical parcellation. For functional tasks, the VWFA localizer task was acquired with echo-planar imaging (EPI) sequence, TR = 2000ms, TE = 30ms, and 172 TRs. For N = 50 participants, 2-mm isotropic voxels were acquired with 100 × 100 base resolution; 25 slices approximately parallel to the base of the temporal lobe to cover the entire ventral temporal cortex. For the rest of the participants, slightly larger voxels (2.2-mm isotropic) were acquired for whole brain coverage (54 slices) with the same number of TRs to cover the whole brain. All other tasks were also collected with EPI sequence, but with TR = 1000ms, TE = 28ms, voxel resolution of 2 × 2 × 3 mm^3, 120 × 120 base resolution, 56 slices for the whole-brain coverage, and 244 TRs for the language localizer, 234 TRs for the dynamic localizer, and 400 TR for the spatial working memory localizer. To ensure that our results are not due to differences in scanning acquisition parameters, we collected an additional 2 runs of the VWFA task on 8 participants in our sample using identical parameters to the other 3 tasks, finding similar results (see Figure S6 and Table S5) for all analyses.

All tasks were preprocessed in the same manner. Functional data were motion corrected (all timepoints aligned to the first timepoint in the scan) and timepoints with greater than 1mm total vector motion between consecutive timepoints were identified (and later included in first level GLM as nuisance regressors). Data were distortion corrected, detrended, spatially smoothed (3mm FWHM

kernel for the static visual and 4mm for the other tasks and 4mm for the static visual task with identical scanning parameters), and then registered from functional space to anatomical space (using bbregister from FreeSurfer). We also replicated the definition of the VTC fROIs and the VWFA's functional profile with unsmoothed data, and confirmed that a small amount of smoothing does not affect individual fROI analysis (Figure S7). A block design with a standard boxcar function (events on/off) was used to convolve the canonical HRF (standard g function, d = 2.25 and t = 1.25), and experimental conditions for each task were included as explanatory variables. Six orthogonalized motion measures from the preprocessing stage were included as additional nuisance regressors for each task individually. Resulting beta estimates, contrast maps and preprocessed resting-state data were registered from functional, volumetric space to anatomical, surface space (using FreeSurfer's mri_vol2surf with trilinear interpolation), and then to FsAverage surface space (using FreeSurfer's mri_surf2surf) for subsequent analyses.

Functional regions of interest (fROIs)

Given the individual variability in the precise location of high-level visual regions across individuals, we used the group-constrained subject-specific method[30] to define subject-specific fROIs. Previously defined atlas (parcels) that show the typical location of the regions across large samples of neurotypical adults were used as our search spaces. The VWFA parcel was from Saygin et al.,[31] and specifically, it includes the lateral proportion of the fusiform gyrus, straddles the occipitotemporal sulcus and expands laterally to the inferior temporal gyrus, and anteriorly it covers the front end of the FG4[108] (Figure S10, panel A). This parcel marks the probable location of word-selective responses with anterior and posterior boundaries that match the search space used in similar studies[38,85] and also match the 'visual word' association map from Neurosynth, a neuroimaging meta-analysis website. All other VTC parcels were from Julian et al.[106] (Figure S10, panel B-F) (except for lOFA and lFBA, see below). Additionally, language functional parcels in the frontotemporal cortex were from Fedorenko et al.[30] (we used an updated version based on 220 participants) and MD functional parcels in the frontoparietal cortex were original parcels from Fedorenko et al.[32] (we used an updated version based on 197 participants) (see https://tinyurl.com/5e4tp67w for more details for language and MD parcels) (Figure S5). For language and MD fROIs, we averaged results by lobes and reported results in frontal and temporal language regions, as well as frontal and parietal MD regions. All parcels were originally in volumetric spaces and moved to FsAverage surface with the same method mentioned above. Note that the individual activation maps showed that for most of the subjects, the VWFA may not be contiguous, with variable number of patches that showed word selectivity. Therefore, when defining the fROIs, we did not require continuity and or cluster-based thresholding and instead selected the most significant/responsive voxels within a given search space using the statistical maps of the contrast of interest. Table 1 shows the corresponding localizer tasks and contrasts we used to define fROIs. Importantly, the most significant 150 voxels were selected and any voxels that responded significantly to multiple conditions were assigned to the contrast that they were most responsive to. To confirm our results were robust regardless of ways to define the fROIs, we also used two other widely used methods to select the voxels: a. applying a hard threshold to select the significant voxels; b. choosing the top 10% voxels within the search space. We replicated our main results in Tables S1 and S2. After fROIs were defined with one run of data, an independent run of data was used to extract the percent signal change (PSC) within each individual's fROIs to all conditions of interest (Table 1). To calculate PSC, we took each condition's beta estimate from the GLM, divided by the baseline (i.e., beta estimate for the entire task), then multiply by 100. Any PSC values that exceed +/− 3 standard deviation across subjects were marked as outliers and removed from subsequent analyses.

Based on the PSC values, we additionally calculated category selectivity indices to test for their preferred and any unexpected non-preferred. Specifically, for their preferred category selectivity, the difference between the response to the preferred condition and the average of all other conditions was calculated; for the selectivity of the non-preferred category, we calculated the difference between the condition of interest vs. all other conditions after excluding the response to their preferred category. For example, the VWFA was tested for a preference for objects (average of dynamic and static objects – average of faces (dynamics and static), bodies, scenes, scrambled words, scrambled objects divided by the sum of all conditions).

In addition to canonical frontotemporal language-selective regions, to further explore the functional properties of the clusters within the VTC that respond to auditory language and compare them to the VWFA, we additionally defined two VTC language fROIs (mLang-VTC and aLang-VTC). Specifically, we first created the search spaces for the fROIs based on the Sn-Tx probabilistic map (see below). The same GcSS method was then used to define subject-specific mLang-VTC and aLang-VTC fROIs by selecting the top 150 language-responsive (Sn>Tx) voxels within the medial and anterior parcels we created below. Note that we identified these two fROIs along with the other category-selective VTC regions so that we could assign vertices to the condition to which they showed the strongest response (to achieve better spatial specificity, e.g., the mLang-VTC parcel overlapped with the VWFA parcel but the fROIs selected within these parcels showed the highest responses to the category of interest and do not overlap).

Response time-course

We also visualize the average time course for each of the experimental conditions based on the response time series. For each subject, we extracted responses from the preprocessed 4-D time series for the corresponding time points of each condition. Data from different blocks of the same conditions was first averaged, and resulting values were normalized by the baseline responses (mean activation across all conditions in the given task). The time course for each block of a given condition is characterized by the

activation for every 2 s. For example, the TR for the VWFA localizer is 2s, and each block is 18s, therefore, there are 9 data points in total. Note that for other tasks, the TR is 1s, so we averaged the responses every 2 TR.

Functional profile comparison

To further characterize the functional distinctiveness and reliability of the VWFA's responses, we directly compare the functional profile of the VWFA vs. other VTC regions. Specifically, Pearson correlation was used to correlate each individual's VWFA functional response pattern (to all 14 experimental conditions) with the average VWFA response profile of the remaining participants. We then calculated the within-subject functional pattern similarity between the VWFA and other fROIs, and compared the between-subject VWFA-VWFA correlation (Fisher's z transformed) to the within-subject VWFA-to-other fROI correlation (Fisher's z transformed) with paired t-test.

Probabilistic map

As a sanity check and to justify the use of the functional parcels from independent studies as our search spaces when defining the functional regions of interest (see **Definition of the functional region of interest** below), we created probabilistic maps for different functions of interest. Using the contrast maps from the GLM analysis for each subject, we created probabilistic activation maps for different contrasts of interest. Specifically, for each subject, the statistical map based on all available runs of a task from a given contrast was minimally thresholded at $p < 0.01$ (uncorrected) and resulting maps were binarized. The binarized maps of all subjects were then summed. This resulted in a map where the value at each vertex indicates the number of subjects who showed significant activation for that contrast at that location. We presented any vertex that showed a consistent significant effect of the tested contrast across at least 5 subjects (which is approximately 10% of the participants - the exact percentile varied because the total number of subjects who completed each task was slightly different).

The probabilistic map for words (words vs. the average of other conditions in the task) was made with the static VWFA localizer. The probabilistic maps for other high-level visual categories were created with the dynamic localizer: face (faces vs. objects), object (objects vs. scrambled objects), body (bodies vs. objects), and scene (scenes vs. objects) (Figure S10). Our probabilistic heat maps agreed well with most of the reference parcels with a few exceptions: lFFA parcel from Julian et al.[106] was too small to cover the hot spots of face-selective activation around the left fusiform gyrus that we observed in our face probabilistic map (Figure S10). Therefore, we flipped the rFFA parcel to the left hemisphere. Additionally, given the original lOFA and lFBA parcels were missing the hot spots in the face and body probabilistic maps, we further created new lOFA and lFBA parcels based on the probabilistic maps. Specifically, we chose vertices that were consistently significant ($p < 0.01$ at the individual level) across at least 5 subjects; and we used the cluster function from FSL (5.0.10) to obtain the cluster pass this criterion. Our main results use the updated lFFA, lOFA and lFBA parcels.

Moreover, to explore possible amodal language and attentional load activation within the VTC, we used the auditory language and spatial working memory tasks to generate probabilistic maps within the VTC for language (contrasts: Sn>Tx) and attentional load (contrast: hard vs. easy). While Sn>Ns speech yielded similar hotspots as the Sn>Tx (Figure S8), we used this relatively liberal contrast so that the resulting parcel serves as a loose spatial estimation that allows us to search any possible language-selective voxels (see below). Just as how we made the lOFA and lFBA parcels, the summed map across individuals was thresholded so that at least 5 subjects showed significant Sn>Tx at a given vertex, which resulted in two continuous parcels (with the cluster function from the FSL) at the medial lVTC (mLang-VTC, anterior to the fusiform gyrus) and superior anterior lVTC (aLang-VTC).

Gradient analysis

We complemented the fROI analysis with a gradient analysis, where we explored the category selectivity of the entire VTC. Specifically, using the Desikan-Killiany cortical parcellation,[109] we first divided the VTC into the fusiform and the inferior temporal cortex, and then within each anatomical parcel, we created 10 equal sections from the posterior to the anterior (i.e., the size of each slice along the y axis is equal to the one-tenth of the parcel along that axis). With one run of the data, we identified the potential category-selective voxels within each section with a hard threshold ($p < 0.01$), and in the independent run, we extracted those voxels' responses (i.e., PSC) to each of the high-level visual categories in each section (for faces and objects, responses in the static and dynamic localizers were averaged). Note that we did not include scene selectivity in this analysis as this scene-selective in the VTC is usually more medial to all of the other VTC regions (i.e., medial to the fusiform cortex).

QUANTIFICATION AND STATISTICAL ANALYSIS

Mostly, we used paired t-tests (two-sided) to compare between conditions or fROIs to establish response selectivity and specificity. Bonferroni-Holm multiple comparison correction was used to correct the number of comparisons for each analysis. Moreover, one-way repeated-measure ANOVA (rmANOVA) was used to compare responses between fROIs as well as to examine the visual category specificity of the aLang-VTC and mLang-VTC. Follow-up post-hoc paired t-tests were used to identify the differences (and corrected). All analyses were performed on subjects who completed at least 2 runs of the two visual localizers ($N = 37$),

unless otherwise noted in the relevant Results sections. Paired t-test was run with the MATLAB ttest function or pairwise_t_test in R (for post-hoc tests after ANOVAs) and all ANOVAs were completed in RStudio (V 1.4.1717) using the anova_test function, with the subject number and fROI (or condition where applicable) as within-subject factors. Data are presented as mean ± SEM in the figures, with individual data points shown as gray circles. Asterisks indicate either the preferred condition(s) of a category-selective region (with horizontal lines indicating significant differences between bars) or significant differences between two bars (indicated by curly brackets).