Cerebral Cortex, November 2017;27: 5197–5210

doi: 10.1093/cercor/bhw300
Advance Access Publication Date: 24 September 2016
Original Article

ORIGINAL ARTICLE

Orthographic and Phonological Representations in the Fusiform Cortex

Libo Zhao[1,2], Chunhui Chen[1,2], Luying Shao[1,2], Yapeng Wang[1,2], Xiaoqian Xiao[1,2], Chuansheng Chen[3], Jianfeng Yang[4], Jason Zevin[5], and Gui Xue[1,2]

[1]State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, PR China, [2]Center for Collaboration and Innovation in Brain and Learning Sciences, Beijing Normal University, Beijing 100875, PR China, [3]Department of Psychology and Social Behavior, University of California, Irvine, CA 92697, USA, [4]School of Psychology, Shanxi Normal University, Xi'an 710062, PR China, and [5]Department of Linguistics, University of Southern California, Los Angeles, CA 90089, USA

Address correspondence to Gui Xue. Email: guixue@gmail.com

Abstract

Mental and neural representations of words are at the core of understanding the cognitive and neural mechanisms of reading. Despite extensive studies, the nature of visual word representation remains highly controversial due to methodological limitations. In particular, it is unclear whether the fusiform cortex contains only abstract orthographic representation, or represents both lower and higher level orthography as well as phonology. Using representational similarity analysis, we integrated behavioral ratings, computational models of reading and visual object recognition, and neuroimaging data to examine the nature of visual word representations in the fusiform cortex. Our results provided clear evidence that the middle and anterior fusiform represented both phonological and orthographic information. Whereas lower level orthographic information was represented at every stage of the ventral visual stream, abstract orthographic information was increasingly represented along the posterior-to-anterior axis. Furthermore, the left and right hemispheres were tuned to high- and low-frequency orthographic information, respectively. These results help to resolve the long-standing debates regarding the role of the fusiform in reading, and have significant implications for the development of psychological, neural, and computational theories of reading.

Key words: fMRI, H-Max model, reading, representational similarity analysis, VWFA

Introduction

Representation is at the core of a mechanistic understanding of cognition (Marr, 1982). In the field of reading research, the nature of lexical representation (i.e., orthography, phonology, and semantics) has been studied at the behavioral (Goswami, 2000; Ramus and Szenkovits, 2008), neural (Dehaene and Cohen, 2011; Price and Devlin, 2011), and computational (Plaut et al. 1996; Coltheart et al. 2001) levels. Nevertheless, due to methodological limitations, these levels of investigation have been carried out separately. Using functional imaging and recently developed multiple voxel pattern analysis (MVPA), in particular, representation similarity analysis (RSA), cumulative studies have investigated the nature of neural representation associated with reading and linked multiple levels of analyses

(Glezer and Riesenhuber, 2009; Rauschecker et al. 2012; Nestor et al. 2013; Rothlein and Rapp, 2014; Baeck et al. 2015; Hirshorn et al. 2016).

Using RSA, this study aimed to investigate the nature of lexical representations in the fusiform cortex, a brain region that has been implicated in visual word processing (Cohen et al. 2002; Dehaene and Cohen, 2011; Price and Devlin, 2011; Wandell, 2011), but its precise role has been much debated (Vinckier et al. 2007; Glezer et al. 2009; Rauschecker et al. 2012; Hirshorn et al. 2016). One central debate is whether the fusiform represents phonological information in addition to orthographic information (Dehaene and Cohen, 2011; Price and Devlin, 2011). The strong visual word form area (VWFA) hypothesis posits that this brain area's central role is abstract representation of visual word forms (Cohen et al. 2002; McCandliss et al. 2003). Many studies, however, have found VWFA activation even when there is no explicit visual input, such as during verbal repetition (Price et al. 2003), spelling (Booth et al. 2002), auditory rhyming judgment (Yoncheva et al. 2010), and auditory lexical decision (Dehaene et al. 2010). Furthermore, this area's activation is modulated by phonological learning (Xue et al. 2006) and the transparency level of form-sound correspondence (Mei et al. 2013).

This debate is ongoing because it is difficult to disentangle orthographic and phonological representations due to their compulsive coactivation during reading. Arguably, the phonology-related activations in the VWFA in the above studies could be attributed to top-down modulation of orthographic processing by phonology, but may not necessarily result from phonological processing per se (Dehaene and Cohen, 2011). Although studies using rapid adaptation (Glezer et al. 2009, 2015, 2016) or MVPA (Rauschecker et al. 2012; Nestor et al. 2013; Rothlein and Rapp, 2014; Baeck et al. 2015) hold the promise of pinpointing the neural representation in the fusiform region, these studies did not separately manipulate orthographic similarity and phonological similarity. Chinese language is a particularly good orthography for this purpose because it has complex visual structure and irregular grapheme-to-phoneme correspondence, which allow for the selection of stimuli whose phonological and orthographic similarities are essentially uncorrelated.

Another unresolved debate concerns the nature of orthographic representation in the fusiform. In particular, it is unclear whether the VWFA represents lower level orthographic information in addition to abstract visual form representation. Based on the hierarchical organization of the visual system (Riesenhuber and Poggio, 1999), the fusiform has been proposed to have local combination detectors (LCD) that are sensitive to increasingly larger fragments of words (Dehaene et al. 2005; Dehaene and Cohen, 2011), and the representation in the anterior region is abstract and invariant to font, size, position, and even case (Dehaene et al. 2005; Dehaene and Cohen, 2011). Consistently, univariate studies have revealed increasing sensitivity to larger word segments along the posterior-to-anterior axis of the visual stream (Vinckier et al. 2007). The anterior fusiform (aFG) region shows whole-word tuning (Glezer et al. 2009), and significant cross-case (Dehaene et al. 2001, 2004), cross-font (Qiao et al. 2010), and cross-script priming (Chee et al. 2003). Inconsistent with these results based on univariate analyses, a recent MVPA study found that the VWFA represents lower level information (i.e., position) in addition to higher level information (i.e., identity) (Rauschecker et al. 2012).

This study aimed to determine the level of information represented in the fusiform by linking neural pattern similarity with the multilevel representations generated from the hierarchical object recognition model (i.e., H-MAX) (Riesenhuber and Poggio, 1999; Serre et al. 2005). The H-MAX model simulates the neuronal responses to visual objects at 4 successively more abstract levels (S1, C1, S2, and C2), and these model-based representations have been linked to neural representation during object recognition (Kriegeskorte et al. 2008b). In a newer version of H-MAX, the S2 features are learned from sample images rather than taken from a static feature dictionary (Serre et al. 2005). This enables H-MAX to flexibly stimulate neural responses to all kinds of images, including those of Chinese characters. This study took advantage of this newer version of H-MAX and abstracted S2 features of Chinese characters from a large character pool. By relating neural pattern similarity with C1 and C2 representations generated from H-MAX, lower and higher levels of orthographic information could be estimated for each subregion of the fusiform and further compared across these subregions.

In addition to addressing the two debates on the role of the fusiform in lexical representations, data from this study also allowed us to test how well a computational model of Chinese word reading (Xing and Li, 2004; Yang et al. 2009) fit neural imaging and behavioral data. The computational model maps orthography to phonology, both of which are represented as distributed activation patterns over a number of binary feature units. In particular, orthographic representations reflect abstractions that take into account the organization of Chinese characters, particularly structures that contribute to the mapping from orthography to phonology. By correlating representations defined by the computational model with the neural data, we could test the neural feasibility of the computational model and provides insights for the development of future models.

Materials and Methods

Participants

Twenty (8 males, 12 females) healthy Chinese college students (18–26 years old, mean age = 21.5 years, SD = 0.98; SD, standard deviation) were paid participants of this study. They were all native Chinese speakers. None had any history of psychiatric or neurological disease. They all had normal or corrected-to-normal vision and were right-handed as judged by Snyder and Harris's handedness inventory (1993). Written consent approved by Institutional Review Board of Beijing Normal University was obtained from each participant after a full explanation of the purpose and procedures of the study.

Materials and Functional Magnetic Resonance Imaging Procedure

Two sets of 34 high-frequency Chinese characters were used (Fig. 1). Characters in each set were chosen with the constraint that the orthographic similarity matrix and the phonological similarity matrix, based on subjective ratings (see below), were not correlated. Ten subjects received word set 1 and the other 10 received word set 2.

Participants finished 6 runs of the naming task in the scanner, each containing all 34 characters in a set and lasting for 408 s. Each character was presented only once within a run to reduce the repetition priming effect. The order of the 34

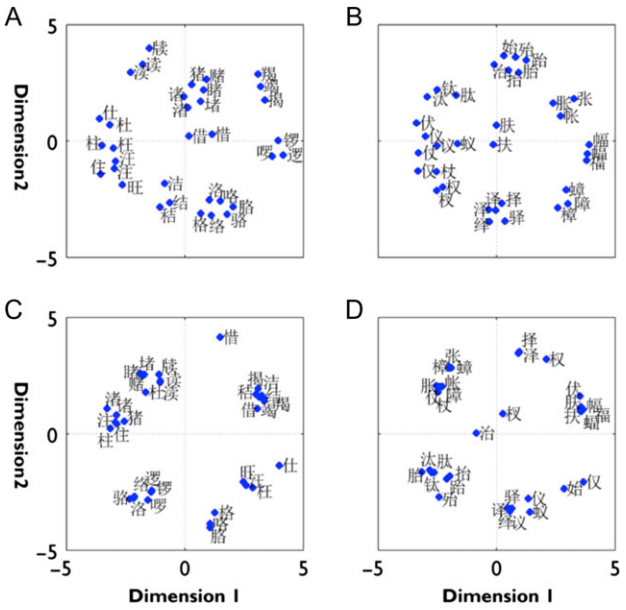

Figure 1. Similarity space of the Chinese characters based on orthographic and phonological ratings as revealed by multiple dimensional scaling. (*A*) and (*B*) Orthographic similarity structure for set 1 and set 2 characters, respectively, is depicted. (*C*) and (*D*) Phonological similarity structure for set 1 and set 2 characters, respectively, is depicted. Each dot represents a Chinese character with its label by its side, and the distance between a pair of dots represents the similarity of them. The farther away 2 dots are, the more dissimilar they are to each other.

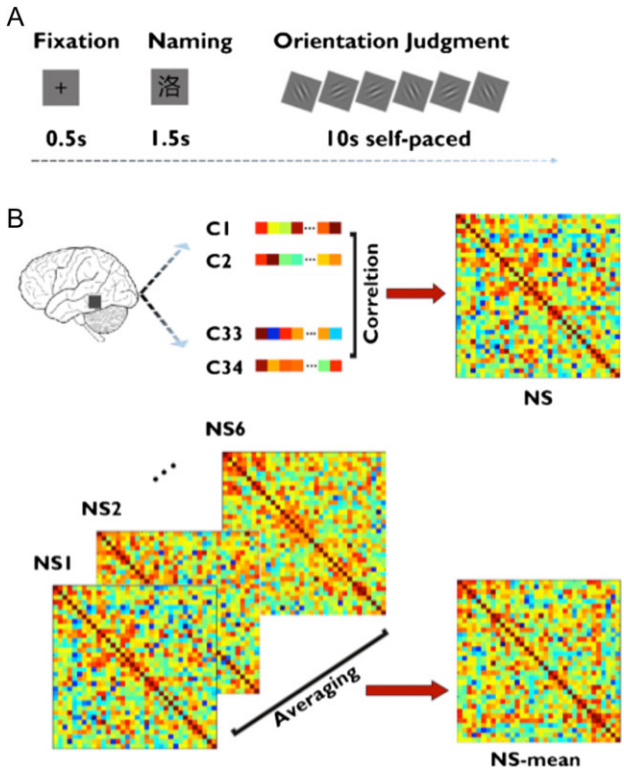

Figure 2. Experiment design and calculation of neural pattern similarity. (*A*) A slow event-related design (12 s per trial) was used to improve the accuracy of the estimation of single-trial responses. Each trial started with a fixation cross of 0.5 s, and then a Chinese character was presented for 1.5 s. Participants were asked to name the character. To prevent further encoding between trials, a series of Gabor images were presented during the 10 s intertrial-interval, and participants were asked to judge the orientation of each image as quickly and accurately as possible. (*B*) For a given ROI, neural pattern similarity among the characters was computed for each run and for each participant separately. Specifically, the neural activation patterns of every possible character pair were subjected to Pearson correlation analysis, which resulted in a neural pattern similarity matrix of 34 × 34 for a given run (NS) (upper panel). Then for each ROI and each participant, the mean neural pattern similarity matrix (NS_{mean}) was computed by averaging these NS matrices of all 6 runs (NS_1–NS_6) (lower panel). These correlations were then Fisher transformed for further analysis.

naming trials was pseudorandomized for each run, with the restriction that no two consecutive characters were orthographically or phonologically similar. A slow event-related design (12 s per trial) was used to enable accurate estimation of the BOLD response pattern associated with each trial (see Fig. 2A). The trial started with a 500 ms fixation cross, followed by a Chinese character presented in the center of the screen for 1500 ms, with a measured visual angle of 2.85° × 2.85°. The characters were presented in black color against a white background. Participants were asked to name overtly each character while keeping their head still. To prevent participants from rehearsing the character that just disappeared, the naming task was followed by a visual orientation judgment task for 10 s, where participants were asked to identify whether the Gabor image presented on the screen tilted to the left or the right of the vertical axis (tilting 45°). Participants were asked to respond as quickly and accurately as possible. Once participants responded by pressing 1 of the 2 buttons, the current image disappeared and, after a blank interval of 100 ms, the next Gabor image was shown. This self-paced procedure was used to make this task engaging.

Rating-based Orthographic and Phonological Similarities

After the naming task, participants were asked to rate orthographic and phonological similarities for every possible pair of the characters they just named (34 × (34–1)/2 = 561 pairs). A 7-point scale was used, with 1 = "Identical" and 7 = "Completely different". These scores were reversed in data analysis to generate an intuitive index of similarity, with 1 = Completely different and 7 = Identical. The order of rating these 2 types of similarities (orthographic and phonological similarities) was counterbalanced across participants. Ratings of a given pair from the 10 participants who received the same set of stimuli were averaged, resulting in two, 34 × 34 matrices of averaged ratings of orthographic (RO) and phonological (RP) similarities.

To make sure that participants understand what we meant by orthographic and phonological similarities, we provided examples of character pairs that were Identical and Completely different for both rating tasks. In addition, we pointed out critical features participants should consider. Specifically, for orthographic similarity, we asked participants to evaluate the amount of overlap in terms of the radicals included, number of strokes, and spatial organization (left-right or top-down). For phonological similarity, we asked participants to evaluate how similar the characters were to each other in terms of the beginning, the middle, and the end of the sounds as well as their tones.

Intersubject agreement in subjective ratings was estimated for each set of the characters and each type of similarities. The results revealed high Cronbach's α for both orthographic (set 1: 0.94, set 2: 0.95) and phonological ratings (set 1: 0.97, set 2: 0.97), suggesting a very high level of interrater agreement.

Model-based Orthographic Similarity

The connectionist model of Chinese character reading we adopted (Yang et al. 2009) maps orthographic input of a Chinese character to its phonological output via a hidden layer. Based on the linguistic description of Chinese orthography (Xing and Li, 2004), the representation scheme of orthography incorporates 3 kinds of visual–spatial features: structural features of the character (e.g., the overall shape of 骆 is left-right), orthographic features of the radicals (e.g., the position of the radical 马 in the character of 骆 is left), and a finer-grained description of the radicals in terms of strokes (e.g., the first stroke of the radical 马 is 𠃌)(see Supplementary Fig. 1). Orthographic representation was organized into a set of 9 slots, with 1 slot for structural features of the character, 7 slots for radicals (1 per radical), and 1 slot for random features. Each slot was further organized into a variable number of subgroups, each of which represents an orthographic feature of radicals or a stroke by the constituent units that took on either 0 or 1. In total, 270 such binary units were included. Therefore, using the above scheme, the orthography of each character used in this study (2 sets of 34 characters) was coded as a binary vector of 270 binary units. By calculating the cosine similarity for all possible 561 pairs of such vectors within each set, we obtained another 34 × 34 matrix of orthographic similarity. These correlation values were Fisher transformed. Examples of highly similar pairs on the orthographic dimension included 骆 and 咯 (0.73), 杜 and 仕 (0.65) (the C1 similarities for these 2 pairs were 0.84–0.82, respectively, and the corresponding C2 similarities for these 2 pairs were 0.97–0.98).

It should be noted that the orthographic representations used in the model may diverge substantially from the visually based similarity scores because, in addition to visual similarity, the model encodes features of characters that are important for orthography-to-phonology translation (Yang et al. 2009; Yang et al. 2013), as well as some arbitrary decisions that were required to create a reasonably compact and systematic representation. Taking the highly similar pairs of characters above as examples, these characters have the typical structure of "phonograms" with 2 components: the left component (typically a single radical) that provides probabilistic cues to semantics and the right component that provides probabilistic cues to pronunciation (i.e., a "phonetic component") (Zhou and Marslen-Wilson, 1999). The arrangement of the orthographic representation used by Yang et al. (2009, 2013) ensures that phonograms with the same phonetic component are represented as such in orthographic representation. This was accomplished by ordering the slots in a particular way that is similar to the process of "vowel-centering" orthographic representations in models of reading in English (Harm and Seidenberg, 1999). On the other hand, some visually salient features, such as the overall shape of a character, are encoded by a very small number of units, resulting in a substantial reduction in the correlation between visual similarity and the similarity based on the abstract coding scheme described above.

Model-based Phonological Similarity

Phonological representation in this model was based on a phonetic description of Standard Mandarin (Huang and Liao, 2002). As with orthographic representation, phonological representation was developed to reflect important aspects of the similarity space for Mandarin syllables. Despite long-standing debates about how speech patterns are represented (Dell and Juliano, 1996; McMurray and Jongman, 2010), segment-and-feature systems used in reading models are widely considered acceptable for descriptive purposes, and have previously been shown to align well with speakers' similarity judgments (Hahn and Bailey, 2005). The representation comprised 5 slots, 1 for the onset (in Mandarin, the onset unit, or "shengmu" contains only the initial consonant), 3 for the rime (the "yunmu", which in this case includes any semivowels, nuclear vowels, and codas), and a fifth slot composed of 4 units to represent lexical tone (see Supplementary Fig. 1). Each of the 4 slots for phonemes comprised 3 subgroups with a variable number of units, representing the manner of articulation, place of articulation, and impressionistic vowel quality. Multiple units were used to encode each of these dimensions, in order to capture graded patterns of similarity. For example, the sounds /p/ and /k/ are more distinct from one another than the sounds /p/ and /t/, although all 3 sounds differ in their "place," a relation that is not possible to capture if each place of articulation (bilabial for /p/, alveolar for /t/, and velar for /k/) were encoded as a binary feature. Finally, 4 units were used to represent the 4 Mandarin tones. Thus, in total, 92 binary units were used.

Similar to orthographic similarity, the phonology of each character used in this study was coded as a binary vector of 92 binary units, and a 34 × 34 matrix of phonological similarity was obtained for each set by computing cosine similarity for all possible pairs of these vectors. These correlation values were Fisher transformed. Examples of highly similar pairs in phonology include 驿/yi4/ and 蚁 /yi3/ (0.88), 赌/du3/ and 杜 /du4/ (0.89).

Similarity based on the H-MAX Model

To further examine the nature of orthographic representation in the fusiform, we adopted the classic H-MAX model that implemented a 4-level hierarchy of visual processing, S1, C1, S2, and C2 (Serre et al. 2005). The gray scale image of each character (120 × 120 pixels) was first fed as input to the S1 stage in the model, where simple cells were simulated by Gabor filters of 4 different orientations (–45°, 0°, 45°, 90°) and 12 receptive field sizes (7–29 pixels). The C1 stage simulated responses of complex cells, by performing the MAX operation on outputs of simple cells with the same orientation selectivity in 1 of the 6 scale groups (7 & 9, 11 & 13, 15 & 17, 19 & 21, 23 & 25, 27 & 29 pixels) in the corresponding spatial range of pooling (8 × 8, 10 × 10, 12 × 12, 14 × 14, 16 × 16, and 18 × 18 pixels for the 6 scale groups, respectively). Due to the poor spatial resolution of functional magnetic resonance imaging (fMRI) scan, the C1 responses were smoothed using a Gaussian kernel size of 15 and sigma 5 to better associate with the imaging data (Kriegeskorte et al. 2008a).

The S2 units were learned prototypes at the C1 level from a large pool of gray scale images of Chinese characters (2390 from the a Chinese database (Liu et al. 2008)). Specifically, 1000 patches of 4 sizes (4 × 4, 8 × 8, 12 × 12, and 16 × 16 pixels) were extracted at random positions from the C1 maps at each scale and direction, with each patch being set as the prototype or the center of one S2 unit. The S2 responses of each target character were computed by convolving its C1 input image at all scales and all directions with the stored S2 units. That is, each S2 unit behaved like a radial basis function, whose response was dependent on the Euclidean distance between a new C1 input patch (at a particular location and scale) and the stored prototype. The C2 responses were obtained by taking the strongest response from each S2 map across all scales and positions, that is, a vector of 4000 (1000 per direction) for each character.

For representations at the C1 and the C2 levels, the Pearson correlation between pairs of these vectors was calculated for each set, yielding the similarity matrix for C1 and C2, respectively.

fMRI Acquisition and Preprocessing

Magnetic resonance imaging data were collected with a 3 T Siemens Trio Tim scanner at the MRI center, Beijing Normal University. Single-shot T2*-weighted echo planar image (EPI) sequence was used for functional imaging acquisition (TR = 2000 ms; TE = 25 ms; flip angle = 90 °; field of view (FOV) = 192 × 192 mm; matrix = 64 × 64; slice thickness = 3 mm). Forty-one contiguous axial slices parallel to anterior commissure–posterior commissure plane were obtained to cover the whole cerebrum and partial cerebellum. Anatomical fMRI was acquired using a T1-weighted, 3D, gradient echo pulse-sequence (TR = 2530 ms; TE = 3.39 ms; flip angle = 7 °; FOV = 256 × 256 mm; matrix = 192 × 256; slice thickness = 1.34 mm). One hundred and twenty-eight sagittal slices were acquired to provide a high-resolution structural image of the whole brain.

Image preprocessing and statistical analyses were carried out using FEAT (FMRI Expert Analysis Tool) version 5.98. The first 5 volumes before the task were discarded to allow for T1 equilibrium. The remaining images were then realigned and temporally filtered (nonlinear high-pass filter with a 60 s cut-off). The EPI images were first registered to the magnetization-prepared rapid gradient echo (MPRAGE) structural image and then into the standard Montreal Neurological Institute (MNI) space, using affine transformations. Registration from structural images to the standard space was further refined using FSL nonlinear image registration tool nonlinear registration.

fMRI Data Analysis

The general linear model within the FILM module of FMRIB Software Library (FSL) was used to model the data. To estimate the BOLD response associated with Chinese character naming, all characters were modeled as one regressor. The error trials in the orientation task were encoded as one nuisance variable, whereas the correct trials were not coded and thus treated as the implicit baseline. The hemodynamic delay was modeled by convolving delta functions with the canonical double gamma hemodynamic response function. To estimate the single-trial neural response, a least-square single-trial method was used (Mumford et al. 2012), which has been shown to provide robust result in single-trial model.

Representational Similarity Analysis of fMRI Data

We used RSA to examine the neural activation similarity across trials (Kriegeskorte et al. 2008a). As our main goal was to elucidate the representations in the fusiform gyrus, we defined the anterior, middle, and posterior parts of the lateral fusiform gyrus as the ROIs, which corresponded to the posterior temporal fusiform gyrus, the temporal occipital fusiform gyrus, and the occipital fusiform gyrus, respectively, in the Harvard–Oxford probabilistic atlases implemented in FSL. In addition, bilateral cuneal cortex (CC, part of the early visual cortex) was also defined using the same atlas to serve as the control regions.

The VWFA was defined based on a previous study on Chinese reading (Xue et al. 2006), which found the peak of activation at [−39 −59 −12] in the MNI standard space. A 6 mm sphere centered at this voxel was defined as the VWFA. The right homolog was also defined. Individual subjects' functional VWFAs were defined based on the contrast of characters > baseline ($P < 0.05$, corrected). The peak voxel within the 12 mm sphere of group VWFA was selected, and a 6 mm sphere around the peak voxel was then defined. This was done separately for the left and right hemispheres. The average locations were −41 ± 5, −60 ± 6, −11 ± 6 (the left VWFA) and 42 ± 7, −61 ± 5, −14 ± 6 (the right VWFA) in MNI coordinates.

For a given region of interest (ROI), the multivoxel response pattern (beta estimates) for each of the 34 characters was extracted for each run and for each participant separately (Fig. 2*B*). Then all possible pairs of these 34 activation patterns (561 in total) were subjected to Pearson correlation analysis, resulting in a 34 × 34 matrix of the neural similarity (NS) in each run (NS_1–NS_6). The NS matrices of all 6 runs were then averaged and Fisher transformed to obtain the mean NS (NS mean).

Testing the Functional Gradient of the Fusiform

To reveal the finer posterior–anterior functional dissociation of the fusiform, we did a searchlight analysis (Kriegeskorte et al. 2008a, 2008b) within the anatomically defined fusiform region (bilateral aFG, mFG, and pFG). For each voxel, a cube with a 3-voxel radius was defined. Similar to ROI-based analysis, NS for all character pairs was calculated within that cube for each run. Fisher's *Z* transformed correlation coefficients were then averaged across runs and correlated with indices of the behavioral and model-based similarities of orthography and phonology (RO, RP, MO (model-based orthographic), MP (model-based phonological), C1, and C2). The correlation coefficients were then assigned to the center voxel of the cube. Finally, random-effects group analysis was performed for each map to identify voxels with correlation coefficients significantly larger than zero based on *t*-tests.

To test the functional gradient of the fusiform along the posterior–anterior axis, Spearman correlations were obtained between each voxel's group-averaged *t*-values and the corresponding MNI *y* coordinate (Wang et al. 2014). A positive correlation would suggest a linear increase of representation along the posterior–anterior axis. The voxels included in this analysis were in the range of −90 to −30 (MNI *y* coordinate), which spanned the entire fusiform. The ranges of *z* and *x* coordinates were based on the VWFA coordinates, with *z* coordinate from −6 to −18, and *x* coordinate from −50 to −30 for the left hemisphere and 30–50 for the right hemisphere.

Linking Rating-based, Model-based, and Neural Similarities

To link different levels of analysis, Spearman correlations were obtained for each participant and Fisher transformed. To test the hypothesis that these correlations would be positive (i.e., a unidirectional hypothesis), one-tailed statistical test was used (Peelen et al. 2014).

Representational Connectivity Analysis

Following Kriegeskorte et al. (2008a), we examined representational connectivity among the eight structural ROIs, using Spearman correlations between the NS patterns of every pair of ROIs. The network was visualized using BrainNet Viewer (Xia et al. 2013).

Noise Ceiling Calculation

To estimate the expected performance given the noise in the neural data, noise ceiling was calculated following Nili et al. (2014).

For any given model, the upper bound was estimated by computing the averaged Spearman correlation between each subject's NS and the mean NS of all subjects. The lower bound was estimated using the leave-one-subject-out method, that is calculating the Spearman correlation between each subject's NS and the mean NS of all other subjects and then averaging these correlations. The upper and lower bounds are reported in the top part of Table 1.

Permutation Test of Significance Level

Permutation tests were performed to determine the significance level of correlations. In this analysis, word labels of the multivoxel neural response pattern were permuted 1000 times. Neural similarity was then calculated and correlated with rating- and model-based similarities. The maximum correlation value over all 8 regions was chosen to construct a distribution under the null hypothesis. A nonparametric statistical test was obtained by calculating the proportion of randomized test statistics that exceeded the observed statistics.

Results

Orthographical and Phonological Representations in the Fusiform Cortex

Two sets of 34 characters were carefully selected so that their orthographic similarity and phonological similarity were minimally correlated. Indeed, this expectation was confirmed by the subjective ratings from both the participants (word set 1, $r = 0.12$; word set 2, $r = 0.01$) (Fig. 1*A–D*) and a separate group of 8 participants (word set 1, $r = 0.09$; word set 2, $r = 0.03$). The use of 2 sets of characters and random assignment of subjects to each set helped to minimize the effect of materials and increase the generalibility of the results. To rule out the possibility that these low correlations were artifacts of averaging across subjects, we also calculated the RO–RP correlation for each individual subject. The results revealed that RO-RP correlations ranged from 0.02 to 0.19 (mean = 0.12, SD = 0.04) for set 1, and 0.006 to 0.13 (mean = 0.05, SD = 0.04) for set 2.

Subjects were asked to name the characters in the scanner using a single-trial fMRI design (Fig. 2*A*). Univariate analysis revealed a widely distributed network for reading Chinese (Supplementary Fig. 2). Focusing on the fusiform, we found strong bilateral activation, and a significant interaction between subregion (posterior, middle, and anterior) and hemisphere (left vs. right) ($F(3,57) = 12.04$, $P < 0.0001$). Hemispheric asymmetry index revealed left lateralization in the middle fusiform (mFG), but right lateralization in the aFG and the posterior fusiform (pFG) as well as in the CC (Supplementary Fig. 3).

To examine whether orthographic and/or phonological information was represented in the fusiform, we calculated the Spearman correlation for each subject, between the NS in the fusiform ROIs (i.e., the bilateral aFG, mFG, and pFG) and the subjective RO and RP. These correlations were Fisher's Z transformed and subjected to one-sample *t*-test. This analysis revealed significant correlations between RO and NS in the bilateral aFG and the bilateral mFG, but not in the pFG or CC. Significant correlations between RP and NS were found in the bilateral aFG and the right mFG (Fig. 3; Table 1), which remained essentially the same after controlling for the behavioral and model-based similarities (RO, MO, C1, C2) (Fig. 3; RP_p in Table 1). In addition, these results remained essentially the same when NS was correlated with the ratings of

Table 1 Spearman correlations (and associated statistics) between NS and psychological and model similarity in the fusiform subregions and the CC

Index	Tests	laFG	raFG	lmFG	rmFG	lpFG	rpFG	lCC	rCC
Noise ceiling	Lower bound	0.16	0.16	0.11	0.15	0.09	0.07	0.17	0.14
	Upper bound	0.31	0.31	0.28	0.30	0.27	0.26	0.32	0.30
RO	Mean *r*	0.06	0.07	0.03	0.04	0.01	0.02	0.02	0.02
	P_t-test	0.0001	0.0001	0.01	0.01	0.33	0.11	0.13	0.12
	P_perm	0.0001	0.0001	0.011	0.003	0.25	0.09	0.17	0.20
RP	Mean *r*	0.03	0.03	0.02	0.03	0.01	0.01	0.002	0.01
	P_t-test	0.02	0.02	0.13	0.003	0.21	0.22	0.45	0.25
	P_perm	0.02	0.02	0.07	0.02	0.27	0.22	0.49	0.28
RP_p	Mean *r*	0.03	0.03	0.01	0.03	0.02	0.01	0.02	0.01
	P_t-test	0.02	0.03	0.17	0.008	0.18	0.30	0.17	0.24
	P_perm	0.03	0.03	0.13	0.01	0.19	0.28	0.14	0.24
C1	Mean *r*	0.03	0.05	0.02	0.03	0.02	0.04	0.03	0.04
	P_t-test	0.03	0.001	0.08	0.01	0.04	0.01	0.03	0.003
	P_perm	0.002	0.001	0.07	0.001	0.05	0.001	0.005	0.002
C2	Mean *r*	0.03	0.07	0.02	0.05	0.01	0.02	−0.01	0.01
	P_t-test	0.004	0.001	0.05	0.003	0.37	0.13	0.33	0.28
	P_perm	0.01	0.001	0.04	0.001	0.38	0.11	0.46	0.37
MO	Mean *r*	0.02	0.03	0.01	0.02	−0.01	−0.01	0.01	−0.03
	P_t-test	0.14	0.04	0.38	0.11	0.35	0.32	0.48	0.03
	P_perm	0.11	0.04	0.51	0.13	0.69	0.71	0.47	0.84
MP	Mean *r*	0.04	0.06	0.02	0.05	0.02	0.02	0.02	0.03
	P_t-test	0.005	0.0001	0.10	0.0001	0.10	0.15	0.08	0.06
	P_perm	0.006	0.0001	0.10	0.0001	0.11	0.10	0.15	0.09

Notes: Mean *r* indicates the average Spearman correlation between NS and the similarity matrix of each index of all subjects. P_*t*-test indicates the *P*-values of the simple *t*-tests, and the *P*_perm indicates the *P*-values based on the permutation tests. One-tailed tests were used. The estimates of the noise ceiling ranges are reported on the top, and the rest are the *t*-test results. RP_p, the rating-based phonological similarity correlated with NS with all the other visual and orthographic similarity being controlled for; MO, reading model-based orthographic similarity; MP, reading model-based phonological similarity; C1, C1 similarity in the H-MAX model; C2, C2 similarity in the H-MAX model.

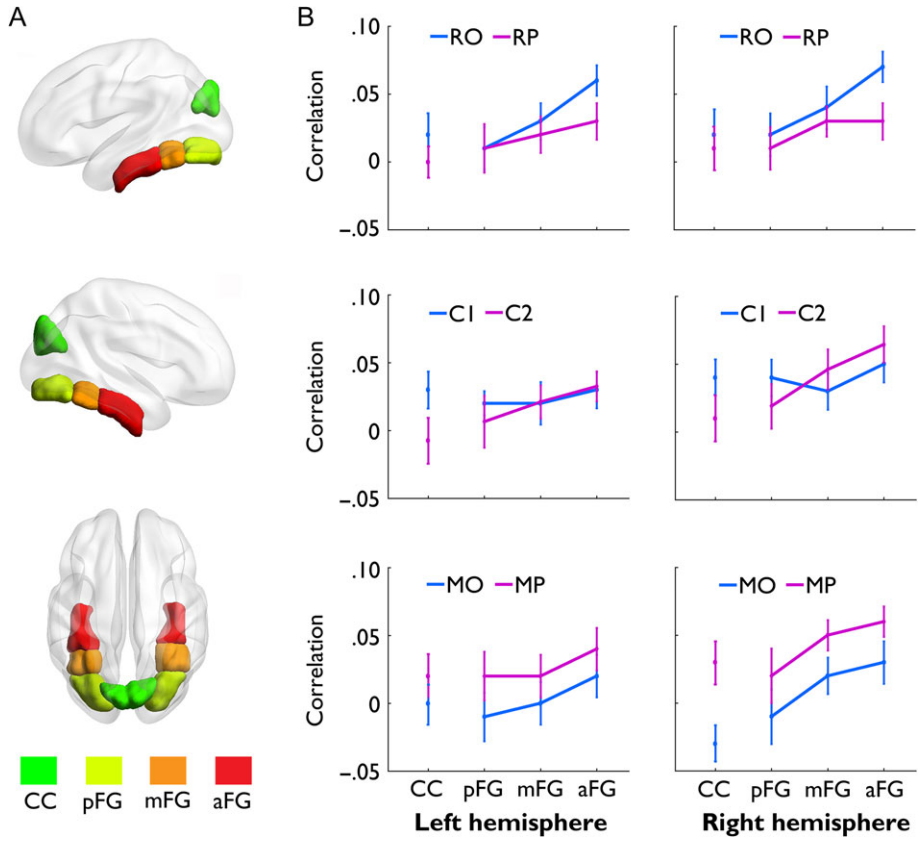

Figure 3. Representation of orthographic and phonological information in the CC, aFG, mFG, and pFG. (*A*) Left lateral (top), right lateral (middle), and bilateral medial view (bottom) of these ROIs in the standard brain is shown. (*B*) Orthographic information and phonological information specified by the subjective ratings (RO; RP) is shown in the top, the middle shows C1 and C2 information specified by the corresponding levels in the H-MAX model, and the bottom show orthographic information and phonological information specified by the reading models.

each individual subject rather than the mean ratings (Supplementary Table 1). The *t*-tests showed greater correlations with NS for RP than RO in the right aFG ($t(19) = 2.74$, $P = 0.01$) and left aFG ($t(19) = 1.82$, $P = 0.08$), but not in the mFG or pFG (Ps > 0.50).

The above analysis showed that the mFG and aFG represented both orthographic and phonological information. We further examined whether NS in these fusiform subregions showed a better fit to the joint orthographic and phonologic similarity than either orthographic or phonological similarity alone. Specifically, for each of these 4 ROIs, we compared 3 linear mixed-effect models with the fixed factor(s) being RO, RP, or both, respectively (subject was the random factor), using the likelihood ratio test. The results showed that the model including both RO and RP produced significantly better fit than the RP-only models (all Ps < 0.001), and the RO-only models (Ps < 0.03) except for the lmFG ($P = 0.39$).

The above results were based on structurally defined fusiform cortex. We also functionally defined the VWFA both at the group level and at the individual level (see Materials and Methods). As shown in Table 2, NS was significantly correlated with RO and RP, and the NS–RP correlation was significant after controlling for the behavioral and model similarities (RO, MO, C1, and C2). Using the same method of model comparison, we found that the model with both RO and RP as predictors produced significantly better fit to the data than for either the RO or RP models alone (all Ps < 0.01). To note, the head motion parameters (for more information of this, please refer to p. 1 of the supplementary materials) were controlled for in this analysis.

Increasing Lexical Representation Along the Posterior-to-Anterior Axis

Previous studies found that the fusiform gyrus showed increasingly stronger response to lexical information and fragments of words along the posterior-to-anterior axis (Vinckier et al. 2007; Xue and Poldrack, 2007), suggesting increasing lexical representation along that axis. To examine such a pattern with the RSA approach, linear trend analyses were conducted. In the first analysis, we examined whether there was a linear trend in the structurally defined pFG-to-aFG ROIs. The results showed a significant linear trend for orthographic information in both hemispheres (left: $F(1,19) = 8.35$, $P = 0.009$; right: $F(1,19) = 8.81$, $P = 0.008$), and a marginally significant trend for phonology in the left ($F(1,19) = 3.86$, $P = 0.07$), but not the right hemisphere ($F(1,19) = 1.36$, $P = 0.26$).

To examine the fine-grained posterior–anterior functional dissociation in the fusiform, we did a searchlight analysis to examine the representations of each voxel within the fusiform cortex. The extent of representations (*t*-values) was then correlated with their corresponding *y* coordinate. As shown in Figure 4, for both hemispheres, there was a significant linear increase from pFG-to-aFG in orthographic (left: $r = 0.44$, $P < 0.0001$; right: $r = 0.68$, $P < 0.0001$) and phonological representation (left: $r = 0.46$, $P < 0.0001$; right: $r = 0.58$, $P < 0.0001$).

Lower Level and Abstract Orthographic Representations in the Fusiform

We then examined whether, in addition to abstract word representations, lower level orthographic information is represented in the fusiform. We first tested whether there was increasingly abstract orthographic representation along the posterior-to-anterior axis. In this analysis, we correlated the NS in the fusiform subregions with the pattern similarity based on the H-MAX model (Riesenhuber and Poggio, 1999; Serre et al. 2005). In particular, S1 and C1 correspond to the simple and complex cells in V1, whereas and S2 and C2 correspond to neurons in V4 and posterior IT that represent composite complex features. Using functional imaging and RSA on scene and object processing, previous studies have revealed increasingly C2 representation along the visual stream, and strong C1 representation in the early visual cortex (Kriegeskorte et al. 2008a). If abstract orthographic representation increases along the posterior-to-anterior axis (Dehaene et al. 2005), we should expect that the NS is increasingly correlated with C2 representation and decreasingly correlated with C1 representation along that axis.

Table 2 Spearman correlations (and associated statistics) between NS and psychological and model similarities in the group-averaged and individual VWFAs

Index	Tests	RO	RP	RP_p	MO	MP	C1	C2
lgVWFA	Mean *r*	0.03	0.03	0.02	0.008	0.03	0.02	0.02
	P_*t*-test	0.02	0.02	0.03	0.32	0.02	0.05	0.08
	P_perm	0.03	0.02	0.02	0.28	0.03	0.04	0.06
rgVWFA	Mean *r*	0.03	0.03	0.03	0.002	0.03	0.03	0.03
	P_*t*-test	0.03	0.01	0.02	0.46	0.03	0.07	0.05
	P_perm	0.04	0.02	0.04	0.49	0.02	0.06	0.04
liVWFA	Mean *r*	0.04	0.03	0.03	0.004	0.04	0.03	0.03
	P_*t*-test	0.005	0.01	0.02	0.40	0.001	0.04	0.03
	P_perm	0.002	0.01	0.03	0.38	0.002	0.02	0.04
riVWFA	Mean *r*	0.03	0.04	0.04	0.004	0.04	0.03	0.03
	P_*t*-test	0.02	0.008	0.005	0.40	0.005	0.03	0.02
	P_perm	0.04	0.005	0.01	0.44	0.004	0.05	0.01

Notes: Mean *r* indicates the average Spearman correlation between NS and the similarity matrix of each index of all subjects. P_*t*-test indicates the *P*-values of the simple *t*-tests, and the P_perm indicates the *P*-values based on the permutation tests. One-tailed tests were used. lgVWFA, left group-averaged VWFA; rgVWFA, right group-averaged VWFA; liVWFA, left individual VWFA; riVWFA, right individual VWFA.

One-sample *t*-tests revealed significant correlations between C1 and NS in all 8 ROIs, but significant correlations between C2 and NS were found only in the bilateral aFG and mFG (Table 1). Linear trend analysis based on structural ROI results suggests increasing C2 representation along the posterior-to-anterior axis in both hemispheres (left: $F(1,19) = 4.19$, $P = 0.05$; right: $F(1,19) = 5.31$, $P = 0.03$), but no significant linear trend for C1 (left: $F(1,19) = 0.50$, $P = 0.49$; right: $F(1,19) = 0.84$, $P = 0.44$). Linear trend analysis based on searchlight results (Fig. 4) showed a strong increase in C2 representation (left: $r = 0.54$, $P < 0.0001$; right: $r = 0.72$, $P < 0.0001$), but a relatively weak increase in C1 representation (left: $r = 0.41$, $P < 0.001$; right: $r = 0.22$, $P < 0.001$). Direct comparison of Fisher's Z transformed correlational coefficients revealed greater linear trend for C2 than for C1 (left: $z = 2.48$, $P < 0.01$; right: $z = 10.06$, $P < 0.0001$). Taken together, these results suggest that whereas abstract orthographic information was increasingly represented along the posterior-to-anterior axis, lower level orthographic information was represented more evenly at every stage of the ventral visual stream.

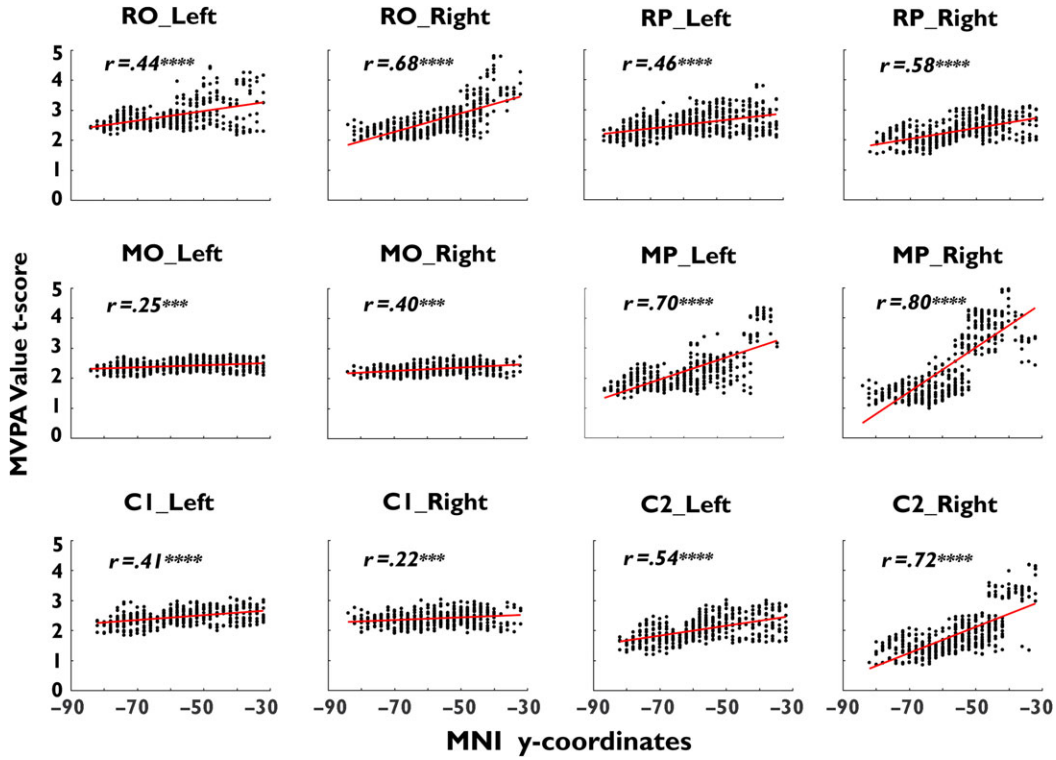

Figure 4. Linear gradient analysis based on the searchlight results of RO, RP, MO, MP, C1, and C2 information in the left and right fusiform. In each small panel, the *x*-axis denotes the MNI *y* coordinates, and the *y*-axis denotes the *t* scores of random-effects group analysis of the MVPA values for a given voxel. **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$.

Representational Connectivity in the Fusiform Cortex

The above analyses suggested that lower level and abstract orthographic, as well as phonological information were distributively represented across several brain regions. We further used representational connectivity analysis to examine how the representational structure in different regions were related to one another (Kriegeskorte et al. 2008a, 2008b). It should be noted that the connectivity pattern might have been affected by the global signal change or signal overlap between regions, so the results should be interpreted with caution as they might not necessarily reflect the true connectivity strength.

Our results revealed 3 interesting patterns (Figure 5A,B). First, there was strong representational connectivity between pFG and mFG, but low connectivity between the cuneus and pFG, and moderate connectivity between mFG and aFG. Second, strong cross-hemisphere and cross-region representational connectivity was found between pFG and contralateral mFG, suggesting that these regions play an important role in cross-hemisphere integration. Third and more importantly, we found that along the posterior-to-anterior axis, the left and right hemispheres represent increasingly different information ($F(2,38) = 9.96$, $P < 0.0001$).

One hypothesis for the low representational connectivity for the more aFG is that the 2 hemispheres are sensitive to different bands of visual frequencies (high frequency for the left and low frequency for the right) (Ivry and Robertson, 1998; Seghier and Price, 2011; Woodhead et al. 2011). To test this hypothesis, we performed high- and low-pass filtering on the images of Chinese characters (Fig. 5C) and correlated the C2 representation of the filtered images with NS in bilateral aFG. Consistent with the above hypothesis, we found significant hemisphere-by-frequency interaction ($F(1,19) = 4.66$, $P = 0.044$) (Fig. 5D). Although the left aFG showed a trend of better fit to C2 of the high-pass filtered images than the low-pass filtered images ($t(19) = 1.76$, $P = 0.08$), the raFG showed a reversed pattern ($t(19) = -1.42$, $P = 0.12$).

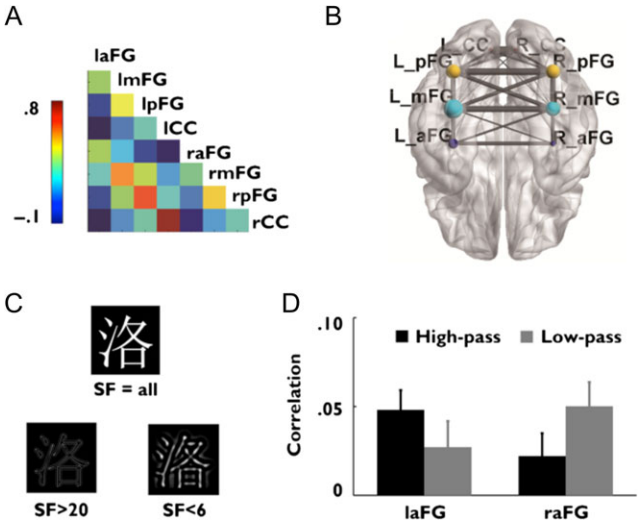

Figure 5. Representational connectivity of the 8 ROIs. (*A*) The Spearman correlations between the NS patterns during the functional state are shown. (*B*) The connectivity patterns in the standard brain template, with the thickness of the line between every 2 regions representing the strength of the connection are shown. (*C*) One-sample character (洛, luo4) that was untreated (top), high-pass filtered (bottom left), and low-pass filtered (bottom right) is shown. (*D*) C2 information in the left and the right aFG as specified by the C2 responses of the high-pass filtered and low-pass filtered images in the H-MAX model is shown.

Testing the Isomorphism of Model, Psychological, and Neural Representations

In the following analysis, we examined whether the computational model's representations of Chinese characters (Xing and Li, 2004; Yang et al. 2009) are isomorphic to the psychological and neural representations. For orthographic representation, we first examined the correspondence between the representation in the reading model (i.e., MO, see Materials and Methods for details), the C2 layer of the H-MAX model, and the subjective ratings (RO). As shown in Fig. 6, there were significant correlations between MO and RO (word set 1: $r = 0.50$, $P < 0.0001$; word set 2: $r = 0.46$, $P < .0001$), between C2 and RO (word set 1: $r = .61$, $P < 0.0001$; word set 2: $r = 0.53$, $P < 0.0001$), and between MO and C2 (word set 1: $r = 0.48$, $P < 0.0001$; word set 2: $r = 0.35$, $P < 0.0001$). Direct comparison of Fisher's Z transformed correlational coefficients revealed stronger correlations between RO and C2 than those between MO and RO (word set 1: $z = 2.44$, $P = 0.007$; word set 2: $z = 1.55$, $P = 0.06$) and between MO and C2 (word set 1: $z = 3.11$, $P = 0.0001$; word set 2: $z = 3.75$, $P = 0.0001$).

Second, we examined whether the MO, RO, and C2 representations were equally isomorphic to the NS in the fusiform. MO was only significantly associated with NS in the right aFG (Table 1). The NS's correlations with MO were significantly lower than those with RO in all 4 regions, and significantly or marginally lower than those with C2 in the left aFG and the left mFG (Table 3).

Finally, we tested whether there was a similar posterior-to-anterior gradient in the representation of orthographic information in MO. In contrast to RO and C2, which showed a significant linear increase in both hemispheres, the analysis based on the structural ROI results showed only a marginally significant linear increase along the posterior-to-anterior axis in the

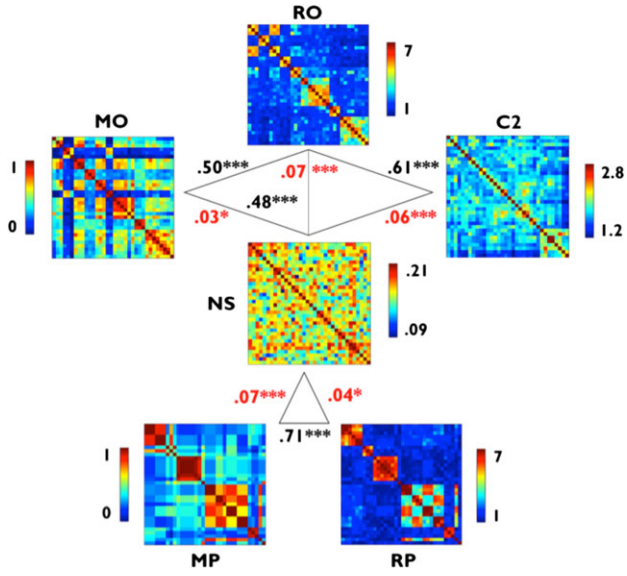

Figure 6. The isomorphism of model, psychological, and neural representations. For orthography (upper part), model-based similarity (MO) was correlated with RO, C2 similarity (C2), and NS. C2 was also correlated with RO. For phonology (lower part), model-based similarity (MP) was correlated with RP and NS. The correlations between NS and the behavioral and model similarities are shown in red, and those between behavioral and model similarities are shown in black. The asterisks indicate significance levels of the correlations relative to zero; *$P < 0.05$; ***$P < 0.001$. Please note that due to space limitation, this figure presents the data from the word set 1. The result for word set 2 is very consistent with this and is presented in Supplementary Fig. 4.

Table 3 Comparing psychological similarities with similarities based on Chinese reading models and visual object recognition model

Region	MO versus RO		MO versus C2		C2 versus RO		MP versus RP	
	t(19)	*P*	*t*(19)	*P*	*t*(19)	*P*	*t*(19)	*P*
Left aFG	−2.88	0.004	−0.89	0.19	−2.55	0.01	0.92	0.38
Right aFG	−1.83	0.04	−1.42	0.09	−0.53	0.30	2.36	0.02
Left mFG	−3.62	0.002	−1.33	0.10	−100.05	0.15	0.30	0.76
Right mFG	−1.67	0.06	−1.59	0.06	−0.25	0.31	200.20	0.04

right hemisphere ($F(1,19) = 2.76$, $P = 0.08$), but not in the left hemisphere ($F(1,19) = 1.52$, $P = 0.23$). The analysis based on the searchlight results (Fig. 4) also showed a significant linear trend for both hemispheres (left: $r = 0.25$, $P < 0.001$; right: $r = 0.40$, $P < 0.001$). However, direct comparison of Fisher's Z transformed correlational coefficients revealed a greater linear increase for RO than for MO (left: $z = 3.19$, $P < 0.001$; right: $z = 5.97$, $P < 0.0001$), and a greater increase for C2 than MO as well (left: $z = 5.13$, $P < 0.0001$; right: $z = 7.12$, $P < 0.0001$). Together, these results suggest that the orthographic representation in the reading model did not fit well with the representations from psychological data, neural data, or the H-MAX model.

For phonological representation, we found that the representation in the reading model (MP) showed strong correlations with RP (word set 1: $r = 0.76$, $P < 0.0001$; word set 2: $r = 0.58$, $P < 0.0001$). One-sample *t*-tests revealed that MP showed a similar pattern with RP, that is, it was significantly correlated with NS in the bilateral aFG and the right mFG, but not in the left mFG, bilateral pFG or cuneus cortex (Table 1). However, paired-samples *t*-tests revealed that MP showed stronger correlations with NS than RP in the right aFG and the right mFG (Table 3), suggesting that the model outperformed subjective ratings in capturing the phonological representations in the brain. Linear trend analysis found stronger representation of MP along the posterior-to-anterior axis in the right hemisphere ($F(1,19) = 3.23$, $P = 0.009$), but not in the left hemisphere ($F(1,19) = 1.14$, $P = 0.29$). The analysis based on the searchlight results (Fig. 4) showed a significant linear trend for both hemispheres (left: $r = 0.70$, $P < 0.0001$; right: $r = 0.80$, $P < 0.0001$).

Discussion

Given the central role of representation in human cognition, a growing number of studies have strived to examine the nature of neural representations in areas such as vision (Kriegeskorte et al. 2008b; Connolly et al. 2012), categorization (Mack et al. 2013), memory (Xue et al. 2010; Davis et al. 2014), and language (Tyler et al. 2013; Clarke and Tyler, 2014). By bridging different levels of analysis, these studies not only deepen our understanding of the nature of neural representations, but also provide important neural constraints to adjudicate competing computational models (Forstmann et al. 2011; Kriegeskorte and Kievit, 2013; Mack et al. 2013). This study extended this line of research to orthographic and phonological representations in the fusiform gyrus, and tested a computational model of Chinese word reading. By relating NS pattern with the similarity patterns based on psychological ratings and computational models, our results help to resolve 2 central debates regarding the role of the fusiform in reading, and also inform the development of psychological, neural, and computational theories of reading.

Our study obtained clear evidence that the mid-fusiform represents not only orthography but also phonology. By independently manipulating the similarity structure of each component (orthography vs. phonology) among carefully selected Chinese characters, and associating NS pattern in the fusiform with the phonological similarity matrices from psychological ratings and computational models, we found significant associations between phonological similarity and NS in the fusiform. Moreover, this representation is specific to the mFG and aFG, but not the pFG or the primary visual cortex. Although previous studies using univariate methods have found elevated fusiform activation in tasks that require phonological processing (Booth et al. 2002; Price et al. 2003; Cohen and Dehaene, 2004; Xue et al. 2006; Dehaene et al. 2010; Desroches et al. 2010; Yoncheva et al. 2010), it is not clear whether it reflects phonological processing per se or the increased demand for orthographic processing to aid the phonological task (Dehaene and Cohen, 2011), because the 2 processes could not be separated according to the connectionist perspective (Seidenberg and McClelland, 1989). The RSA approach used in this study effectively addressed this limitation and the results help to resolve the long-standing debate regarding the role of the left fusiform in orthographic and phonological processing (Xue et al. 2006; Dehaene and Cohen, 2011; Price and Devlin, 2011).

The joint representation of phonology and orthography in the fusiform gyrus may provide a neural basis for fluent and efficient visual word decoding in adults. First, this observation suggests that the fusiform plays a more central role in integrating high-order visual form with phonology (Price and Devlin, 2011), thus serves as the reading skill area (Pugh et al. 2001). Consequently, activation in this region reflects the effects of orthographic–phonological training (Sandak et al. 2004; Xue et al. 2006) and this region is less active for people with reading difficulties (Pugh et al. 2001; McCandliss et al. 2003; McCrory et al. 2005). Second, this joint representation of orthography and phonology, presumably acquired by extensive exposure, could lead to greater word-level specificity (Glezer et al. 2015), thus less interference by words with similar orthography or phonology, and consequently more fluent reading (Regier, 2005; McMurray et al. 2010). Third, this idea of joint representations is also consistent with the observations that the form-sound correspondence rules shape the functional asymmetry of the fusiform (Mei et al. 2013, 2015). Fourth, this joint representation also provides a neural basis for the case-insensitive activity (Polk and Farah, 2002) and representation (Rothlein and Rapp, 2014), at least when the lower and upper cases differ significantly in visual appearances.

We also found that there is increased representation of orthography along the posterior-to-anterior axis in the fusiform region. This finding is consistent with existing activation-based results, which show a posterior-to-anterior progression of selectivity from individual letters to bigrams and morphemes (Vinckier et al. 2007). Priming studies also suggest that only the aFG subregion shows strong tuning to whole words (Glezer et al. 2009). Thus, the evidence from activation level, priming, and

multiple voxel pattern representation converges to suggest a hierarchical orthographic representation of visual words in the fusiform cortex. Indeed, this hierarchical representation is a general feature in visual object recognition. For example, higher order categorical representation of objects has been found in the lateral occipital cortex and fusiform, but not in the early visual cortex (Kriegeskorte et al. 2008a; Connolly et al. 2012).

Our results also suggest that orthographic representation becomes more abstract along the posterior-to-anterior gradient. The H-MAX-C2 feature representation, obtained by using a nonlinear max pooling mechanism, showed invariance to size, position, and view. Consistent with the conjecture that the representation in this layer may be comparable to those found in primate V4 and posterior IT, previous studies have found that the H-MAX-C2 representation of natural image patches provides a better fit to the activation pattern in the right FFA than does the C1 representation (Kriegeskorte et al. 2008a, 2008b). Our finding that the bilateral aFG showed the best fit to the H-MAX-C2 representation is consistent with this observation. Also consistent is the finding of cross-font priming effects in the aFG (Gauthier et al. 2000; Qiao et al. 2010).

Unlike the VWFA hypothesis and the LCD proposal that emphasize the abstract orthographic representation in the aFG (Dehaene et al. 2005; Dehaene and Cohen, 2011), our results suggest that the aFG cortex also represents lower level features encoded in the C1 layer of the H-MAX model. This echoes previous findings that the early visual cortex and the right fusiform cortex showed comparable levels of fit to the representation in V1 model during object recognition, and that the representation in the right fusiform cortex also resembled the representation in the early visual cortex (Kriegeskorte et al. 2008a). Our result is also consistent with another study that found position-sensitive response to English words in the fusiform using the multivoxel pattern classification technique (Rauschecker et al. 2012). In a separate line of research, it has been found that the neural patterns in scene-selective regions (e.g., parahippocampus place area) were linked to low-level image properties such as spatial frequency and orientation (Rajimehr et al. 2011; Nasr and Tootell, 2012; Watson et al. 2014). Taken together, these findings suggest that the higher level regions may represent both lower level and higher level information.

The observation that the aFG contains both abstract orthographic representation as well as low-level orthographic features could shed light on the mechanisms of the invariance of word recognition, including font, size, case, and even script, which is an important feature of fluent reading. Based on the hierarchical organization of the visual cortex (Riesenhuber and Poggio, 1999), one influential mechanism involves the LCD are increasingly more sensitive to larger fragments of words (Dehaene et al. 2005; Dehaene and Cohen, 2011). The second mechanism is the storage of multiple interconnected representations in this fusiform subregion, which could correct one another and could thus arrive at the same interpretation despite minor deviations of visual appearance of the words (McClelland and Rumelhart, 1981; Rumelhart and McClelland, 1982; Price and Devlin, 2011). This mechanism could not only account for the case, font, size, and even script invariance, but also is compatible with the observation that lower level features are represented in this subregion. Similar mechanisms have been proposed to account for the view-independent object recognition (Logothetis et al. 1994). Future research should examine how the intrinsic organization of the visual system and learning jointly modulate lexical representation in the fusiform cortex (Xue et al. 2006; Dehaene et al. 2010; Wandell et al. 2012).

One additional interesting finding from this study is that although the orthographic and phonological information was represented in both hemispheres, there was clear hemispheric differentiation in the more anterior subregions. Consistent with the idea that the left and right hemispheres are tuned for high- versus low-frequency information, respectively (Ivry and Robertson, 1998), the representation in left aFG showed better fit to high-pass filtered C2 representation than did that in the right aFG. This corroborates existing observations of leftward lateralization for high-frequency images and gratings (e.g., closer to 7 cycles per degree) and rightward lateralization for low-frequency images and gratings (closer to 5 cycles per degree) (Woodhead et al. 2011). One advantage of our method is that the stimuli shown to the subjects were not high- or low-pass filtered, which avoided confounding factors such as processing strategies or task difficulties.

Our results showed little correspondence between NS and the similarity space defined by a set of orthographic representations used in a computational model of Chinese reading (Yang et al. 2009). One possibility is that these representations diverge in important ways from the neural representations in the aFG, although they are in concord with subjective similarity ratings. As noted in the Materials and Methods section, orthographic representations in the Chinese reading model incorporated some design features that deviated from the actual visual similarity structure. Indeed, the representations in the aFG were better captured by a general model of the ventral visual stream than by the more abstract reading model. The greater correlations of the H-MAX C2 layer representation with both rating-based and neural similarities suggest that the realism of representations for computational models of reading could be improved by considering how visual input is transformed along the ventral visual stream. In particular, future models should simultaneously represent both lower and higher level visual features by storing multiple slightly different orthographic copies of the same words.

Although RSA is a powerful tool for linking different levels of analyses, the correlations we found between model-based or behavioral similarities and NS was generally low. They were in the range of 0.03–0.08, which was consistent with previous studies in various visual and cognitive domains (Kriegeskorte et al. 2008a; Bruffaerts et al. 2013; Clark and Tyler, 2014; Peelen et al. 2014; Baeck et al. 2015). One reason for the low correlations in RSA is that the single-trial response estimation based on the low signal noise ratio (SNR) fMRI data is necessarily very noisy. Consistent with this explanation, our calculation of noise celling suggested that, given the level of noise in the neural data, we could expect the maximum correlations to be around 0.3. Although more repetitions could potentially improve SNR, they could also lead to psychological and neural repetition priming effects, which may contaminate the estimation of mental representation. In addition, as the observed correlations were lower than the expected maximum correlations, we suspect that some additional factors may have further obscured the correlations between NS and behavioral and model similarities. First, the fusiform may represent many different features, including lower and higher level orthographic, phonological, and perhaps semantic information. Second, neural representations could be very sparse and distributed, which creates great challenges for RSA. Finally, fMRI's spatial resolution is still relatively limited for the purpose of examining mental representations. Future studies should use high-resolution imaging, in

combination with better feature selection methods, to study neural representations in the brain.

Several other important questions need to be studied in the future. First, this study only examined Chinese word representation using a naming task, future studies should examine lexical representations in different writing systems (Siok et al. 2004) and with different tasks (Yang et al. 2012; Yang and Zevin, 2014). Second, future research should investigate all lexical information, including orthography, phonology, and semantics (Bruffaerts et al. 2013; Clarke and Tyler, 2014) that may be jointly represented in the fusiform. Finally, future research should examine how word representations in the fusiform develop as a result of reading experience (Glezer et al. 2015) and whether their individual differences are associated with reading abilities (Boets et al. 2013; He et al. 2013).

To summarize, this study provides important novel evidence to further our understanding of the nature of orthographic and phonological representations in the brain, with significant implications for the behavioral, neural, and computational theories of reading. More broadly, our study suggests that the integration of representations at different levels (e.g., behavioral, neural, and model) provides a powerful tool to advance the mechanistic understanding of cognition.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/

Funding

National Natural Science Foundation of China (31130025); the 973 Program (2014CB846102); the 111 Project (B07008); the National Natural Science Foundation of China (31221003, 31200841, 1621136008); and German Research Foundation (DFG TRR-169).

Note

We would also like to thank Dr Malte Rasch for his help with the H-MAX model. *Conflict of Interest*: None declared.

References

Baeck A, Kravitz D, Baker C, de Beeck HPO. 2015. Influence of lexical status and orthographic similarity on the multi-voxel response of the visual word form area. NeuroImage. 111: 321–328.

Boets B, de Beeck HPO, Vandermosten M, Scott SK, Gillebert CR, Mantini D, Bulthé J, Sunaert S, Wouters J, Ghesquière P. 2013. Intact but less accessible phonetic representations in adults with dyslexia. Science. 342:1251–1254.

Booth JR, Burman DD, Meyer JR, Gitelman DR, Parrish TB, Mesulam MM. 2002. Functional anatomy of intra-and cross-modal lexical tasks. Neuroimage. 16:7–22.

Bruffaerts R, Dupont P, Peeters R, De Deyne S, Storms G, Vandenberghe R. 2013. Similarity of fMRI activity patterns in left perirhinal cortex reflects semantic similarity between words. J Neurosci. 33:18597–18607.

Chee MW, Soon CS, Lee HL. 2003. Common and segregated neuronal networks for different languages revealed using functional magnetic resonance adaptation. J Cogn Neurosci. 15:85–97.

Clarke A, Tyler LK. 2014. Object-specific semantic coding in human perirhinal cortex. J Neurosci. 34:4766–4775.

Cohen L, Dehaene S. 2004. Specialization within the ventral stream: the case for the visual word form area. Neuroimage. 22:466–476.

Cohen L, Lehericy S, Chochon F, Lemer C, Rivaud S, Dehaene S. 2002. Language-specific tuning of visual cortex? Functional properties of the visual word form area. Brain. 125: 1054–1069.

Coltheart M, Rastle K, Perry C, Langdon R, Ziegler J. 2001. DRC: a dual route cascaded model of visual word recognition and reading aloud. Psychol Rev. 108:204–256.

Connolly AC, Guntupalli JS, Gors J, Hanke M, Halchenko YO, Wu Y-C, Abdi H, Haxby JV. 2012. The representation of biological classes in the human brain. J Neurosci. 32:2608–2618.

Davis T, Xue G, Love BC, Preston AR, Poldrack RA. 2014. Global neural pattern similarity as a common basis for categorization and recognition memory. J Neurosci. 34:7472–7484.

Dehaene S, Cohen L. 2011. The unique role of the visual word form area in reading. Trends Cogn Sci. 15:254–262.

Dehaene S, Cohen L, Sigman M, Vinckier F. 2005. The neural code for written words: a proposal. Trends Cogn Sci. 9: 335–341.

Dehaene S, Jobert A, Naccache L, Ciuciu P, Poline JB, Le Bihan D, Cohen L. 2004. Letter binding and invariant recognition of masked words: behavioral and neuroimaging evidence. Psychol Sci. 15:307–313.

Dehaene S, Naccache L, Cohen L, Le Bihan D, Mangin JF, Poline JB, Riviere D. 2001. Cerebral mechanisms of word masking and unconscious repetition priming. Nat Neurosci. 4: 752–758.

Dehaene S, Pegado F, Braga LW, Ventura P, Nunes Filho G, Jobert A, Dehaene-Lambertz G, Kolinsky R, Morais J, Cohen L. 2010. How learning to read changes the cortical networks for vision and language. Science. 330:1359–1364.

Dell GS, Juliano C. 1996.Phonological encoding. In: Dijkstra T, DeSmedt K, editors.Computational psycholinguistics: symbolic and connectionist models of language processing. London: Harvester-Wheatsheaf.

Desroches AS, Cone NE, Bolger DJ, Bitan T, Burman DD, Booth JR. 2010. Children with reading difficulties show differences in brain regions associated with orthographic processing during spoken language processing. Brain Res. 1356: 73–84.

Forstmann BU, Wagenmakers E-J, Eichele T, Brown S, Serences JT. 2011. Reciprocal relations between cognitive neuroscience and formal cognitive models: opposites attract? Trends Cogn Sci. 15:272–279.

Gauthier I, Tarr MJ, Moylan J, Skudlarski P, Gore JC, Anderson AW. 2000. The fusiform “face area” is part of a network that processes faces at the individual level. J Cogn Neurosci. 12: 495–504.

Glezer LS, Jiang X, Riesenhuber M. 2009. Evidence for highly selective neuronal tuning to whole words in the “visual word form area”. Neuron. 62:199–204.

Glezer LS, Kim J, Rule J, Jiang X, Riesenhuber M. 2015. Adding words to the brain’s visual dictionary: novel word learning selectively sharpens orthographic representations in the VWFA. J Neurosci. 35:4965–4972.

Glezer LS, Eden G, Jiang X, Luetje M, Napoliello E, Kim J, Riesenhuber M. 2016. Uncovering phonological and orthographic selectivity across the reading network using fMRI-RA. NeuroImage. 138:248–256.

Goswami U. 2000. Phonological representations, reading development and dyslexia: towards a cross-linguistic theoretical framework. Dyslexia. 6:133–151.

Hahn U, Bailey TM. 2005. What makes words sound similar? Cognition. 97:227–267.
Harm MW, Seidenberg MS. 1999. Phonology, reading acquisition, and dyslexia: insights from connectionist models. Psychol Rev. 106:491–528.
He Q, Xue G, Chen C, Chen C, Lu Z-L, Dong Q. 2013. Decoding the neuroanatomical basis of reading ability: a multivoxel morphometric study. J Neurosci. 33:12835–12843.
Hirshorn EA, Li Y, Ward MJ, Richardson RM, Fiez JA, Ghuman AS. 2016. Decoding and disrupting left midfusiform gyrus activity during word reading. Proc Natl Acad Sci USA. doi:10.1073/pnas.1604126113.
Huang B, Liao X. 2002. XianDaiHanYu (3rd Supplement). Beijing: Higher Education Press.
Ivry RB, Robertson LC.1998. The two sides of perception. MIT Press.
Kriegeskorte N, Kievit RA. 2013. Representational geometry: integrating cognition, computation, and the brain. Trends Cogn Sci. 17:401–412.
Kriegeskorte N, Mur M, Bandettini P. 2008a. Representational similarity analysis–connecting the branches of systems neuroscience. Front Syst Neurosci. 2:4.
Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA. 2008b. Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron. 60:1126–1141.
Liu Y, Hao M, Shu H, Tan LH, Weekes BS. 2008. Age-of-acquisition effects on oral reading in Chinese. Psychon B Rev. 15:344–350.
Logothetis NK, Pauls J, Bülthoff HH, Poggio T. 1994. View-dependent object recognition by monkeys. Curr Biol. 4:401–414.
Mack ML, Preston AR, Love BC. 2013. Decoding the brain's algorithm for categorization from its neural implementation. Curr Biol. 23:2023–2027.
Marr D. 1982. A computational investigation into the human representation and processing of visual information. WH San Francisco: Freeman and Company.
McCandliss BD, Cohen L, Dehaene S. 2003. The visual word form area: expertise for reading in the fusiform gyrus. Trends Cogn Sci. 7:293–299.
McClelland JL, Rumelhart DE. 1981. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. Psychol Rev. 88:375–407.
McCrory EJ, Mechelli A, Frith U, Price CJ. 2005. More than words: a common neural basis for reading and naming deficits in developmental dyslexia? Brain. 128:261–267.
McMurray B, Jongman A. 2010. What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. Psycho Rev. 118:219–246.
McMurray B, Samelson VM, Lee SH, Bruce Tomblin J. 2010. Individual differences in online spoken word recognition: implications for SLI. Cogn Psychol. 60:1–39.
Mei L, Xue G, Lu Z-L, Chen C, Wei M, He Q, Dong Q. 2015. Long-term experience with Chinese language shapes the fusiform asymmetry of English reading. NeuroImage. 110:3–10.
Mei L, Xue G, Lu Z-L, He Q, Zhang M, Xue F, Chen C, Dong Q. 2013. Orthographic transparency modulates the functional asymmetry in the fusiform cortex: an artificial language training study. Brain Lang. 125:165–172.
Mumford JA, Turner BO, Ashby FG, Poldrack RA. 2012. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage. 59:2636–2643.
Nasr S, Tootell RBH. 2012. Role of fusiform and anterior temporal cortical areas in facial recognition. Neuroimage. 63:1743–1753.
Nestor A, Behrmann M, Plaut DC. 2013. The neural basis of visual word form processing: a multivariate investigation. Cereb Cortex. 23:1673–1684.
Nili H, Wingfield C, Walther A, Su L, Marslen-Wilson W, Kriegeskorte N. 2014. A toolbox for representational similarity analysis. PLoS Comput Biol. 10 (4):e1003553.
Peelen MV, He C, Han Z, Caramazza A, Bi Y. 2014. Nonvisual and visual object shape representations in occipitotemporal cortex: evidence from congenitally blind and sighted adults. J Neurosci. 34:163–170.
Plaut DC, McClelland JL, Seidenberg MS, Patterson K. 1996. Understanding normal and impaired word reading: computational principles in quasi-regular domains. Psychol Rev. 103:56–115.
Polk TA, Farah ML. 2002. Functional MRI evidence for an abstract, not perceptual, word-form area. J Exp Psychol Gen. 131:65–72.
Price CJ, Devlin JT. 2011. The interactive account of ventral occipitotemporal contributions to reading. Trends Cogn Sci. 15:246–253.
Price CJ, Winterburn D, Giraud AL, Moore CJ, Noppeney U. 2003. Cortical localisation of the visual and auditory word form areas: a reconsideration of the evidence. Brain Lang. 86:272–286.
Pugh KR, Mencl WE, Jenner AR, Katz L, Frost SJ, Lee JR, Shaywitz SE, Shaywitz BA. 2001. Neurobiological studies of reading and reading disability. J Commun Disord. 34:479–492.
Qiao E, Vinckier F, Szwed M, Naccache L, Valabregue R, Dehaene S, Cohen L. 2010. Unconsciously deciphering handwriting: subliminal invariance for handwritten words in the visual word form area. Neuroimage. 49:1786–1799.
Rajimehr R, Devaney KJ, Bilenko NY, Young JC, Tootell RBH. 2011. The "Parahippocampal Place Area" responds preferentially to high spatial frequencies in humans and monkeys. Plos Biol. 9:e1000608.
Ramus F, Szenkovits G. 2008. What phonological deficit? Q J Exp Psychol. 61:129–141.
Rauschecker AM, Bowen RF, Parvizi J, Wandell BA. 2012. Position sensitivity in the visual word form area. Proc Natl Acad Sci USA. 109:E1568–E1577.
Regier T. 2005. The emergence of words: attentional learning in form and meaning. Cogn Sci. 29:819–865.
Riesenhuber M, Poggio T. 1999. Hierarchical models of object recognition in cortex. Nat Neurosci. 2:1019–1025.
Rothlein D, Rapp B. 2014. The similarity structure of distributed neural responses reveals the multiple representations of letters. NeuroImage. 89:331–344.
Rumelhart DE, McClelland JL. 1982. An interactive activation model of context effects in letter perception 2. The contextual enhancement effect and some tests and extensions of the model. Psychol Rev. 89:60–94.
Sandak R, Mencl WE, Frost SJ, Rueckl JG, Katz L, Moore DL, Mason SA, Fulbright RK, Constable RT, Pugh KR. 2004. The neurobiology of adaptive learning in reading: a contrast of different training conditions. Cogn Affect Behav Neurosci. 4:67–88.
Seghier ML, Price CJ. 2011. Explaining left lateralization for words in the ventral occipitotemporal cortex. J Neurosci. 31:14745–14753.
Seidenberg MS, McClelland JL. 1989. A distributed, developmental model of word recognition and naming. Psychol Rev. 96:523–568.

Serre T, Wolf L, Poggio T (2005) Object recognition with features inspired by visual cortex. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, CVPR 2005, pp 994–1000.

Siok WT, Perfetti CA, Jin Z, Tan LH. 2004. Biological abnormality of impaired reading is constrained by culture. Nature. 431: 71–76.

Snyder PJ, Harris LJ. 1993. Handedness, sex, and familial sinistrality effects on spatial tasks. Cortex. 29:115–134.

Tyler LK, Cheung TPL, Devereux BJ, Clarke A. 2013. Syntactic computations in the language network: characterizing dynamic network properties using representational similarity analysis. Front Psychol. 4:271.

Vinckier F, Dehaene S, Jobert A, Dubus JP, 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.

Wandell BA. 2011. The neurobiological basis of seeing words. Ann NY Acad Sci. 1224:63–80.

Wandell BA, Rauschecker AM, Yeatman JD. 2012. Learning to see words. Ann Rev Psychol. 63:31–53.

Wang Q, Luo S, Monterosso J, Zhang J, Fang X, Dong Q, Xue G. 2014. Distributed value representation in the medial prefrontal cortex during intertemporal choices. J Neurosci. 34 (22):7522–7530.

Watson DM, Hartley T, Andrews TJ. 2014. Patterns of response to visual scenes are linked to the low-level properties of the image. Neuroimage. 99:402–410.

Woodhead ZVJ, Wise RJS, Sereno M, Leech R. 2011. Dissociation of sensitivity to spatial frequency in word and face preferential areas of the fusiform gyrus. Cereb Cortex. 21 (10): 2307–2312.

Xia M, Wang J, He Y. 2013. BrainNet Viewer: a network visualization tool for human brain connectomics. PloS one. 8:e68910.

Xing HB, Li P. 2004. The acquisition of Chinese characters: corpus analyses and connectionist simulations. J Cogn Sci. 5:1–49.

Xue G, Chen C, Jin Z, Dong Q. 2006. Language experience shapes fusiform activation in processing a logographic artificial language: an fMRI training study. Neuroimage. 31:1325–1336.

Xue G, Dong Q, Chen C, Lu Z, Mumford JA, Poldrack RA. 2010. Greater neural pattern similarity across repetitions is associated with better memory. Science. 330:97–101.

Xue G, Poldrack RA. 2007. The neural substrates of orthographic learning: implication for the VWFA hypothesis. J Cogn Neurosci. 19:1643–1655.

Yang J, McCandliss BD, Shu H, Zevin JD. 2009. Simulating language-specific and language-general effects in a statistical learning model of Chinese reading. J Mem Lang. 61: 238–257.

Yang J, Shu H, McCandliss BD, Zevin JD. 2013. Orthographic influences on division of labor in learning to read Chinese and English: insights from computational modeling. Biling Lang Cogn. 16:354–366.

Yang J, Wang X, Shu H, Zevin JD. 2012. Task by stimulus interactions in brain responses during Chinese character processing. NeuroImage. 60:979–990.

Yang J, Zevin J. 2014. The impact of task demand on visual word recognition. Neurosci. 272:102–115.

Yoncheva YN, Zevin JD, Maurer U, McCandliss BD. 2010. Auditory selective attention to speech modulates activity in the visual word form area. Cereb Cortex. 20:622–632.

Zhou X, Marslen-Wilson W. 1999. The nature of sublexical processing in reading Chinese characters. J Exp Psychol Learn Mem Cogn. 25:819–837.