NeuroImage 215 (2020) 116838

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Linguistic experience acquisition for novel stimuli selectively activates the neural network of the visual word form area

Mingyang Li [a], Yangwen Xu [b,c], Xiangqi Luo [a], Jiahong Zeng [a], Zaizhu Han [a,*]

[a] State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, 100875, China
[b] Center for Mind/Brain Sciences (CIMeC), University of Trento, Trento, 38123, Italy
[c] International School for Advanced Studies (SISSA), Trieste, 34136, Italy

ARTICLE INFO

Keywords:
Categorical specificity
Connectivity hypothesis
Language experience
Meaningless stimulus
VWFA

ABSTRACT

The human ventral visual cortex is functionally organized into different domains that sensitively respond to different categories, such as words and objects. There is heated debate over what principle constrains the locations of those domains. Taking the visual word form area (VWFA) as an example, we tested whether the word preference in this area originates from the bottom-up processes related to word shape (the shape hypothesis) or top-down connectivity of higher-order language regions (the connectivity hypothesis). We trained subjects to associate identical, meaningless, non-word-like figures with high-level features of either words or objects. We found that the word-feature learning for the figures elicited the neural activation change in the VWFA, and learning performance effectively predicted the activation strength of this area after learning. Word-learning effects were also observed in other language areas (i.e., the left posterior superior temporal gyrus, postcentral gyrus, and supplementary motor area), with increased functional connectivity between the VWFA and the language regions. In contrast, object-feature learning was not associated with obvious activation changes in the language regions. These results indicate that high-level language features of stimuli can modulate the activation of the VWFA, providing supportive evidence for the connectivity hypothesis of words processing in the ventral occipitotemporal cortex.

1. Introduction

Recent neuropsychological and neuroimaging research has revealed that the ventral occipito-temporal visual cortex in the human brain is functionally organized into different domains showing specialization for different categories, such as words, animals, and scenes (Mahon and Caramazza, 2009; Grill-Spector and Weiner, 2014; Bi et al., 2016). One key relevant question is what drives this organizational principle in the cortex. Two influential hypotheses have been proposed (Mahon and Caramazza, 2011; Hannagan et al., 2015). ***The shape hypothesis*** assumes that the category-specific distribution in the ventral visual cortex is determined by the shape information of stimuli (e.g., eccentricity, curvature; Hasson et al., 2002; Nasr et al., 2014; Srihasam et al., 2014), while ***the connectivity hypothesis*** argues that it is constrained by its innate connectivity with high-level regions (Mahon and Caramazza, 2011).

Here, we tested the above two hypotheses for the case of the visual word form area (VWFA), which is located in the middle fusiform gyrus of the left ventral visual cortex (Schlaggar and Mccandliss, 2007; Dehaene and Cohen, 2011; Carreiras et al., 2014) and shows selective activity for words stronger than other categories (faces, buildings and so on). Studies have shown that the high responses for words of the VWFA could be accounted for by physical shapes of words (e.g., line conjunctions, eccentricity; Szwed et al., 2011; Srihasam et al., 2014; Taylor et al., 2019). While other studies found that the word-specific responsivity of the VWFA occurs irrespective of low-level visual properties (Dehaene et al., 2001; Hasson et al., 2002), sensory modality (Reich et al., 2011; Striem-amit et al., 2011) and language type (Bolger et al., 2005; Tan et al., 2005). Furthermore, some studies found that the VWFA is connected to high-level language regions, such as the left superior temporal gyrus (STG)/sylvian fissure for phonological processing (Price, 2012; Dewitt and Rauschecker, 2013; Chen et al., 2019) and the left frontal lobes for semantic and syntactic processing (Sahin, 2009; Price, 2012). The location of word selectivity in the VWFA of literate 8-year-olds could be predicted from their own neuroanatomical "connectivity fingerprints" in preschool, at age 5 (Saygin et al., 2016). Although this evidence could not

* Corresponding author.
E-mail address: zzhhan@bnu.edu.cn (Z. Han).

https://doi.org/10.1016/j.neuroimage.2020.116838
Received 10 February 2020; Received in revised form 24 March 2020; Accepted 6 April 2020
Available online 13 April 2020

deny the role of visual features in VWFA, these results suggested that the activity in the VWFA might be influenced by other high-level areas rather than the pure function of visual input.

However, the previous studies might not have provided conclusive evidence to distinguish the above two hypotheses for the following reasons. (1) Most studies on the VWFA adopted native language words as stimulus materials. Given that the acquisition of a word in natural situations is always simultaneously learned in both the context of its shape and high-level language features, it is difficult to differentiate whether the emergence of the VWFA is driven by its word-like shape or by the high-level language areas. Hence, it becomes particularly important to investigate this issue by using novel stimuli for subjects. (2) Indeed, some studies have revealed that the VWFA appeared when novel visual words of nonnative languages were learned through phonology (Xue et al., 2006; Brem et al., 2014), semantics (Xue et al., 2006; Taylor et al., 2019) or even simple action (Song et al., 2012). Since these studies adopted word-like shapes and language features to be learned at the same time, they could not distinguish the above two hypotheses. (3) Although a few elegant studies have observed the VWFA selectivity of activity by using stimuli with non-word-like shapes (Song et al., 2010, 2012; Moore et al., 2013; Martin et al., 2019), the training procedure always associated a stimulus with a given actual word. It is unclear whether VWFA involvement after word learning was induced by novel stimuli or associated actual words. (4) More importantly, the training literature usually investigated the activation changes in the VWFA only by the contrast between pre- and posttraining activation in the VWFA area, without considering the learning effects for other categories. Thus, it is unknown whether the VWFA emerged only when the stimuli were trained as words (i.e., word specificity) or any objects (i.e., domain generality).

To overcome the above limitations of previous research, the present study conducted a functional MRI (fMRI) experiment to explore whether the appearance of the VWFA could be modulated by the learning of high-level language information for non-word-like novel stimuli regardless of their shapes. Specifically, we adopted a feature-associated learning paradigm in which 16 identical non-word-like meaningless figures were learned as having a word or object identity by associating the figures with word or object features, respectively. We first performed a whole-brain analysis to inspect whether the word-identity acquisition for a nonsense shape could change the activation of the VWFA area as well as other language areas. Then, to more closely examine the word-learning effects in language regions, we carried out a region of interest (ROI) analysis in the VWFA and other high-level language regions. This analysis explored whether word learning relative to object learning modulated the ROIs' activation intensity, activation pattern, and functional connectivity strength. Because our stimuli, consisting of meaningless, non-word-like shapes, were simultaneously learned as representations of words or objects but were not actual words or objects, the influence of stimuli shapes could be excluded. In this case, if word-specific learning effects could still be observed in the VWFA and the language network connecting with it, it can be inferred that the effects resulted from the language experience rather than the stimulus shape. To our knowledge, this is the first well-controlled study directly examining the connectivity hypothesis for the VWFA.

2. Materials and methods

2.1. Participants

Seventy-five college students were recruited. All were right-handed native speakers of Chinese and had a normal or corrected-to-normal vision. Forty-nine of the subjects participated in the assessments of the meaninglessness and categorical bias for novel figure stimuli. The remaining 26 subjects (average age: 22 ± 1.98 years old; 5 males) took part in the fMRI experiment and were randomly and equally divided into two groups (the word- and object-learning groups). This study protocol was approved by Beijing Normal University's Institutional Review Board. Written informed consent was obtained from all participants prior to the experiments.

2.2. Stimuli

Sixteen novel figures were created (see Fig. 1A). Each consisted of lines or arcs and was surrounded by a circle with a 150-pixel radius. To confirm that the figures were not meaningful and were not real words or real objects before they were learned, we had subjects evaluate the meaninglessness and categorical bias of the stimuli by the 49 subjects. The evaluation was performed using a 7-point grading scale (1 = very high, 4 = medium, 7 = very low) and answering three questions: how likely was the figure to be (1) a meaningful thing, (2) a real word, and (3) a real object? To compare the evaluation scores of the novel figures with those of standard stimuli for each question, we additionally added three other types of assessment materials: 5 meaningless symbols, 10 real Chinese words, and 20 real objects. Moreover, to further avoid subjects' response bias for the above stimuli, we also added other types of fillers (5 pseudo-Chinese words, 5 Korean words, 5 ancient words, and 5 abstract line drawings) (Fig. 1B). All 71 stimuli were pseudorandomly presented. The evaluation scores of the novel figures for the meaningfulness question were close to those for meaningless symbols (4.84 ± 1.02 vs. 5.07 ± 0.89; $t < 1.37$, $p > 0.15$; Fig. 1B). Moreover, the scores of the novel figures for the word- and object-likelihood questions were significantly higher than the scores of the real words (5.79 ± 1.10 vs. 1.51 ± 0.47; $t = 3.03$, $p < 0.01$) and real objects (4.84 ± 1.01 vs. 1.41 ± 0.32; $t = 4.71$, $p < 0.01$), respectively. These results indicate that the novel figures were meaningless and quite dissimilar to real words or objects.

2.3. Behavioral training procedure

A feature-based associative learning paradigm was adopted to train subjects to learn the 16 meaningless figures as word or object identities (Fig. 1C). Specifically, when a figure was learned as a word identity, it was associated with two linguistic features: one of four grammatical categories (i.e., adjective, adverb, preposition, auxiliary) and one of 16 pronunciations (*/bazo/*, */biba/*, */desa/*, */dogu/*, */gike/*, */gapo/*, */haga/*, */keso/*, */mate/*, */mude/*, */pehi/*, */poni/*, */ponu/*, */yuri/*, */woke/*, and */zume/*). In order to prevent the pronunciations from being close to real Chinese words, these pronunciations were created as bisyllabic pseudowords by combining 12 consonants and 5 vowels of Esperanto using the "espeakers" speech synthesizer software (http://espeak.sourceforge.net/). Similarly, when a figure was learned as an object identity, it was associated with two object features: one of four material properties (i.e., metal, wood, glass, and plastics) and one of 16 nonvisual semantic features from Binder et al. (2016) (12 sensorimotor features: cold, foot-operated, hand-operated, heavy, hot, insipid, light, noisy, quiet, rough, scented, and smooth; 2 spatial features: indoor, and outdoor; and 2 emotional features: delightful, and frightening).

We developed two ***feature-learning*** tasks (i.e., association and matching) in which subjects were trained to learn the figures as word or object identities and two ***feature-testing*** tasks (i.e., recognition and generation) in which subjects were assessed on how well the figures were learned (Fig. 1C). In ***the feature association learning task***, subjects were instructed to associate each figure with its two corresponding features as accurately as possible. Each figure was visually presented on the screen with the written names of the two features, except that the pronunciation feature was listened to through clicking a trumpet logo on the screen. ***The feature matching learning task*** involved simultaneously presenting a figure and two features; subjects were required to judge which feature matched the figure based on the above association task. The correct answer was provided following the subject's response. ***The feature recognition testing task*** was identical to the above matching task, except that no feedback on the correct answer was given. ***The feature generation testing task*** required the subjects to write down or speak the two learned features for each of the visually presented figures.

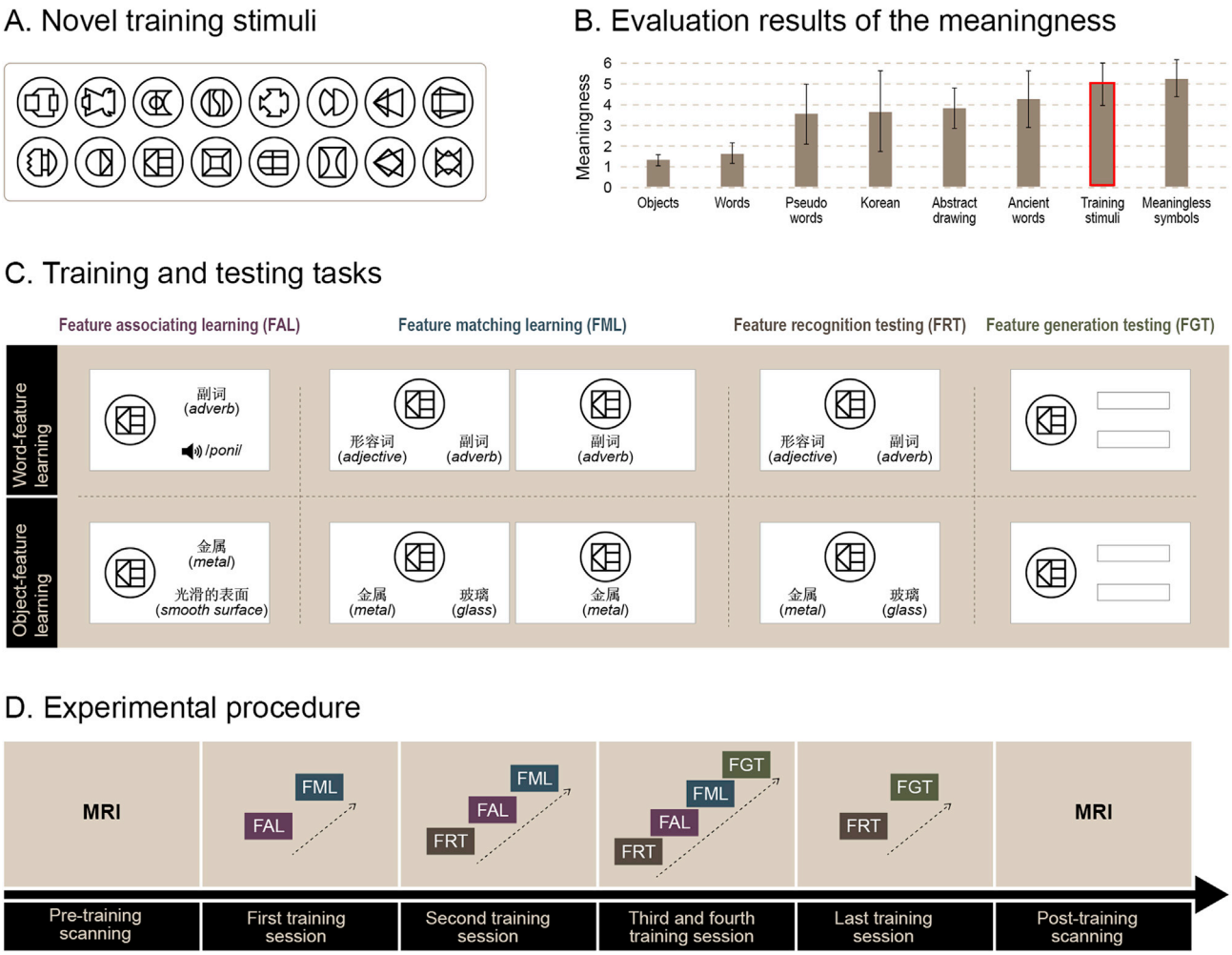

Fig. 1. Experimental stimuli and paradigm. The highlighted training stimuli in Figure B are the ones in Figure A. The full names of the tasks in Figure D are given in Figure C.

The training was completed between two MRI scans in a self-paced manner in 5 separate sessions on separate days within a span of 2 weeks (Fig. 1D). During each of the first four sessions, the two learning tasks were performed several times. The recognition testing task was also completed before the learning tasks from the second session. The generation testing task was added at the end of training in the third sessions. The final session included only the two testing tasks and occurred the day before the second MRI scan. A subject was trained for approximately 1 h per session, with the criterion that he or she mastered more than a quarter of the figures presented in the previous session. The word- and object-learning subject groups ($n = 13$, each group) performed the word- and object-learning/testing tasks, respectively.

Note that we enabled the associated features to be presented in an identical way (i.e., verbal words) between object and word categories. If, in this case, a neutrally relevant distinction between the categories was still observed, it might mainly originate from the abstract meanings of the learned features rather than the physical shapes or training their ways.

2.4. Neuroimaging data acquisition

The 26 subjects who participated in behavioral training were scanned twice (Fig. 1D) using a 3T Siemens Trio Tim scanner at the MRI center, Beijing Normal University. The images included two task-fMRI data [main experiment for investigating the learning effects and localizer experiment for defining the language ROIs] and three other types of data [3D sagittal T1-weighted magnetization prepared rapid gradient-echo imaging (3D imaging), resting-state fMRI, and diffusion-weighted imaging (DWI)]. Both the pre- and posttraining scans collected all the data with identical procedures, except that the posttraining scan did not include the language ROI localizer experiment. The resting-state fMRI and DWI data were not used in the present study.

Task-fMRI experiments. The main experiment was used to examine the learning effects when the 16 meaningless shapes were learned as word identities relative to object identities. We also designed two filler conditions: fixation and color bar trials. These conditions were incorporated into 2 runs, each consisting of 16 blocks (6 meaningless figure blocks, 6 color bar blocks, and 4 fixation blocks). Sets of figure and color bar blocks were separated by one fixation block. Each shape block included 10 stimuli, and each stimulus was visually presented 800 ms followed by a blank screen for 1200 ms. The participants performed a one-back repetition detection task in which a stimulus was presented on each trial, and the participants were instructed to press a button with their right index finger whenever the stimulus was repeated from the previous trial. In total, each figure or color bar was presented 7–8 times. In the fixation block, a fixation stimulus appeared for 14 s in the center of the screen and was passively viewed. ***The language ROI localizer experiment*** was used to predefine the language ROIs. This experiment had two runs, and each run included three main conditions (Chinese characters, line-drawing objects, scrambled line-drawing objects) and one baseline condition (fixation stimuli). The scrambled stimuli were created by randomizing the square cells of the gridded line-drawing object pictures. Each run contained 5 blocks of each main condition

and 6 blocks of the fixation condition. Each of the stimuli in the main conditions was visually presented with a fixation at the center of the screen for 350 ms followed by a 350-ms blank screen. In the fixation block, a fixation stimulus was presented for 14 s. For every condition, participants were asked to passively view the fixation stimulus.

MRI acquisition parameters. The task and resting-state fMRI data were collected with the T2*-weighted echo-planar imaging (EPI) sequence: repetition time (TR) = 2000 ms, echo time (TE) = 30 ms, flip angle = 90°, field of view (FOV) = 192 × 192 mm, slice thickness = 3.5 mm, and voxel size = 3.0 × 3.0 × 3.5 mm. ***Structural 3D data*** were obtained with the following parameters: 144 slices, TR = 2530 ms, TE = 3.39 ms, flip angle = 7°, FOV = 256 × 256 mm, slice thickness = 1.33 mm, and voxel size = 1.3 × 1.0 × 1.3 mm. ***DWI data*** were acquired in two runs. Each run contained 30 diffusion weighting directions with a b value of 1000 s/mm^2 and 1 null b value data (TR = 8000 ms, TE = 89 ms, FOV = 282 × 282 mm, slice thickness = 2 mm, and voxel size = 2.2 × 2.2 × 2.2 mm).

2.5. Functional MRI data preprocessing

The task-state fMRI data were analyzed with SPM12 (http://www.fil.ion.ucl.ac.uk/spm/). The first 10 s of each run were excluded from the analysis to allow for the initial stabilization of the fMRI signal. The preprocessing procedure included slice timing, motion correction, normalization to the Montreal Neurological Institute (MNI) space and smoothing with an isotropic 8-mm full width at half-maximum (FWHM) Gaussian kernel. For each individual subject, the data were first high-pass filtered and then fitted by a general linear model (GLM) modeled by a boxcar regressor matching its time course to estimate the condition effect. Two runs from 2 subjects (one per subject group) were excluded from our analyses due to excessive head motion (>2 mm) and rotation (>2°).

2.6. Word-learning effects in the whole brain

For each voxel in the gray matter atlas of the DPABI package (http://rfmri.org/dpabi), we performed a 2 * 2 mixed analysis of variance (ANOVA). The independent variables had a within-subject factor (i.e., training session: pre-vs. posttraining) and a between-subject factor (i.e., subject group: word-vs. object-learning group). The dependent variable was the percent signal change (PSC) of average blood oxygenation level-dependent (BOLD) on the voxel in the main fMRI experiment. The false discovery rate (FDR) was applied to correct the multiple comparisons ($q < 0.05$, cluster size > 30 voxels).

2.7. Word-learning effects in the language ROIs

For a more detailed investigation of the word-learning effects in language regions, we inspected whether the word learning relative to the object learning changed 1) activation intensity of the VWFA and other language ROIs, 2) activation pattern of them, and 3) functional connectivity strength between them. The following will first describe how the language ROIs were obtained and then describe how the effects were analyzed.

Definition of language ROIs. The ROIs included the VWFA and other high-level language regions in the left hemisphere. These ROIs were extracted through our localizer fMRI data using common approaches (Glezer et al., 2009). ***1) The VWFA ROI.*** This ROI was first created at the individual subject level with original size (individual i.e., original-sized individual VWFA, or original-sized iVWFA) and extracted as the largest cluster responsive to words versus scrambled objects (FDR corrected, $q < 0.05$) masked by the contrast of words versus line-drawing objects in the left ventral visual cortex for each subject ($p < 0.05$, uncorrected). If the VWFA cluster of a subject could not be defined or had less than 5 voxels under this threshold, we extracted the cluster responsive to words versus scrambled objects located around the left middle fusiform gyrus (FDR corrected, $q < 0.05$). In addition, we also created 6 other ROIs for the VWFA. To further investigate whether the above effects observed in the original ROI were confounded by the different VWFA ROI sizes across subjects (which ranged from 5 to 63 voxels), each ROI was created as three same sizes of spheres whose centers were located in the peak voxel of the corresponding cluster: 6-mm–, 9-mm– and 12-mm–radius spheres (i.e., spherical individual VWFA, or spherical iVWFA). To further delineate the functional differences across the ventral visual cortex, we also anatomically created a set of 9-mm radius spherical VWFA ROIs along the *y*-axis of the ventral visual cortex on the basis of the literature (Cohen et al., 2000; Lerma-Usabiaga et al., 2018): the posterior VWFA (pVWFA; center coordinates: 45 -72 -10), classic VWFA (cVWFA: 43 -54 -12), and anterior VWFA (aVWFA: 43 -48 -12). ***2) Other language ROIs.*** They were extracted at the group level on the localizer fMRI data and were the clusters for the contrast of words versus line-drawing objects in the left hemisphere (FDR corrected, $q < 0.05$; cluster size > 20 voxels; Vinckier et al., 2007).

Activation intensity in language ROIs: voxelwise univariate analysis. This analysis was used to find the ROIs whose activation magnitude significantly changed due to the acquisition of language experience. For each ROI, we first separately investigated the training effect (pre- vs. posttraining session) in each subject group. Using a *paired-samples t-test*, we compared the mean values of the BOLD PSC in all voxels of the ROI between the two sessions. We then compared the training effect between the two subject groups. Using 2 * 2 mixed ANOVA, we examined the interaction effects between training sessions and subject groups.

Activation patterns in language ROIs: multivoxel pattern analysis (MVPA). MVPA was used to further complement and confirm the findings in the above univariate analysis (Haxby et al., 2001; Norman et al., 2006; Jiang et al., 2018). It examined whether the word-identity learning for novel figures elicited a more similar activation pattern to that of real words in each ROI. This was done by comparing the similarity values of activation patterns when viewing real words and novel figures between pretraining and posttraining. Specifically, for each ROI, we first extracted four BOLD PSC values for viewing novel figures (from two factors) for each voxel in the ROI on each subject: 2 training sessions (pre-vs. posttraining) * 2 subject groups (word-vs. object learning). Then, we extracted the BOLD PSC value for viewing real words in the localizer fMRI experiment. Third, each value of novel figures was correlated with the value of real words for each subject across the voxels in the ROI, and the correlation efficiencies were Fisher *z*-transformed. Fourth, the training effects of each subject group were examined by comparing the transformed coefficients between the pre- and posttraining scans in each subject group. Finally, the interaction between the two factors was tested.

Because this and the following analyses required an ROI with the same size across subjects, the original-sized individual VWFAs were not used.

Connectivity strength between the VWFA and other language ROIs: psychophysiological interaction (PPI) analysis. During a cognitive fMRI task, PPI analysis can be implemented to find task-related changes in the functional connectivity between two brain areas (Friston et al., 1997; Reilly et al., 2012). It was adopted here to investigate whether word-learning training modulated the strength of functional connectivity between the VWFA and other language ROIs. For each spherical VWFA ROI, we first extracted the time course of each main fMRI experimental condition in the ROI as a seed using the SPM12 PPI Toolbox. Then, the time course of each experimental contrast was created for the training stimuli in the pre- and posttraining scans (training stimuli vs. baseline, respectively). The time course of the experimental contrast was then convolved with the hemodynamic response function (HRF) to generate an HRF-convolved-task-time course. Second, the PPI variables were created as the element-by-element product of the HRF-convolved-task-time course and the seed ROI time course. Third, a new GLM was established with the first regressor as the PPI variable created in the above procedure. The significance of this regressor

indicated whether activity in the region was more correlated with activity in the VWFA when viewing the novel figures. Fourth, for each subject group, we performed a planned contrast to investigate the training effect, in which we compared the mean regressor value of the PPI variable for all voxels in the higher-order language ROI individually for each subject between pre- and posttraining scans. Finally, a mixed two-factor ANOVA was applied to compare the training effect in the two subject groups.

3. Results

3.1. Behavioral performance

Two groups of subjects ($n = 26$) were trained to associate the same 16 figures with high-level features of either words (pronunciation and word class) or objects (material property and other semantic features of objects) (Fig. 1). Behavioral performance in the two testing tasks is reported in Fig. 2. For the accuracy of the feature recognition testing task, the training of two weeks engendered a gradually increasing tendency across training sessions (the main effect of the training session: $F = 18.48$, Greenhouse-Geisser corrected, $p < 10^{-7}$). The subjects could almost make correct responses for all the items in the final session (word-learning group: 94% ± 7%; object-learning group: 98% ± 3%). The main effect of subject group was not significant ($F = 2.45$, $p > 0.13$). The interaction between the training session and the subject group was also nonsignificant (Greenhouse-Geisser corrected, $F = 2.38$, $p > 0.10$). A similar result pattern was also observed for the reaction time in the feature recognition testing task and the accuracy in the feature generation testing task. Note that the reaction time in the generation task was not analyzed here because this task needed to spend a long time to complete, and reaction time on this task might not meaningfully reflect subjects' learning ability.

These results suggest that our subjects could successfully acquire word and object identities for the novel figures before the second scan. Their learning efficiency was not distinctive between the two learning groups.

3.2. Word-learning effects in the whole brain

We performed a whole-brain two-way ANOVA: 2 training sessions (pre vs. posttraining) * 2 subject groups (word-vs. object-learning group) (FDR corrected, $q < 0.05$, cluster size > 30 voxels). The main effect of the training session obtained eight survival clusters (Fig. 3). Five clusters showed increased BOLD signal strength: bilateral precuneus (peak MNI coordinates: 18 −57 27, $t = 7.25$, 429 voxels), bilateral postcingulate area (peak coordinates: 3 −3 42, $t = 4.41$, 41 voxels), left inferior frontal (peak coordinates: −36 33 12, $t = 4.92$, 72 voxels), left supplementary motor area (coordinates: −6 9 69, $t = 4.40$, 47 voxels) and right middle frontal area (peak coordinates: 33 48 27, $t = 4.62$, 33 voxels). The other three clusters presented decreased BOLD signal strength: the left inferior occipital gyrus (peak coordinates: 45 −69 −3, $t = -5.78$, 141 voxels), right posteroinferior temporal area (peak coordinates: 45 −66 −6, $t = -5.86$, 146 voxels), and right postCG (peak coordinates: 42 –33 42, $t = -5.86$, 93 voxels). The strongest training effect was observed in the bilateral precuneus. This region is localized in the posterior part of the default mode network (DMN) (Raichle et al., 2001) and is associated with successful recollections of previously processed items (Raichle, 2015). In the present study, the region may contribute to a general recollection for the learned figures regardless of the high-level features, while other significant areas may be involved in the general learning effect, such as the ventral area for visual familiarity and the frontal area for executive control.

However, no significant effect was found in the main effect of the subject group and the subject group * testing session interaction (FDR $q > 0.05$ or cluster size < 30 voxels).

3.3. Activation intensity in language ROIs

Figs. 4 and 5 illustrate the results of voxelwise univariate analysis for activation intensity in the VWFA and other language ROIs, respectively.

The VWFA. The original-sized individual VWFA ROI (mean MNI coordinates of peak points: $x = -42.81 \pm 4.99$, $y = -59.08 \pm 8.77$, $z = -15.69 \pm 3.09$) (Fig. 4A) was close to the classic VWFA in the literature (MNI coordinates: −45, −57, −12; Cohen et al., 2000; Lerma-Usabiaga et al., 2018). Furthermore, the coordinate values of the VWFA peaks between the two subject groups were insignificantly different in all three directions (word-learning group: −42.75 ± 3.86, −58.50 ± 10.35, −15.25 ± 3.49; object-learning group: −42.75 ± 6.23, −59.00 ± 9.73, −16.50 ± 2.88; $ps > 0.30$). Similarly, the cluster sizes of the VWFA also had no significant group difference (27.00 ± 18.16 voxels vs. 27.50 ± 20.83 voxels; $p > 0.95$). These results demonstrate that the original VWFA ROI defined by our localizer fMRI data was reliable and homogenous between the two subject cohorts.

In the above original-sized individual VWFA ROI, the activation intensity (i.e., BOLD PSC value) for viewing the 16 figures significantly decreased when the figures were learned as a word identity (pretraining: 2.08 ± 1.29; posttraining: 1.39 ± 1.04; $t = -3.81$, $p < 0.003$). In contrast, the PSC value did not change when the figures were learned as an object identity (pretraining: 2.09 ± 1.11; posttraining: 1.76 ± 0.95; $t = 1.20$, $p > 0.25$). There was no significant interaction between training session and subject group ($F = 1.26$, $p > 0.25$; Fig. 4B). Furthermore, the PSC value of posttraining in the ROI was significantly correlated with the behavioral reaction time in the final session during the feature matching task when the novel stimuli had been learned as words ($r = -0.65$, $p < 0.03$) but not as objects ($r = -0.01$, $p > 0.70$). However, the correlations of word learning were not significant for the behavioral accuracy in the two testing tasks ($r = -0.41$ and 0.29, $ps > 0.15$). These null results might be because the accuracies in the two tasks were so high (word learning: 94% ± 7%; object learning: 97% ± 3%) that they had not enough variation to yield obvious correlations with brain signal.

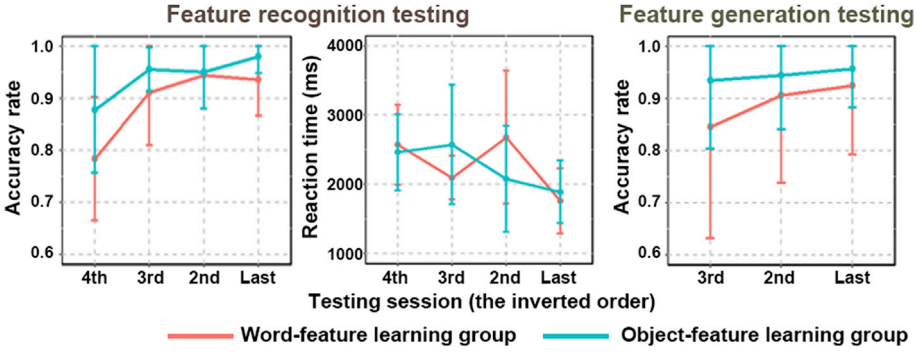

Fig. 2. The training performance in two testing tasks. The x-axis is ordered by the inverted time from the last testing. The error bars represent the standard deviations.

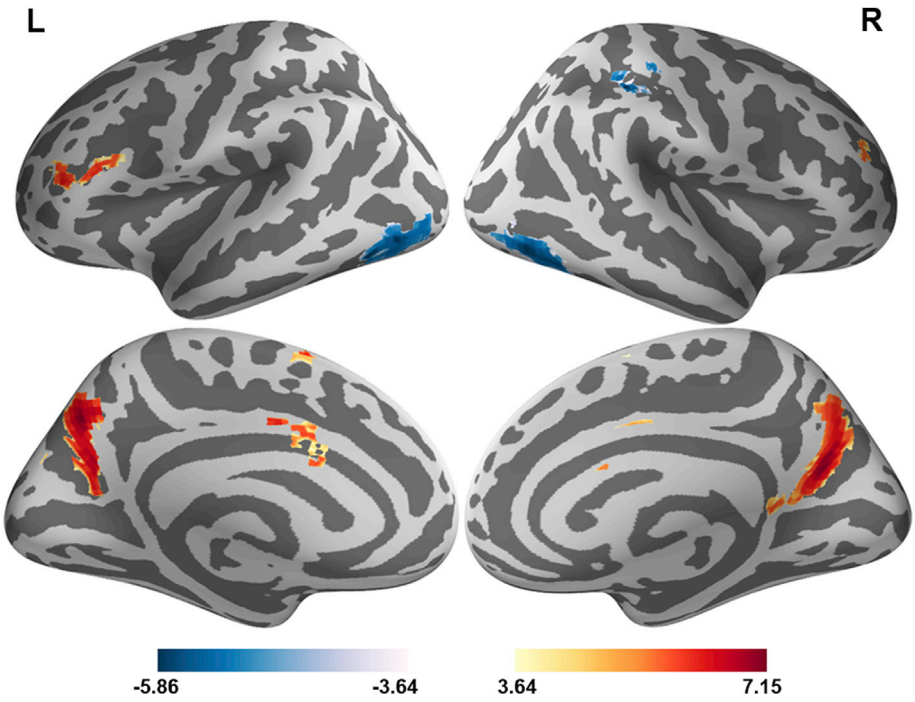

Fig. 3. The main effects of training (pretraining vs. posttraining) in the whole-brain ANOVA.

Furthermore, the effects observed in the above original-sized individual VWFA were maintained well when the VWFA was used as spherical individual VWFA with 6-mm–, 9-mm– or 12-mm–radius. The difference in the PSC values between the pre- and posttraining sessions was still significant for word learning in the new ROIs ($t = -3.76$ to -3.62; $ps < 0.004$) but not for object learning ($t = -1.01$ to -0.50; $ps > 0.30$) or for the interaction between the two groups ($F = 1.57$ to 2.11; $ps > 0.15$). The PSC value during the posttraining in each spherical ROI was also significantly correlated with the reaction time of the final session in the feature matching task for the word-learning group ($r = -0.81$ to -0.67, $ps < 0.02$) but not the object-learning group ($r = -0.08$ to 0.03, $ps > 0.80$). Other behavior-brain correlations were also nonsignificant ($r = -0.32$ to 0.26, $ps > 0.30$). We illustrate only the results of the 9-mm sphere for simplicity (Fig. 4B and C), and it is also true for the following result presentation at the individual subject level since the result patterns of the spherical VWFA ROIs across the three different sizes were highly similar.

For the three literature-based VWFA ROIs, the main effects of training session for activation intensity in each ROI were significant or marginally significant for word-training subjects (aVWFA: $t = -1.97$, $p < 0.075$; cVWFA: $t = -2.77$, $p < 0.015$; pVWFA: $t = -4.45$, $p < 0.001$) but nonsignificant for object-training subjects ($ts = -1.19$ to -0.12, $ps > 0.25$). The training session and subject group interaction was marginally significant in the pVWFA ($F = 3.71$, $p < 0.07$) but not significant in the other two ROIs (aVWFA: $F = 0.96$, $p > 0.30$; cVWFA: $F = 1.36$, $p > 0.25$). However, the activation intensity in these VWFA ROIs was not significantly correlated with the behavioral performance in the final testing session for each subject group ($ps > 0.30$).

In summary, we observed that the learning of a nonsense word-unlike new shape to have a word identity selectively changed the activation intensity in the VWFA. The learning ability of subjects could predict the activation intensity of the region. Namely, the activation level of the VWFA could be modulated by the language-identity learning experience for unfamiliar stimuli irrespective of their external shapes.

Other language ROIs. Following the approach of Vinckier et al. (2007), the language ROIs were extracted as the clusters in the left hemisphere, which showed a higher activation intensity for viewing words than line-drawing objects in our localizer fMRI data (FDR corrected, $q < 0.05$; cluster size > 20 voxels). We obtained five clusters: the left anterior temporal lobe (ATL; center coordinate: −51 0 −6; 92 voxels), posterior superior temporal gyrus (pSTG; center coordinate: 51 −52 15; 87 voxels), precentral gyrus (preCG; center coordinate: −60 0 21; 71 voxels), postcentral gyrus (postCG; center coordinate: 48 −12 45; 116 voxels), and supplementary motor area (SMA; center coordinate: −3 0 63; 114 voxels). Word-feature learning for the 16 nonsense stimuli induced increased BOLD PCS values in three ROIs (pSTG: -0.01 ± 0.50 vs. 0.51 ± 0.59, $t = 2.66$, $p < 0.025$; postCG: 0.60 ± 0.64 vs. 1.28 ± 0.56, $t = 2.89$, $p < 0.015$; SMA: 0.50 ± 0.66 vs. 1.33 ± 0.64, $t = 3.26$, $p < 0.007$) but not in the other two ROIs (ATL: -0.25 ± 0.71 vs. -0.04 ± 0.64, $t = 1.22$, $p > 0.20$; preCG: 0.20 ± 0.46 vs. 0.34 ± 0.56, $t = 0.66$, $p > 0.50$). In contrast, object-feature learning for the stimuli did not give rise to a significant change in the PCS value in any ROIs ($t = 0.25$ to 1.77; $ps > 0.10$). Among these ROIs, only the postCG showed a marginally significant interactive effect between the subject group and training session ($F = 3.40$, $p < 0.08$; other ROIs: $Fs < 2.27$, $ps > 0.14$). These results indicate that word-feature learning improved the activation level in three high-level language regions (i.e., the left pSTG, postCG, and SMA).

3.4. Activation pattern in language ROIs

Fig. 6 displays the results of activation pattern in language ROIs.

The VWFA ROI. For each of the six spherical ROIs (three from the individual level, three from the literature), the correlation coefficients of the activation pattern in the ROI between viewing novel figures and real words were marginally significant in the posterior VWFA ($t = 2.03$, $p < 0.07$), 6-mm-radius individual ROI ($t = 2.08$, $p < 0.07$) and 12-mm-radius individual ROI ($t = 1.96$, $p < 0.08$). Object identity learning showed a significant or trending decrease from pretraining to posttraining in all three individual ROIs (6-mm-radius: $t = 1.84$, $p < 0.10$; 9-mm-radius: $t = 2.36$, $p < 0.04$; 12 mm-radius: $t = 2.39$, $p < 0.04$). The training session * subject group interaction did not reach the significance level for any ROIs ($Fs < 0.38$, $ps > 0.50$). This suggests that the learning for novel figures might enlarge the difference of activation pattern between new figures and real words in the ventral visual cortex in both conditions.

Other language ROIs. Word-identity learning caused a significant

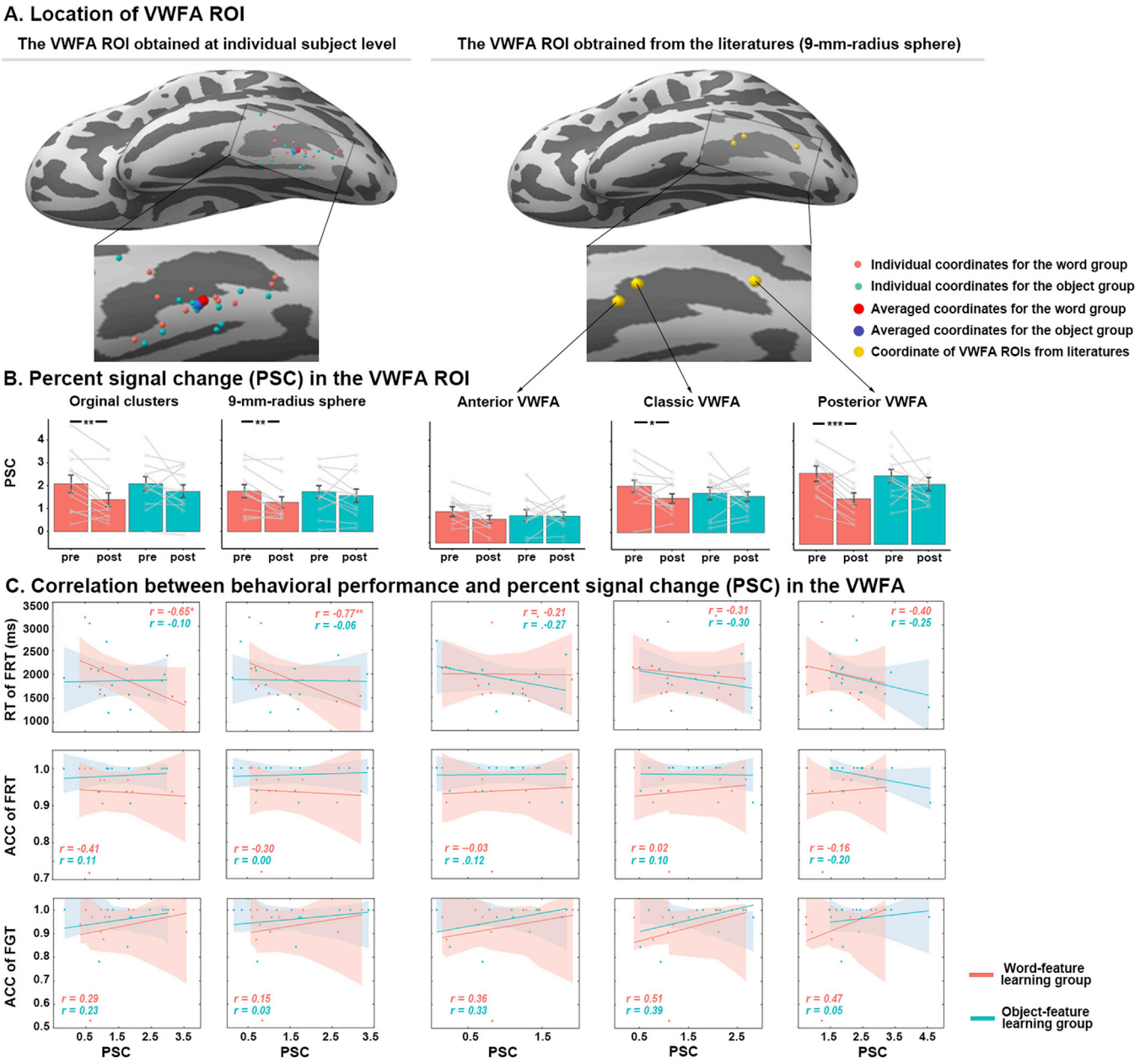

Fig. 4. The VWFA ROIs obtained by different approaches and the training effects on activity in the ROIs. In Figure C, the gray lines show the trend of PSC from pre-to posttraining scans for each subject, and the error bars show the standard error (s.e.m). *: $p < 0.05$; **: $p < 0.01$. In Figure C, the shadow shows the 95% confidence interval of the fitted line. ACC = accuracy; FGT = feature generation testing; FRT = feature recognition testing; PSC = percent signal change; ROI = region of interest; RT = reaction time; VWFA = visual word form area.

increase in the correlation coefficients of the activation pattern between novel figures and real words in the left SMA (pretraining: 0.46 ± 0.19; posttraining: 0.55 ± 0.26; $t = 2.40$, $p < 0.04$), the postCG (pretraining: 0.32 ± 0.32; posttraining: 0.58 ± 0.25; $t = 3.46$, $p < 0.005$), and the pSTG (pretraining: 0.21 ± 0.27; posttraining: 0.39 ± 0.26; $t = 2.56$, $p < 0.03$). The correlation coefficients did not obviously change in the other two regions ($ts = 0.75$ and 1.27, $ps > 0.45$). Object-feature learning did not elicit any significant change in the correlation coefficients in each ROI ($ts = 0.14$ to 0.79, $ps > 0.50$). The interaction between training session and subject group was significant in the postCG ($F = 5.63$, $p < 0.03$) but not in the other four ROIs ($Fs = 0.004$ to 1.66, $p > 0.20$). These results demonstrate that language experience acquisition for a novel figure made the activation in the language regions (the left SMA, postCG, and pSTG) not only higher but also more similar to the activation elicited by a real word.

3.5. Connectivity strength between the VWFA and other language ROIs

The above analyses have revealed that the word-identity acquisition for novel figures selectively modulated the activation intensity/pattern in both the VWFA and other high-level language regions. PPI analysis here further explored whether this word-learning selectivity also occurred in the functional connectivity strength between the VWFA and the language ROIs (Fig. 7). For the individual spherical VWFA, the word learning for novel figures led to a significantly increased functional connectivity strength of this region with the left pSTG (pretraining: 0.25 ± 0.45, posttraining: 0.48 ± 0.47; $t = 2.32$, $p < 0.05$) but not with each of the other four language ROIs ($ts = -0.62$ to 0.83, $ps > 0.40$). The object learning for the figures caused marginally significant increases in the connection strength of the VWFA with the left pSTG ($t = 1.97$, $p < 0.08$) and SMA ($t = 2.18$, $p < 0.06$) but not the other three ROIs ($ts = 0.79$ to

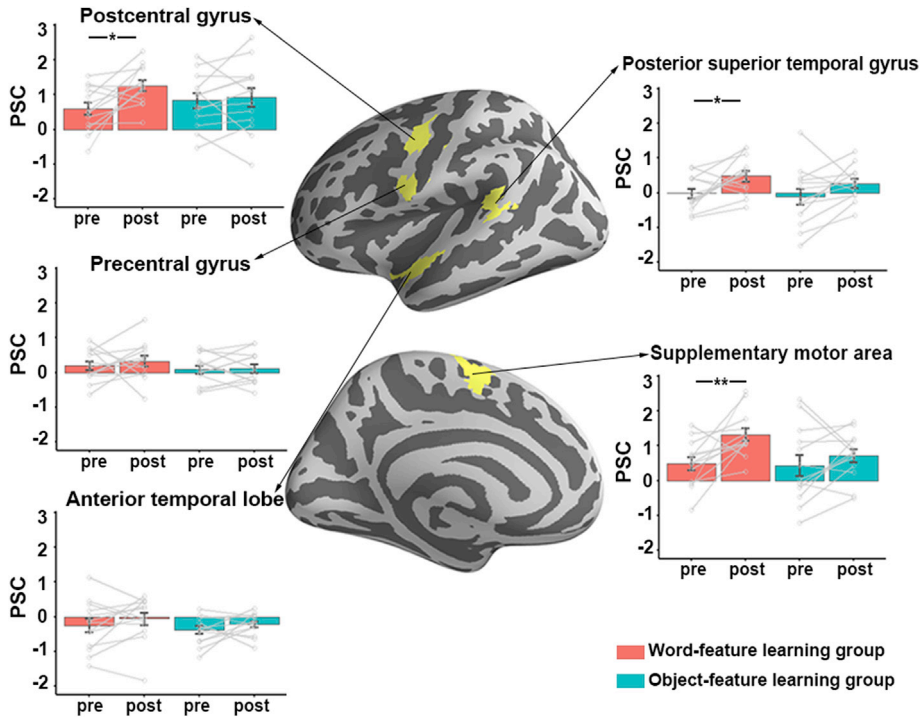

Fig. 5. The training effects on activity in the language ROIs. The ROIs (yellow patches) were defined by the contrast of words versus objects (*FDR* corrected $q < 0.05$, cluster size > 50 voxels). The gray lines show the trend of PSC from pre-to posttraining scans for each subject, and the error bars show the standard error (s.e.m). *: $p < 0.05$; **: $p < 0.01$. *FDR* = false discovery rate; PSC = percent signal change; ROI = region of interest.

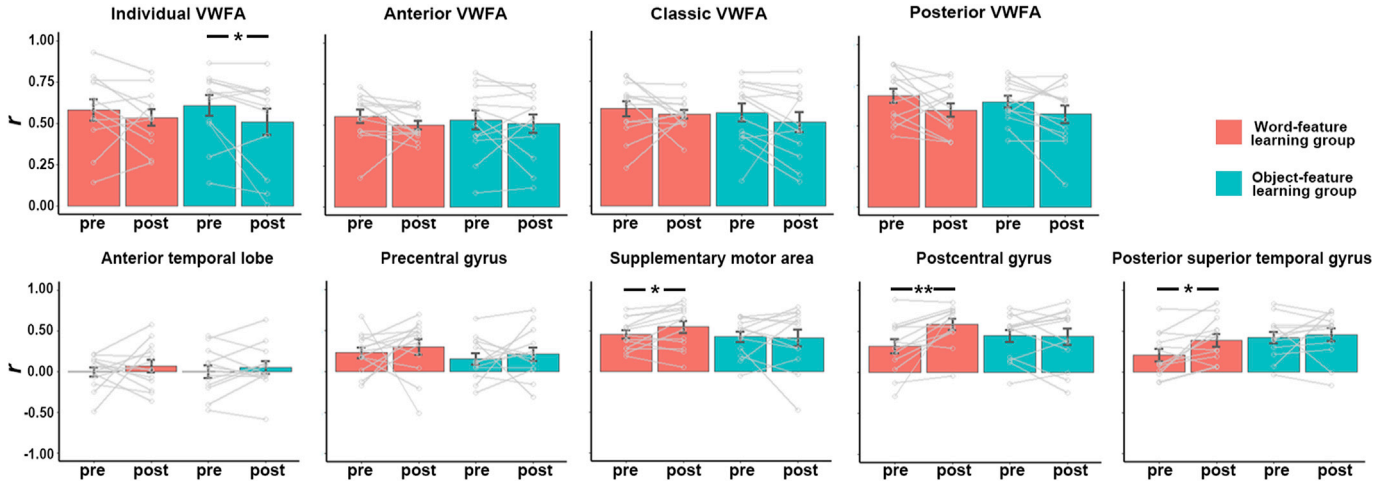

Fig. 6. Results of the MVPA in the language ROIs. Each bar represents the correlation coefficient of PSC values of the voxels in the area across subjects between viewing novel training figures (pretraining or posttraining) and real stimuli (words or objects). The gray lines show the trend in correlation coefficients from pre-to posttraining scans for each subject, and the error bars show the standard error (s.e.m). *: $p < 0.05$. MVPA = multivoxel pattern analysis; PSC = percent signal change; ROI = region of interest.

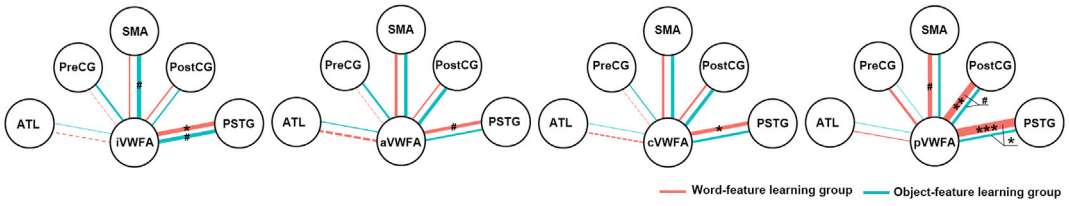

Fig. 7. Results of psychophysiological interaction (PPI) analysis for the connection strength between the VWFA and other language ROIs. Each line represents the paired-sample *t* value between pre- and posttraining PPI connectivity between the VWFA and the language ROI. The solid and dashed lines denote the positive and negative values, respectively. The width of the line reflects the absolute value of the value. The lines that point to two connections in two training groups indicate the interactive effect between training groups and training sessions. ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, #: $p < 0.10$.

1.13, $ps > 0.25$). The interaction between training session and subject group was not significant for any connection between the VWFA and each of the other five language ROIs ($Fs = 0.004$ to 1.46, $p > 0.20$).

For the three literature-obtained VWFA ROIs, the word learning for novel figures gave rise to a gradually reduced tendency for the functional connection strength along the posterior-to-anterior direction. For the connection between the VWFA and the left pSTG, a significant increase in strength occurred in the pVWFA ($t = 4.57$, $p < 0.001$), followed by the cVWFA ($t = 2.26$, $p < 0.05$) and the aVWFA ($t = 1.87$, $p < 0.09$). Similarly, for the VWFA-postCG connection, a significant strength increase was present only in the pVWFA ($t = 3.35$, $p < 0.01$) and not in the cVWFA ($t = 0.93$, $p > 0.35$) or aVWFA ($t = 0.74$, $p > 0.45$). For the VWFA-SMA connection, we observed only a marginally significant strength increase in the pVWFA ($t = 2.13$, $p < 0.06$) but not for the cVWFA ($t = 1.21$, $p > 0.25$) and aVWFA ($t = 1.48$, $p > 0.16$). The strength of other connections did not significantly change from word learning ($ts = -1.17$ to 1.34, $p > 0.25$). In contrast, the object learning for novel figures did not change the strength of connectivity between the VWFA and the language ROIs ($ts = 0.05$ to 1.71, $p > 0.10$). More importantly, word learning induced more significantly increased connectivity strength of the pVWFA with the pSTG ($F = 5.83$, $p < 0.03$) and postCG ($F = 3.70$, $p < 0.07$) than object learning for the same figure.

These results indicate that word-identity acquisition for a meaningless shape not only increased the activation level of the three language regions themselves (i.e., the left pSTG, postCG, and SMA) but also enhanced the connection strength between each of those regions and the VWFA.

4. Discussion

The word-processing selectivity of the VWFA has been speculated to originate from bottom-up physical information, i.e., the word-like physical shape of stimuli (shape hypothesis), or from top-down linguistic information, i.e., feedback connectivity from high-level language regions (connectivity hypothesis) (Hannagan et al., 2015). To examine the two hypotheses, we trained two groups of subjects to associate the same word-unlike meaningless figures with the high-level word or object features. We observed that word-feature learning for the figures elicited changes in neural activation intensity and pattern in the VWFA. Moreover, the word-learning efficiency of the subjects could successfully predict the activation intensity of this area after training. Word-learning effects were also observed in three high-level language areas (the left pSTG, postCG, and SMA). Furthermore, this word learning increased the strength of functional connectivity between the VWFA and the language regions. The growth of connection strength presented a gradual declining tendency from posterior to anterior along the VWFA. In contrast, object-feature learning did not result in obvious activity changes in the language regions. These findings demonstrate that the neural activation of the VWFA can be modulated by the linguistic information acquisition of visual stimuli independent of their physical shape, providing supportive evidence for the connectivity hypothesis.

4.1. The role of the VWFA in reading

Although previous training studies have found the selective activation of the VWFA for a new stimulus that was learned as a word, the stimulus had a word shape (e.g., foreign words; Xue et al., 2006; Song et al., 2012; Brem et al., 2014) or was learned as a real actual native word even though its shape was unlike a word (e.g., face, house, building, and meaningless figure; Song et al., 2010, 2012; Moore et al., 2013; Martin et al., 2019). In these situations, it is difficult to differentiate whether the activity change in the VWFA for a trained stimulus was from the word shape of the stimulus itself, the word physical shape of the native word, or language experience acquisition. Furthermore, the studies always merely considered the training change in the VWFA in the word domain without contrasting the change with that in other domains (e.g., objects). It is unclear whether the VWFA change was domain-general rather than word-specific. The current study trained subjects to learn identical nonsense word-unlike shapes to a word identity or an object identity. This training should control for the influence of the stimuli shape and actual words or objects. Moreover, we found that the VWFA change occurred only in word learning and not in object learning, suggesting the word specificity of the change. This could be interpreted by considering that the formation of a word-specific domain in the VWFA might be determined by the connectivity fingerprint of this area (Hannagan et al., 2015; Saygin et al., 2016; Spunt and Adolphs, 2017; Chen et al., 2019), and its activation changes might come from linguistic feedback information, which is transformed from high-order language regions that connect to the VWFA.

We also found that the VWFA ROIs at different cerebral locations presented distinctive sensitivity for word learning. First, the BOLD PSC value of the posttraining session in the individual-based VWFA ROIs showed a significant correlation with the behavioral performance of word-feature training, while the correlation was not significant in the three literature-based VWFA ROIs. This is likely because the literature-based ROIs could not capture sufficient behavioral-related neutral signals in our subjects. Second, the strongest word-learning effects for functional connectivity occurred in the posterior VWFA among the three literature-based VWFA ROIs (see Fig. 7). The posterior region has been argued to engage in letter-level information processing (Vinckier et al., 2007; Turkeltaub et al., 2008) and to play a critical role in visual feature extraction for visual stimuli (Lerma-Usabiaga et al., 2018). In our study, subjects associated a single stimulus with a bisyllabic pseudoword. This might need to combine the two syllables into a word identity. Each of the components was processed similar to a letter. It is possible that the region transferred the visual information of the stimuli to the language regions (i.e., the left pSTG, postCG, and SMA), and in turn, the syllabic information in the language regions was returned as feedback to the posterior VWFA. In contrast, the anterior part of the VWFA exhibited a relatively weak effect on word-identity learning. This may be because the anterior part began to process a whole word identity and received only feedback information about the word identity from the language regions.

4.2. Other language regions

Consistent with the connectivity hypothesis (Hannagan et al., 2015; Saygin et al., 2016; Spunt and Adolphs, 2017), we observed that the acquisition of language knowledge for novel figures also selectively induced increased activation magnitude in three language areas (the left pSTG, postCG and SMA) and a more similar activation pattern to that of read words. This indicates that word-feature learning for novel stimuli is involved in a large-scale brain network including the VWFA. Indeed, a rich body of literature has documented that these three language regions are critical nodes in the language dorsal pathway (Fedorenko and Thompson-schill, 2014; Price, 2012; Schlaggar and Mccandliss, 2007). They are responsible for processing pronunciation, articulation or grammatical information (Jobard and Crivello, 2003; Turkeltaub et al., 2003; Price, 2012; Wu et al., 2012). Given that the learned high-level language features in the present study were related to pronunciation and word class, the language areas might be involved in processing the phonetic features, automatic articulation, or grammatical information of figures. Moreover, this information can be used as a feedback source to assist the more efficient recognition of visual stimuli in the VWFA.

4.3. Feedback modulation of language areas to the VWFA

Although the present study could not provide direct evidence for the feedback flow from high-level language regions to the VWFA area, we found that the activation of the VWFA was modulated by high-level features of language. One possible account is that the activation in this area can be modulated by high-order linguistic properties such as phonology and semantics (Price and Devlin, 2011; Whaley et al., 2016),

and that the location of the VWFA is the interactive interface between visual feedforward and high-level language feedback flows (Price and Devlin, 2003, 2011; Carreiras et al., 2014). Recent tractographic studies have also revealed that the VWFA has strong structural white-matter connections with high-level language areas, such as perisylvian language areas, inferior frontal areas and the posterior superior temporal region (Bouhali et al., 2014; Schotten et al., 2014; Hannagan et al., 2015). Moreover, intracranial electroencephalography (EEG) studies have reported strong functional connectivity between the ventral visual cortex and high-level language areas during word processing (Thesen et al., 2012; Vidal et al., 2012; Whaley et al., 2016). In our experiment, the activation of the VWFA could have possibly been regulated by high-level language information resulting in selective effects in the word-identity learning group.

Note that the present study merely provided supportive evidence for the connectivity hypothesis of the VWFA. This does not mean that we disproved the shape hypothesis. In other words, the emergence of the VWFA in the ventral occipitotemporal cortex might be a consequence of the innate connectivity of the area with other language regions or of the interaction between the connectivity and physical visual shape of the stimuli.

4.4. Learning effects related to objects

Although our subjects could learn equivalently well regarding word and object identities in behavioral performance (Fig. 2), we observed significant brain signal changes only for word learning and not for object learning. The weak effects of object learning in neural subtracts might be due to the following reasons. First, we mainly investigated the learning effects in language ROIs, which led to the weak effects in the ROIs. The object-learning effects might become stronger in object ROIs. Second, our stimuli were trained into a specific category of words or into the general category of many types of objects (e.g., tool, arbitrarily manipulated object, nonmanipulable object, animal, and building) (Mahon et al., 2009). The ambiguous category specificity for objects led to the imprecise involvement of the cerebral cortices. Third, the method of learning knowledge (i.e., written visual description) was more similar to that of new word learning but not new object learning. The word-described features for object learning might also activate the brain network of word processing in addition to that of object processing. As a result, the neural difference of the learning effects between the two categories might be reduced. Finally, representational feature types of words are fewer than those of objects. The learning of two types of stimuli led to more exact representations for the learned words than for the learned objects. These reasons need to be clarified in future research.

4.5. Limitations

The current study has at least the following limitations. First, the words and objects were learned using two different subject groups, and the difference in subjects might confound the learning effects between words and objects. Hence, the observed effects in the present study need to be further verified by within-subject design studies in the future. Second, we failed to observe a double dissociation between the two learning groups in brain activity. Even the word-learning effects were not very strong. This might be due to our training paradigms, analysis approaches, the low number of subjects in each training group ($n = 13$), or short training duration (approximately 5 h). Finally, our training procedure for novel materials was rigid and included a few types of trained features. Our results should be validated by adopting a more flexible training procedure with more feature types.

5. Conclusion

To investigate what causes the appearance of the VWFA, we adopted a feature-based associative learning paradigm to train subjects to learn meaningless word-unlike figures as words or objects. We observed that word-feature but not object-feature learning could modulate the activation magnitude of the VWFA and higher-order areas (i.e., the left pSTG, postCG, and SMA) and the functional connectivity strength between these regions. Moreover, the growth of connection strength presented a gradual increasing tendency from the anterior to posterior axis along the VWFA. These identified a language network connecting to the VWFA that could be activated by language experience acquisition independent of stimuli shape. In other words, the emergence of the VWFA might be triggered by the feedback information from high-level language regions that are innately connected to it. This offers important insights into the underlying neural principles of word processing in the ventral occipitotemporal cortex.

Data and codes accessibility statement

The data and codes that support the finding of this study are available from the corresponding author (zzhhan@bnu.edu.cn), upon reasonable request. We are glad to share our data with anyone who shows interest with our current research, or someone who want use our data to do further analysis. And we will link the data with the research when we upload our data on public dataset later. To get the data recently, one can write a formal E-mail to the corresponding author with clear indication of the purpose with the data.

Declaration of competing interest

The authors declare no competing financial interests.

CRediT authorship contribution statement

Mingyang Li: Conceptualization, Data curation, Formal analysis, Investigation, Visualization, Writing - original draft. **Yangwen Xu:** Conceptualization, Investigation. **Xiangqi Luo:** Data curation. **Jiahong Zeng:** Data curation. **Zaizhu Han:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Supervision, Writing - review & editing.

Acknowledgements

We would like to thank BNU-CNLab members for data collection. We are also grateful to all research participants. This work was supported by the National Key R&D Program of China (2018YFC1315200; 2017YFF0207400), the National Natural Science Foundation of China (31872785; 81171019; 81972144) and the Beijing Natural Science Foundation (7182088).

References

Bi, Y., Wang, X., Caramazza, A., 2016. Object domain and modality in the ventral visual pathway. Trends Cognit. Sci. 20, 282–290. https://doi.org/10.1016/j.tics.2016.02.002.

Binder, J.R., Conant, L.L., Humphries, C.J., Fernandino, L., Simons, S.B., Aguilar, M., Desai, R.H., 2016. Toward a brain-based componential semantic representation. Cogn. Neuropsychol. 33, 130–174. https://doi.org/10.1080/02643294.2016.1147426.

Bolger, D.J., Perfetti, C.A., Schneider, W., 2005. Cross-cultural effect on the brain eevisited : universal structures plus writing system variation. Hum. Brain Mapp. 104, 92–104. https://doi.org/10.1002/hbm.20124.

Bouhali, F., Schotten, M.T. De, Pinel, P., Poupon, C., Universite, S., Atomique, E., Bm, I., Yvette, G., 2014. Anatomical connections of the visual word form area. J. Neurosci. 34, 15402–15414. https://doi.org/10.1523/JNEUROSCI.4918-13.2014.

Brem, S., Bach, S., Kucian, K., Kujala, J.V., Guttorm, T.K., Martin, E., Lyytinen, H., Brandeis, D., Richardson, U., 2014. Brain sensitivity to print emerges when children learn letter–speech sound correspondences. Proc. Natl. Acad. Sci. Unit. States Am. 107, 7939–7944. https://doi.org/10.1073/pnas.1421835111.

Carreiras, M., Armstrong, B.C., Perea, M., Frost, R., 2014. The what, when, where, and how of visual word recognition. Trends Cognit. Sci. 18, 90–98. https://doi.org/10.1016/j.tics.2013.11.005.

Chen, L., Wassermann, D., Abrams, D.A., Kochalka, J., Gallardo-Diez, G., Menon, V., 2019. The visual word form area is part of both language and attention circuitry. Nat. Commun. 10, 1–12. https://doi.org/10.1038/s41467-019-13634-z.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.A., 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. https://doi.org/10.1016/j.brainres.2014.08.050.

Dehaene, S., Cohen, L., 2011. The unique role of the visual word form area in reading. Trends Cognit. Sci. 15, 254–262. https://doi.org/10.1016/j.tics.2011.04.003.

Dehaene, S., Naccache, L., Cohen, L., Bihan, D. Le, Mangin, J., Poline, J., Rivière, D., 2001. Cerebral mechanisms of word masking and unconscious repetition priming. Nat. Neurosci. 4, 752–758. https://doi.org/10.1038/89551.

Dewitt, I., Rauschecker, J.P., 2013. Wernicke's area revisited: parallel streams and word processing. Brain Lang. 127, 181–191. https://doi.org/10.1016/j.bandl.2013.09.014.

Fedorenko, E., Thompson-schill, S.L., 2014. Reworking the language network. Trends Cognit. Sci. 18, 120–126. https://doi.org/10.1016/j.tics.2013.12.006.

Friston, K.J., Buechel, C., Fink, G.R., Morris, J., Rolls, E., Dolan, R.J., 1997. Psychophysiological and modulatory interactions in neuroimaging. Neuroimage 229, 218–229. https://doi.org/10.1006/nimg.1997.0291.

Glezer, L.S., Jiang, X., Riesenhuber, M., 2009. Evidence for highly selective neuronal tuning to whole words in the "visual word form area. Neuron 62, 199–204. https://doi.org/10.1016/j.neuron.2009.03.017.

Grill-Spector, K., Weiner, K.S., 2014. The functional architecture of the ventral temporal cortex and its role in categorization. Nat. Rev. Neurosci. 15, 536–548. https://doi.org/10.1038/nrn3747.

Hannagan, T., Amedi, A., Cohen, L., Dehaene-Lambertz, G., Dehaene, S., 2015. Origins of the specialization for letters and numbers in ventral occipitotemporal cortex. Trends Cognit. Sci. 19, 374–382. https://doi.org/10.1016/j.tics.2015.05.006.

Hasson, U., Levy, I., Behrmann, M., Hendler, T., Malach, R., 2002. Eccentricity bias as an organizing principle for human high-order object areas. Neuron 34, 479–490. https://doi.org/10.1016/S0896-6273(02)00662-1.

Jobard, G., Crivello, F., 2003. Evaluation of the dual route theory of reading : a metanalysis of 35 neuroimaging studies. Neuroimage 20, 693–712. https://doi.org/10.1016/S1053-8119(03)00343-4.

Lerma-Usabiaga, G., Carreiras, M., Paz-Alonso, P.M., 2018. Converging evidence for functional and structural segregation within the left ventral occipitotemporal cortex in reading. Proc. Natl. Acad. Sci. Unit. States Am. 115, E9981–E9990. https://doi.org/10.1073/pnas.1803003115.

Mahon, B.Z., Anzellotti, S., Schwarzbach, J., Zampini, M., Caramazza, A., 2009. Category-specific organization in the human brain does not require visual experience. Neuron 63, 397–405. https://doi.org/10.1016/j.neuron.2009.07.012.

Mahon, B.Z., Caramazza, A., 2011. What drives the organization of object knowledge in the brain? Trends Cognit. Sci. 15, 97–103. https://doi.org/10.1016/j.tics.2011.01.004.

Mahon, B.Z., Caramazza, A., 2009. Concepts and categories: a cognitive neuropsychological perspective. Annu. Rev. Psychol. 60, 27–51. https://doi.org/10.1146/annurev.psych.60.110707.163532.

Martin, L., Durisko, C., Moore, M.W., Fiez, J.A., Coutanche, M.N., Chen, D., 2019. The VWFA Is the Home of orthographic learning when houses are used as letters. eNeuro 6, 1–13. https://doi.org/10.1523/eneuro.0425-17.2019.

Moore, M.W., Durisko, C., Perfetti, C.A., Fiez, J.A., 2013. Learning to read an alphabet of human faces produces left-lateralized training effects in the fusiform gyrus. J. Cognit. Neurosci. 26, 896–913. https://doi.org/10.1162/jocn.

Nasr, S., Echavarria, C.E., Tootell, R.B.H., 2014. Thinking outside the box : rectilinear shapes selectively activate scene-selective cortex. J. Neurosci. 34, 6721–6735. https://doi.org/10.1523/JNEUROSCI.4802-13.2014.

Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cognit. Sci. 10, 424–430. https://doi.org/10.1016/j.tics.2006.07.005.

Price, C.J., 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech , spoken language and reading. Neuroimage 62, 816–847. https://doi.org/10.1016/j.neuroimage.2012.04.062.

Price, C.J., Devlin, J.T., 2011. The Interactive account of ventral occipitotemporal contributions to reading. Trends Cognit. Sci. https://doi.org/10.1016/j.tics.2011.04.001.

Price, C.J., Devlin, J.T., 2003. The Myth of the Visual Word Form Area, vol. 19, pp. 473–481. https://doi.org/10.1016/S1053-8119(03)00084-3.

Raichle, M.E., 2015. The brain's default mode network. Annu. Rev. Neurosci. 38, 433–447. https://doi.org/10.1146/annurev-neuro-071013-014030.

Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L., 2001. A default mode of brain function. Proc. Natl. Acad. Sci. U.S.A. 98, 676–682. https://doi.org/10.1073/pnas.98.2.676.

Reich, L., Szwed, M., Cohen, L., Curie, M., 2011. A ventral visual stream reading center independent of visual experience. Curr. Biol. 21, 363–368. https://doi.org/10.1016/j.cub.2011.01.040.

Reilly, J.X.O., Woolrich, M.W., Behrens, T.E.J., Smith, S.M., Johansen-berg, H., 2012. Tools of the trade : psychophysiological interactions and functional connectivity. Soc. Cognit. Affect Neurosci. 7, 604–609. https://doi.org/10.1093/scan/nss055.

Sahin, N.T., 2009. Sequential processing of lexical, grammatical, and phonological information within Broca's area. Science 326, 445–449. https://doi.org/10.1126/science.1174481.

Saygin, Z.M., Osher, D.E., Norton, E.S., Youssoufian, D.A., Beach, S.D., Feather, J., Gaab, N., Gabrieli, J.D.E., Kanwisher, N., 2016. Connectivity precedes function in the development of the visual word form area. Nat. Neurosci. 19, 1250–1255. https://doi.org/10.1038/nn.4354.

Schlaggar, B.L., Mccandliss, B.D., 2007. Development of neural systems for reading. Annu. Rev. Neurosci. 475–503. https://doi.org/10.1146/annurev.neuro.28.061604.135645.

Schotten, M.T. De, Cohen, L., Amemiya, E., Braga, L.W., Dehaene, S., Yvette, G., Paris-sud, F.U., 2014. Learning to read improves the structure of the arcuate fasciculus. Cerebr. Cortex 24, 989–995. https://doi.org/10.1093/cercor/bhs383.

Song, Y., Hu, S., Li, X., Li, W., Liu, J., 2010. The role of top-down task context in learning to perceive objects. J. Neurosci. 30, 9869–9876. https://doi.org/10.1523/JNEUROSCI.0140-10.2010.

Song, Y., Tian, M., Liu, J., 2012. Top-down processing of symbolic meanings modulates the visual word form area. J. Neurosci. 32, 12277–12283. https://doi.org/10.1523/JNEUROSCI.1874-12.2012.

Spunt, R.P., Adolphs, R., 2017. A new look at domain specificity: insights from social neuroscience. Nat. Rev. Neurosci. 18, 559–567. https://doi.org/10.1038/nrn.2017.76.

Srihasam, K., Vincent, J.L., Livingstone, M.S., 2014. Novel domain formation reveals proto-architecture in inferotemporal cortex. Nat. Neurosci. 17, 1776–1783. https://doi.org/10.1038/nn.3855.

Striem-amit, E., Cohen, L., Dehaene, S., Amedi, A., 2011. Reading with sounds: sensory substitution selectively activates the visual word form area in the blind. Neuron 76, 640–652. https://doi.org/10.1016/j.neuron.2012.08.026.

Szwed, M., Dehaene, S., Kleinschmidt, A., Eger, E., Valabrègue, R., Amadon, A., Cohen, L., 2011. Specialization for written words over objects in the visual cortex. Neuroimage 56, 330–344. https://doi.org/10.1016/j.neuroimage.2011.01.073.

Tan, L.H., Laird, A.R., Li, K., Fox, P.T., 2005. Neuroanatomical correlates of phonological processing of Chinese characters and alphabetic words : a meta-analysis. Hum. Brain Mapp. 91, 83–91. https://doi.org/10.1002/hbm.20134.

Taylor, J.S.H., Davis, M.H., Rastle, K., 2019. Mapping visual symbols onto spoken language along the ventral visual stream. Proc. Natl. Acad. Sci. Unit. States Am. 116, 17723–17728. https://doi.org/10.1073/pnas.1818575116.

Thesen, T., McDonald, C.R., Carlson, C., Doyle, W., Cash, S., Sherfey, J., Felsovalyi, O., Girard, H., Barr, W., Devinsky, O., Kuzniecky, R., Halgren, E., 2012. Sequential then interactive processing of letters and words in the left fusiform gyrus. Nat. Commun. 3, 1284–1288. https://doi.org/10.1038/ncomms2220.

Turkeltaub, P.E., Flowers, D.L., Lyon, L.G., Eden, G.F., 2008. Development of ventral stream representations for single letters. Proc. Natl. Acad. Sci. Unit. States Am. 1145, 13–29. https://doi.org/10.1196/annals.1416.026.

Turkeltaub, P.E., Gareau, L., Flowers, D.L., Zeffiro, T.A., Eden, G.F., 2003. Development of neural mechanisms for reading. Nat. Neurosci. 6, 767–773. https://doi.org/10.1038/nn1065.

Vidal, J.R., Freyermuth, S., Jerbi, K., Hamame, C.M., Ossandon, T., Bertrand, O., Minotti, L., Kahane, P., Berthoz, A., Lachaux, J.-P., 2012. Long-distance amplitude correlations in the high gamma band reveal segregation and integration within the reading network. J. Neurosci. 32, 6421–6434. https://doi.org/10.1523/JNEUROSCI.4363-11.2012.

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J.P., Sigman, M., 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. https://doi.org/10.1016/j.neuron.2007.05.031.

Whaley, M.L., Kadipasaoglu, C.M., Cox, S.J., Tandon, N., 2016. Modulation of orthographic decoding by frontal cortex. J. Neurosci. 36, 1173–1184. https://doi.org/10.1523/JNEUROSCI.2985-15.2016.

Wu, C., Ho, M.R., Chen, S.A., 2012. A meta-analysis of fMRI studies on Chinese orthographic , phonological , and semantic processing. Neuroimage 63, 381–391. https://doi.org/10.1016/j.neuroimage.2012.06.047.

Xue, G., Chen, C., Jin, Z., Dong, Q., 2006. Language experience shapes fusiform activation when processing a logographic artificial language: an fMRI training study. Neuroimage 31, 1315–1326. https://doi.org/10.1016/j.neuroimage.2005.11.055.