Contents lists available at ScienceDirect

NeuroImage

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

Mechanisms underlying category learning in the human ventral occipito-temporal cortex

Xiangqi Luo [a,1], Mingyang Li [b,1], Jiahong Zeng [a], Zhiyun Dai [a], Zhenjiang Cui [a], Minhong Zhu [a], Mengxin Tian [a], Jiahao Wu [a], Zaizhu Han [a,*]

[a] State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, PR China
[b] Key Laboratory for Biomedical Engineering of Ministry of Education, Department of Biomedical Engineering, College of Biomedical Engineering & Instrument Science, Zhejiang University, Hangzhou 310027, PR China

ARTICLE INFO

Keywords:
VOTC
Category learning
Meaningless stimulus
Nonvisual features
Category dissociation

ABSTRACT

The human ventral occipito-temporal cortex (VOTC) has evolved into specialized regions that process specific categories, such as words, tools, and animals. The formation of these areas is driven by bottom-up visual and top-down nonvisual experiences. However, the specific mechanisms through which top-down nonvisual experiences modulate category-specific regions in the VOTC are still unknown. To address this question, we conducted a study in which participants were trained for approximately 13 h to associate three sets of novel meaningless figures with different top-down nonvisual features: the wordlike category with word features, the non-wordlike category with nonword features, and the visual familiarity condition with no nonvisual features. Pre- and post-training functional MRI (fMRI) experiments were used to measure brain activity during stimulus presentation. Our results revealed that training induced a categorical preference for the two training categories within the VOTC. Moreover, the locations of two training category-specific regions exhibited a notable overlap. Remarkably, within the overlapping category-specific region, training resulted in a dissociation in activation intensity and pattern between the two training categories. These findings provide important insights into how different nonvisual categorical information is encoded in the human VOTC.

1. Introduction

Rapid and precise categorization of a visual stimulus is usually crucial for the survival and reproduction of animals (Thorpe et al., 1996). Evidence has shown that the ventral occipitotemporal cortex (VOTC), an important area for processing visual stimuli in humans and primates, can efficiently discriminate stimuli from different categories (e.g., tools, animals, faces, and words; Mahon and Caramazza, 2009; Grill-Spector and Weiner, 2014; Bi et al., 2016). A hot and important scientific question is how the VOTC achieves this functionality.

Previous studies have indicated that the formation of these regions is driven not only by bottom-up visual modalities (Hasson et al., 2002; Nasr et al., 2014; Srihasam et al., 2014; Arcaro et al., 2017) but also by top-down nonvisual sensory and motor modalities (Price and Devlin, 2011; van den Hurk et al., 2017; Taylor et al., 2019). For example, braille reading in congenitally blind individuals activates the visual word form area (VWFA) (Reich et al., 2011; Kim et al., 2017; Mattioni et al., 2020). Furthermore, evidence suggests that these regions are modulated by top-down information from other high-level regions (Chen et al., 2019; Li et al., 2020; Liu et al., 2021). This modulation occurs through connectivity between category-specific regions in the VOTC and high-level regions, known as the connectivity hypothesis (Mahon and Caramazza, 2011; Hannagan et al., 2015; Li et al., 2018; Op de Beeck et al., 2019). Crucially, the structure of this connectivity network is innate and present from birth (Saygin et al., 2011; Osher et al., 2016; Saygin et al., 2016; Mars et al., 2018). However, this network does not manifest unless top-down information is learned (Li et al., 2020). For instance, word-specific areas in the VOTC do not emerge in humans without literacy (Dehaene et al., 2010; Dehaene et al., 2015; Dehaene-Lambertz et al., 2018). In essence, category-specific regions in the VOTC are innate at birth, but their initiation is influenced by postbirth learning experiences (Bracci and Op de Beeck, 2016; Op de

* Corresponding author at: State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, PR China.
E-mail address: zzhhan@bnu.edu.cn (Z. Han).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.neuroimage.2024.120520
Received 3 August 2023; Received in revised form 7 January 2024; Accepted 17 January 2024
Available online 18 January 2024

Beeck et al., 2019).

However, the precise manner in which learning experiences from top-down processing modulate category-specific regions in the VOTC has not been determined. Two hypotheses have been proposed. The first hypothesis suggests that different top-down nonvisual features were processed by neurons from distinct locations, which means stimuli with different nonvisual features will activate different areas in VOTC even when the visual features are controlled (Rauschecker et al., 2011; Saygin et al., 2011; Ekstrand et al., 2020). The second hypothesis proposes that the location of training-related areas will be formed mainly by visual features (Hasson et al., 2002; Malach et al., 2002; Nasr et al., 2014; Srihasam et al., 2014). Therefore, stimuli with different nonvisual but controlled visual features may be represented in the same area in VOTC. Additionally, for the second hypothesis, different nonvisual features may be more likely to influence this category-specific region on VOTC by modulating brain activation in different ways (e.g., the whole activation intensity and the pattern of activity across multiple voxels; Ishai et al., 1999; Haxby et al., 2001; Song et al., 2012; Carreiras et al., 2014; Coggan et al., 2016). The most significant difference between these two hypotheses lies in whether the activated areas of different category stimuli with the controlled visual features are situated in the same or different locations on the cortex after training. To our knowledge, no studies have directly differentiated between these two possibilities.

To explore this question, we employed meaningless novel figures

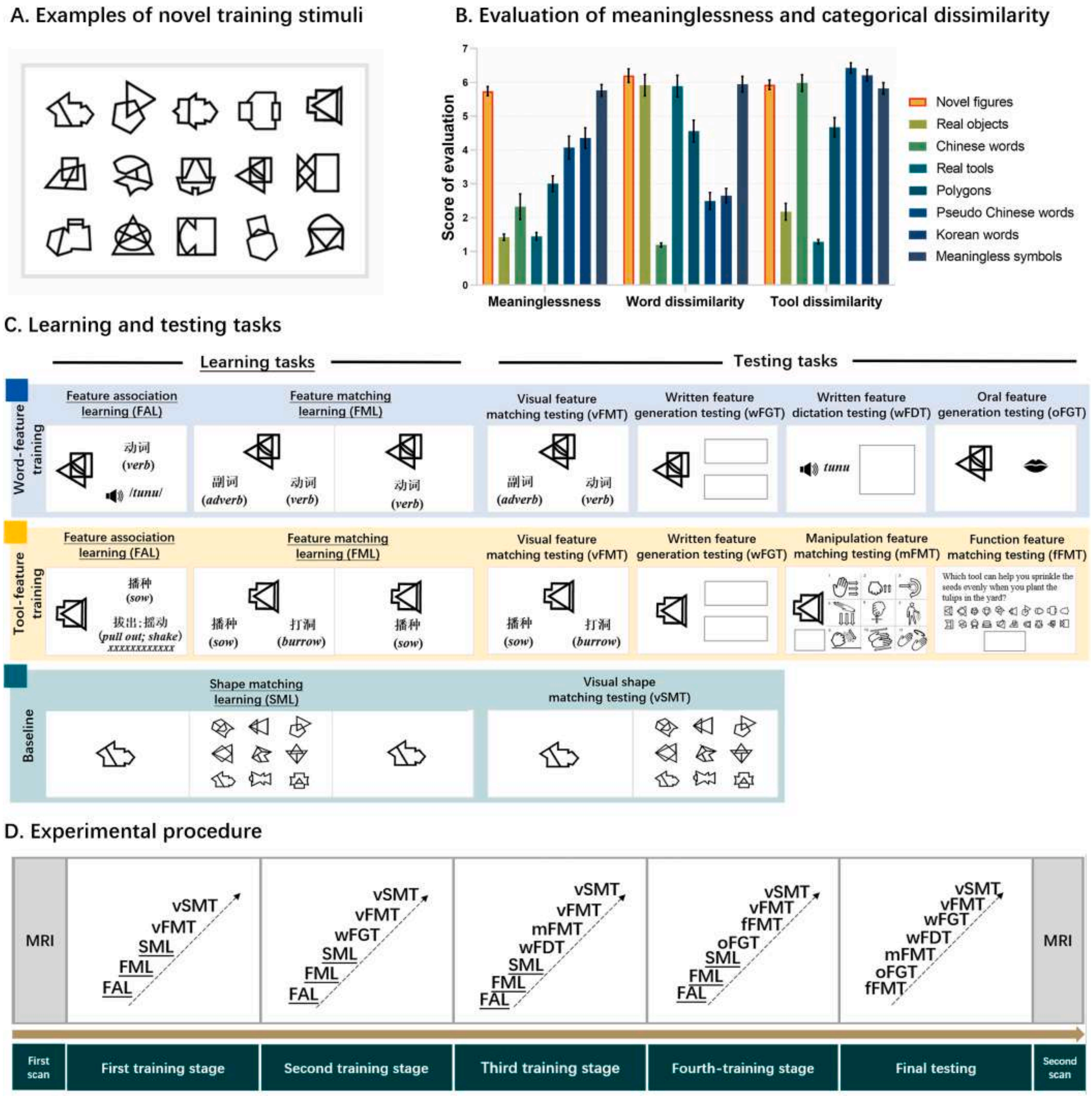
A. Examples of novel training stimuli
B. Evaluation of meaninglessness and categorical dissimilarity
Score of evaluation
0
1
2
3
4
5
6
7
Meaninglessness
Word dissimilarity
Tool dissimilarity
Novel figures
Real objects
Chinese words
Real tools
Polygons
Pseudo Chinese words
Korean words
Meaningless symbols
C. Learning and testing tasks
Learning tasks
Testing tasks
Word-feature training
Feature association learning (FAL)
动词
(verb)
/tunu/
Feature matching learning (FML)
副词
(adverb)
动词
(verb)
动词
(verb)
Visual feature matching testing (vFMT)
副词
(adverb)
动词
(verb)
Written feature generation testing (wFGT)
Written feature dictation testing (wFDT)
tunu
Oral feature generation testing (oFGT)
Tool-feature training
Feature association learning (FAL)
播种
(sow)
拔出；摇动
(pull out; shake)
XXXXXXXXXXXX
Feature matching learning (FML)
播种
(sow)
打洞
(burrow)
播种
(sow)
Visual feature matching testing (vFMT)
播种
(sow)
打洞
(burrow)
Written feature generation testing (wFGT)
Manipulation feature matching testing (mFMT)
Function feature matching testing (fFMT)
Which tool can help you sprinkle the seeds evenly when you plant the tulips in the yard?
Baseline
Shape matching learning (SML)
Visual shape matching testing (vSMT)
D. Experimental procedure
MRI
FAL
FML
SML
vFMT
vSMT
FAL
FML
SML
wFGT
vFMT
vSMT
FAL
FML
SML
wFDT
mFMT
vFMT
vSMT
FAL
FML
SML
oFGT
fFMT
vFMT
vSMT
fFMT
oFGT
mFMT
wFDT
wFGT
vFMT
vSMT
MRI
First scan
First training stage
Second training stage
Third training stage
Fourth-training stage
Final testing
Second scan

Fig. 1. Experimental stimuli and paradigm. The examples of stimuli highlighted in the red boxes in Figure B are those shown in Figure A. The evaluation scores in Figure B are shown as the mean values and standard error of the mean (SEM). The full names of the tasks in Figure D are provided in Figure C. The learning tasks in Figures C and D are underlined to distinguish them from the testing tasks.

with varied nonvisual and controlled visual features as different categories and tested whether learning these different nonvisual features could induce distinctive locations or regulate the activation intensities and patterns in the same region of VOTC. Nineteen healthy participants participated in approximately 13 training sessions, associating three visually homogeneous sets of novel meaningless figures (each set comprising 20 figures) with different nonvisual features, namely, pronunciation and grammatical class for the wordlike condition, manipulation and function for the non-wordlike condition, and no nonvisual features for the visual familiarity condition which served as the baseline. Subjects performed a one-back task for all stimuli in the pre- and post-training MRI scan sessions. The post-training scans also included a categorization task, requiring subjects to identify the training condition to which the stimuli belonged. Compared to the one-back task, this task requires deeper processing from the subjects. The fMRI data were analyzed to examine whether potential categorical differences (wordlike and non-wordlike) existed in the activation location, intensity, and pattern in the VOTC.

2. Materials and methods

2.1. Participants

Forty-nine healthy college students were recruited for this study. All the participants were right-handed native Mandarin Chinese speakers with normal or corrected-to-normal vision. Nineteen of them [age: 20.05 (*mean*) ± 0.27 (*standard error of the mean, SEM*) years old; 10 females] participated in the training and fMRI experiments. The remaining 30 subjects completed experiments to assess the meaninglessness and categorical bias of the novel figures. This study was approved by the Institutional Review Board of the National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University. Written informed consent was obtained from all participants before the experiments.

2.2. Stimuli

A total of 60 novel figures (450 * 450 pixels) were created with simple lines and arcs (Fig. 1A), which was similar to what was done in our recent study (Li et al., 2020). The stimuli were generated collaboratively by two Ph.D. students, each of whom was responsible for creating 30 stimuli. The instructions for this generation process were as follows: “Please utilize Photoshop software to craft a series of meaningless stimuli, each composed of 2–5 basic geometric shapes (e.g., lines, triangles, diamonds, ovals, etc.). These fundamental geometric shapes can be superimposed to create irregular geometries. It is crucial to ensure that the combinations of geometric shapes used in crafting these stimuli are as random as possible and should not resemble or be modeled after any real objects.

To ensure the meaninglessness and absence of association with real-world objects (e.g., words and tools) in our novel figures, we measured both the objective low-level visual dissimilarity and subjective functional dissimilarity compared to real-world objects. Using ShapeComp, a low-level vision model (Morgenstern et al., 2021), we assessed visual dissimilarity. The results showed that the novel stimuli were visually similar to each other, and the real-world object stimuli also exhibited visual similarity within their group, but the novel stimuli were visually dissimilar to the real-world object stimuli (see Supplementary Fig. 1). To measure functional dissimilarity, we instructed 30 healthy subjects to evaluate their meaninglessness and categorical dissimilarity. The evaluation was conducted using a 7-point grading scale (1 = very similar, 4 = medium, 7 = very dissimilar) and involved answering three questions: how unlikely was the figure to be (1) a meaningful object, (2) a real word, or (3) a real tool? To compare the evaluation scores of the novel figures with those of standard stimuli for each question, we added three other assessment materials: 10 real objects, 20 real Chinese words, and 20 real tools. Moreover, to avoid bias in the responses to the above stimuli, we added 40 filler stimuli (10 polygons, 10 pseudo-Chinese words, 10 Korean words, and 10 meaningless symbols) (Fig. 1B). All 150 stimuli were presented in a pseudorandom manner during the assessment. The evaluation scores for the meaninglessness question for the 60 novel figures (5.74 ± 0.13) were close to those for the meaningless symbols (5.76 ± 0.17; $t_{29} = -0.17$, $p = 0.87$) but were significantly greater than those for the other six types of stimuli (score range: 1.42 to 4.35; *ps* < 0.001). Moreover, the scores for the word-likelihood questions of the novel figures (6.20 ± 0.20) were close to those of meaningless symbols (5.95 ± 0.23; $t_{29} = 1.65$, $p = 0.11$), real objects (5.92 ± 0.32; $t_{29} = 1.15$, $p = 0.26$), and real tools (5.89 ± 0.32; $t_{29} = 1.23$, $p = 0.23$) but were significantly greater than those of the other four types of stimuli (score range: 1.19 to 4.56; *ps* < 0.001). Similarly, the scores for the tool-likelihood questions of the novel figures (5.93 ± 0.14) were close to those of meaningless symbols (5.82 ± 0.17; $t_{29} = 0.87$, $p = 0.39$), Chinese words (5.98 ± 0.24; $t_{29} = -0.19$, $p = 0.85$), and Korean words (6.21 ± 0.17; $t_{29} = -1.39$, $p = 0.18$) and were significantly lower than those of the pseudo-Chinese words (6.43 ± 0.15; $t_{29} = -2.48$, $p = 0.02$). However, these values were significantly greater than those for the other three types of stimuli (score range: 1.29 to 4.67; *ps* < 0.001) (Fig. 1B). All the results above indicate that the novel figures were visually and functionally meaningless and sufficiently dissimilar to real words or tools before training.

2.3. Behavioral training procedure

We designed three training conditions with varying nonvisual features and controlled visual features: wordlike, non-wordlike, and visual familiarity conditions. First, each figure in the wordlike condition was associated with two kinds of word features (i.e., pronunciation and grammatical class). It is particularly important to emphasize that the wordlike condition is only partially but not entirely related to real words because of its relatively low ecological validity. Therefore, we use the term "wordlike" stimuli to distinguish them from words acquired in real natural situations. Second, to differentiate from the wordlike condition, we chose two nonword features, namely, manipulation and function, for the non-wordlike condition. This selection was influenced by the prototype of the tools, with the aim of establishing associations between meaningless shapes and operational characteristics such as manipulation and resulting functions. Consequently, many subsequent training tasks were centered around tools. However, the lack of actual manipulation in these tasks poses a challenge in forming a genuine tool-related representation. Nonetheless, we believe that this condition is distinct from the wordlike condition and can be characterized as non-wordlike. Third, to obtain the training-related category-specific location, we designated a visual familiarity condition as a baseline (which was subsequently described as the baseline condition for simplicity) for which a procedure similar to that of the wordlike or non-wordlike condition was used, except that the stimuli were not learned with specific features (Fig. 1C).

The 60 novel stimuli were randomly divided into three fixed sets, each containing 20 figures. The assignment of stimuli sets to training conditions followed a Latin-counterbalanced design across subjects. Specifically, seven subjects considered the first stimulus set as the wordlike condition, the second as the non-wordlike condition, and the third as the baseline condition. The other six subjects regarded the second set as the wordlike condition, the third as the non-wordlike condition, and the first as the baseline condition. Additionally, another six subjects were trained to associate the third set with the wordlike condition, the first with the non-wordlike condition, and the second with the baseline condition. By employing meaningless novel stimuli and implementing a counterbalanced design, the study ensured stimulus homogeneity across the different training conditions. Each subject was presented with all the stimuli during the behavioral training and fMRI experiments. The detailed design and procedure are explained as follows.

Learning and testing tasks of wordlike and non-wordlike conditions. When a subject learned the wordlike identity of a figure, the figure was specifically associated with a parameter feature space of two linguistic features: one of 20 pronunciations (e.g.,*/tunu/*,*/dufi/*,*/tepi/*) and one of four grammatical classes (i.e., adjective, adverb, preposition, auxiliary). To avoid pronunciations that were close to those of real Chinese words, we created these pronunciations by combining 10 consonants and 5 vowels of Esperanto to generate bisyllabic pseudowords using "espeakers" speech synthesizer software (http://espeak.sourceforge.net/). Similarly, when the nonword identity of a figure was known, the figure was specifically associated with another parameter feature space of two nonlinguistic features: one of 20 functional properties (e.g., burrow, sow, irrigate) and one of 20 ordered manipulations in which each consisted of two main correlated actions (e.g., press and push, pull out and shake, beat and stir). There were 16 actions in total. Each action was drawn on a card that showed a clear diagram of how the action was performed. These 16 action cards were used in the testing tasks.

To train subjects to learn how to categorize figures as wordlike or non-wordlike, we used two feature learning tasks (i.e., association and matching). In *feature association learning (FAL)*, subjects were instructed to respond to each figure with its two corresponding features as accurately as possible. Each figure was visually presented on the screen. For the wordlike identity, the written name of the grammatical category (i.e., adjective) was visually presented, and the auditory pronunciation was presented when the subject clicked a trumpet logo on the screen. For the non-wordlike identity, we visually presented the written names of the functional properties (e.g., sow) and two manipulation actions (e.g., pull out and shake). To help subjects better understand the features of manipulation actions, we additionally presented a written description of the manipulation process (e.g., pulling out the switch of the stimulus when the seeds are placed on the bottom of the stimulus and then shaking the stimulus clockwise so that the seeds can fall evenly) on the screen. The *feature matching learning (FML)* task involved the presentation of a figure with two feature options. Subjects were required to judge which feature matched the figure based on the above association task, and there was no time limit for the subjects' responses. The correct answer was provided as feedback following the subject's response.

We developed the following six testing tasks to evaluate the training effects comprehensively. 1) The *visual feature matching testing* (*vFMT*) task was identical to the *FML* task, except no feedback providing the correct answer was given. 2) The *written feature generation testing* (*wFGT*) task requires the subject to write down two learned features for each visually presented figure. The pronunciation of each wordlike stimulus was written in *pinyin* form. The above two testing tasks were applied to both wordlike and non-wordlike conditions. We also developed two additional tasks for each wordlike and non-wordlike condition. 3) In *written feature dictation testing (wFDT)* for the wordlike condition, subjects were instructed to write/draw figures when listening to their pronunciation feature via earphones. 4) In *oral feature generation testing (oFGT)* for the wordlike condition, subjects were instructed to verbally report the pronunciation and grammatical class for each visually presented figure. 5) *Manipulation feature matching testing (mFMT)* for the non-wordlike condition required the subjects to select two action cards (e.g., pull out, shake) from the 16 action cards that had manipulation actions corresponding to the visually presented figure. 6) In *function feature matching testing (fFMT)* for the non-wordlike condition, we asked subjects to select the figure whose function was most appropriate for a given situational prompt (e.g., Which figure can help you sprinkle the seeds evenly when you plant the tulips in the yard?). There were no time limits for the response in any of the testing tasks.

Learning and testing tasks of the visual familiarity (baseline) condition. The training task was *shape matching learning (SML)*. A target figure was first presented on the screen for a short time (i.e., a duration ranging from 66.8 ms to 167 ms), after which nine candidate figures were visually presented. The subjects were required to select the target figure among the candidates. The correct answer was provided following each response of the subjects. The *visual shape matching test* (*vSMT*) was used for testing. The *vSMT* was identical to the learning task, except no feedback was provided for the correct answer. There were no time limits for the response in any of the testing tasks.

Training procedure. Each subject was trained in four successive stages (Fig. 1D). All the stages involved completed the same training tasks. The visual familiarity testing task was the same during all stages, but the wordlike and non-wordlike testing tasks varied across stages. Generally, the difficulty of the testing tasks for wordlike and non-wordlike conditions increased as the stages progressed. Each stage involved 2–5 training sessions lasting approximately 1 h per day. On the first day of each training stage, the subjects completed all the learning tasks and subsequently completed specific testing tasks. The learning tasks lasted 40–50 min, followed by a period of testing tasks lasting approximately 10–20 min. During the learning task portion of the session, the subjects spent time on each of the two learning tasks according to their preferences. During the testing portion of the session, more challenging tasks were presented and completed first, followed by easier tasks. No time limit was given for the response to the training tasks, and subjects were encouraged to perform the tasks as precisely as possible. When a subject had more than 80 % accuracy in each testing task on a day, he or she progressed to the next stage of training. Each subject completed 12.74 training sessions on average (*SEM* = 0.59 sessions, range: 9 to 18 sessions). To examine whether the participants had acquired the different categories of all the figures, we designed and implemented a final testing session in which each subject completed all the testing tasks the day before the post-training MRI scan.

2.4. Neuroimaging data acquisition

The 19 individuals who participated in the behavioral training underwent fMRI scanning twice (i.e., before and after training) using a 3T Siemens Magnetom Prisma scanner with a standard 64-channel phased-array head coil at Beijing Normal University. We collected four types of images: task-state fMRI images, 3D T1-weighted images, diffusion-weighted images (DWI), and resting-state images. Task-state fMRI images were acquired for two tasks (i.e., the one-back and categorization tasks) with the 60 novel figures. These tasks involved varying levels of information processing depth for the figures. The one-back task could be completed based on the shape information of the figures regardless of whether the categories of these figures were learned. In contrast, the categorization task was completed based on learned category information after training. As a result, fMRI images for the one-back task were collected at each scanning session, with the categorization task images exclusively obtained during the post-training scans. The T1-weighted images were gathered in pre- and post-training sessions, and DWI images and resting-state were collected before the training.

MRI acquisition parameters. Structural 3D images in the sagittal plane were obtained with the following parameters: repetition time (TR) = 2530 ms, echo time (TE) = 2.27 ms, flip angle (FA) = 7°, field of view (FOV) = 256 × 256 mm^2, slice number = 208, slice thickness = 1.0 mm, and voxel size = 1.0 × 1.0 × 1.0 mm^3. The task fMRI data were collected in the transverse plane with a T2*-weighted echo-planar imaging (EPI) sequence with the following parameters: TR = 2000 ms, TE = 34 ms, FA = 90°, FOV = 200 × 200 mm^2, slice number = 72, slice thickness = 2 mm, and voxel size = 2.0 × 2.0 × 2.0 mm^3. DWI was acquired in the transverse plane with the following parameters: TR = 3900 ms, TE = 65 ms, FA = 90°, FOV = 256 × 256 mm^2, slice thickness = 2.0 mm, and voxel size = 2.0 × 2.0 × 2.0 mm^3. There were a total of 64 diffusion weighting directions with a b value of 2000 s/mm^2 and 10 b_0 images. The resting-state images were acquired in the transverse plane with the following parameters: TR = 1000 ms, TE = 30 ms, FA = 70°, FOV = 192 × 192 mm^2, slice thickness = 3.0 mm, and voxel size = 3.0 × 3.0 × 3.0 mm^3. DWI and resting-state data were not further analyzed in the current study.

One-back fMRI task. This task was a one-back repetition detection task with a block design. In addition to the three training conditions (wordlike, non-wordlike, and baseline), we added two real, meaningful conditions (20 real words, 20 real tools) to avoid response bias in the subjects. Each condition contained 20 black and white line figures. The whole experiment contained four runs, each consisting of 20 blocks (i.e., four blocks per condition). Blocks in a run were presented in a pseudorandom order and were separated by a 6 s fixation window. Each block included 8 stimuli from the same condition, and each stimulus was visually presented at 800 ms, followed by a blank screen for 1200 ms. Each run began with a 10 s fixation and ended with a 2 s fixation to stabilize the signal. The subjects were instructed to respond by pressing a button with their right index finger whenever the presented stimulus matched the preceding one.

Categorization fMRI task. This task was a categorization task using an event-related (ER) design. It included 6 runs. The stimuli in each run consisted of the 60 novel figures that the subjects had learned. The figures were visually presented in a pseudorandom order and were the same across subjects. Each figure was presented for 1 s, followed by a random interval with a fixation duration ranging from 1 s to 9 s. Participants were instructed to identify the category (wordlike, non-wordlike, or baseline) to which each figure was associated. They responded by pressing the button corresponding to the chosen category, and three types of buttons were counterbalanced across subjects. The example of the experiment instruction was as follows: "Please determine the category to which the following figure belongs: if it is a word, press the key with the right index finger; if it is a tool, press the key with the right middle finger; if it does not possess any attributes, press the key with the right ring finger." The sequence of the three conditions in each run and the interval between trials were optimized using the optseq2 algorithm (https://surfer.nmr.mgh.harvard.edu/optseq/). Each run began and ended with a 10 s fixation to stabilize the signal.

The 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 6-mm full width at half-maximum (FWHM) Gaussian kernel. For each subject, the data were first high-pass filtered (the parameter was 0.0078 Hz for both categorization and one-back fMRI tasks) and then fitted by a general linear model (GLM) with a boxcar regressor with a duration matching the response time to estimate the effect of the experimental conditions. We excluded data from a total of two runs from our analysis due to excessive head motion (> 2 mm) or rotation ($> 2^{\circ}$) of the subject. Moreover, because gradient-echo sequences could lead to signal loss in the inferior temporal cortex (Ojemann et al., 1997), each subsequent analysis excluded voxels experiencing signal loss in all three conditions (wordlike, non-wordlike, and baseline).

Compared to the one-back task, the categorization task in the post-training scan required a deeper level of learned information about the categories and might induce stronger category-related training effects in the VOTC. Therefore, the following analyses first investigated the effects of the categorization task and ultimately yielded regions of interest (ROIs) from the group-averaged activation map and peaks of interest (POIs) from the individual activation map. Then, the ROI and POI were used to investigate the effects of the one-back task at two distinct time points (i.e., pre- and post-training). The analyses of the post-training data were used to inspect whether the training effects observed in the categorization task were also evident in the one-back task and were consistent across tasks. Moreover, analyses of the pre-training data were used to determine the original state before training.

2.5. Category-related activation effects in the VOTC

To explore whether learning the features of the two categories (wordlike, non-wordlike) gave rise to different activation effects in the VOTC, we compared the location, intensity, and pattern of activation between the categories in the VOTC using fMRI task data.

Activation location. Activation location refers to the location of significant brain clusters specific to a particular experimental condition. To compare the categorical differences in the location of activation in the VOTC, we first created a VOTC mask that included all the voxels in the inferior occipital gyrus, fusiform gyrus, parahippocampal gyrus, and inferior temporal cortex of the left hemisphere (y-axis ranging from –90 to –30) defined by the IBASPM 116 atlas from the WFU_PickAtlas toolbox (Maldjian et al., 2003). We chose the left hemisphere because we primarily focus on the differences between word and nonword nonvisual features, and word-specific activation is almost strongly located in the left hemisphere due to language dominance (Knecht et al., 2000; Seghier and Price, 2011). Then, we performed two complementary analyses. First, a group-level analysis was performed to examine the degree of overlap between the locations of the wordlike- and non-wordlike-relevant activation clusters in the VOTC mask. The clusters were extracted as the fMRI signals (i.e., the parameter estimate beta values) from the contrast of the *wordlike condition (or non-wordlike condition)* vs. *baseline condition* across subjects (*GRF* corrected, voxel-level $p < 0.001$, cluster-level $p < 0.025$, one-tailed). We then further extracted the common region between the two categories ($\text{ROI}_{\text{common}}$), i.e., the mask of the overlapping voxels between the two clusters. The overlapping degree was measured as the percentage of voxels in the $\text{ROI}_{\text{common}}$ relative to the voxel number in the smaller of two category-specific clusters. A second individual-level analysis was performed to test the degree of closeness of the activation peaks between the two categories in the VOTC mask. Specifically, we extracted the peak in the VOTC mask for each category and each subject (i.e., POI) and then performed a comparison across subjects of all the coordinates of the peaks between the two categories (related samples Wilcoxon signed-ranks test, $p < 0.05$). As we later found that the coordinates of the peaks between the two categories were not significant on any of the three axes, the common peak (i.e., $\text{POI}_{\text{common}}$) for each subject was extracted as an 8 mm-radius sphere centered at the peak of the contrast *wordlike condition* + *non-wordlike condition* > *baseline condition.* Importantly, $\text{ROI}_{\text{common}}$ is a group-level defined region, while $\text{POI}_{\text{common}}$ is an individual-level defined region. The inclusion of both definition methods in the present study emphasizes their complementarity rather than differences in obtaining significant results.

Activation intensity. Activation intensity refers to the average beta values derived from the BOLD signals in the regions of interest within the brain. We separately conducted two analyses to compare the activation intensity among the three conditions (wordlike, non-wordlike, baseline) in the common region (i.e., $\text{ROI}_{\text{common}}$) and the common peak (i.e., $\text{POI}_{\text{common}}$) in the VOTC obtained via the above analysis.

Notably, in subsequent analyses, we used the same procedure for analyzing $\text{ROI}_{\text{common}}$ and $\text{POI}_{\text{common}}$, except that the clusters of interest had different shapes and locations. For simplicity, we only introduce the analysis details for the $\text{ROI}_{\text{common}}$ unless otherwise noted. Furthermore, although the main purpose of the following analyses was to reveal the differences between the two categorical conditions (wordlike and non-wordlike), the differences between the two categorical conditions and the baseline condition were also investigated, as these comparisons shed light on the degree of change in neural signals specifically prompted by categorical training.

We extracted each subject's activation intensity (beta value) in each of the three conditions (each condition vs. rest) for each voxel in the $\text{ROI}_{\text{common}}$. Then, the activity intensity in each condition was averaged across voxels for each subject. Finally, the average intensity between each pair of the three conditions (3 pairs: wordlike vs. non-wordlike, wordlike vs. baseline, and non-wordlike vs. baseline) was compared across subjects using a paired sample *t* test ($p < 0.05$).

Activation pattern. The activation pattern denotes the patterns of neural activity across multiple voxels within the regions of interest in the brain. Multivariate pattern analysis (MVPA) was adopted to compare the similarity of the activation patterns in the ROI_{common} and POI_{common} among the three conditions. To ensure reliable results, we adopted the following three MVPA subanalyses.

Split-half correlation analysis in the ROI_{common}. This analysis was used to examine the similarity of neural patterns in the ROI_{common} between the three training conditions, focusing on the stability of activation patterns across runs within a condition (Haxby et al., 2001; Kang et al., 2021). The data were split in half (i.e., half1 and half2). The hypothesis was that if the activation patterns between the two conditions were significantly different, the activation patterns produced by the two data halves from the same condition (i.e., "within-condition") would exhibit greater stability than the activation patterns produced by half1 from one condition and half2 from a different condition (i.e., "between-condition"). In other words, there was a greater correlation for the within-condition comparison than for the between-condition comparison.

This analysis included the following main steps. 1) For each subject, the preprocessed normalized fMRI data in the ROI_{common} were divided into halves. This process was repeated to form all possible unique divisions. For example, the six runs in the categorization task were split into two halves (each containing three runs) and repeated to form the ten possible divisions (e.g., way1: runs1, 2, and 3 vs. runs 4, 5, and 6; way 2: runs1, 2, and 4 vs. runs 3, 5, and 6). 2) The beta value of each voxel in each of the three conditions was extracted for each half of the data for a given division. The mean beta value across all three conditions within the ROI_{common} was subtracted from the beta value for one condition in each voxel to obtain a response more specific to that category condition. As a result, there were six category-specific beta values for each voxel and each subject. Each half of the data yielded three values corresponding to the three conditions (wordlike, non-wordlike, baseline). 3) For each pair of conditions, the category-specific beta values of two conditions from one-half of the data were correlated with the category-specific beta values of the other half across the voxels in the ROI_{common}. This yielded four correlation coefficients: two within-condition coefficients (correlations between the two halves of the data from the same condition) and two between-condition coefficients (correlations between one-half of the data from one condition and the other half from a different condition). 4) The four coefficients were further Fisher *z*-transformed across subjects to solve the skewed distribution of correlation coefficients. 5) The difference (D value) between the two conditions was calculated by subtracting the average value of two *z*-transformed within-condition coefficients from the average value of two *z*-transformed between-condition correlation coefficients. 6) The above steps were completed for all possible divisions, and the D values of all the divisions were averaged. Finally, the average D values, after inserting 2.5 standard deviations (SD) as the outlier threshold, were statistically tested (compared with zero) across subjects (one-sample *t* test, $p < 0.05$) to investigate whether significant differences emerged in the activation patterns between each pair of conditions. To increase the reliability of our findings, we ran the analysis with no category-specific beta step (i.e., step 2) or Fisher z transfer step (i.e., step 4).

Leave-one-out support vector machine (SVM) analysis of the ROI_{common}. This analysis was also used to investigate the neural pattern similarity in the ROI_{common} between the three training conditions but focused on the classification accuracy of activation patterns between conditions. This analysis was conducted with the Decoding Toolbox (Hebart et al., 2015) based on a library for SVM (Chang and Lin, 2011). The procedure included the following steps. 1) The beta values from unsmoothed beta maps of each voxel within the ROI_{common} for each condition were extracted. 2) The SVM classifier was trained to establish a predictive model using the beta values of a pair of conditions from 5 out of 6 runs. 3) The model's predictive ability (correct vs. incorrect) was examined using the beta values of the same two conditions on the left-out run. This step was repeated six times, each time leaving out a different run (leave-one-out procedure). 4) The average classification accuracy was calculated across repetitions. Finally, the group-level decoding accuracy across subjects was assessed using a one-sample *t* test ($p < 0.05$) with the removal of outliers whose values exceeded 2.5 SD.

Leave-one-out SVM searchlight analysis across the entire VOTC. This analysis was used to explore whether the different activation patterns between the three conditions also appeared in areas of the VOTC other than the ROI_{common}, ultimately investigating the uniqueness of the ROI_{common} in discriminating the categorical patterns. This analysis was implemented in a searchlight manner for the data from the categorization task. First, for each subject, we created a 6 mm-radius sphere centered around every voxel throughout the VOTC in the native space. Second, a leave-one-out SVM analysis was performed in each sphere instead of in the $ROI_{common,}$ and the classification accuracy between each pair of conditions in each voxel of the VOTC was obtained. This analysis yielded a map of classification accuracy for each pair of conditions in the VOTC. Third, the map of each subject was spatially normalized to the MNI space and smoothed using a Gaussian kernel (6 mm FWHM). Finally, we assessed the significant clusters whose classification accuracy was above chance (50 %) across subjects (one-sample *t test, GRF* corrected, voxel-level $p < 0.001$, cluster-level $p < 0.025$, one-tailed).

3. Results

3.1. Behavioral performance

The mean accuracies of 19 subjects for each testing task in the final test reached the ceiling for both the wordlike condition (*vFMT*: 0.99 ± 0.003; *wFGT*: 0.99 ± 0.005; *wFDT*: 0.96 ± 0.01; *oFGT*: 1.00 ± 0.003) and the non-wordlike condition (*vFMT*: 0.99 ± 0.004; *wFGT*: 0.99 ± 0.005; *mFMT*: 0.95 ± 0.02; *fFMT*, 0.99 ± 0.01). In addition, the participants could also effectively recognize the figures in the baseline condition during the visual shape matching test (*vSMT*) (accuracy=0.92 ± 0.02; chance level=0.10) (Fig. 2A). These results showed that the participants could successfully associate the novel stimuli with their corresponding nonvisual categorical features of words and nonwords and be very familiar with the visual shape of the figures in the baseline condition.

To assess the potential differences in familiarity level between the two categorical training conditions, we conducted a comparative analysis of behavioral performance across two commonly used testing tasks for the two categories (i.e., *vFMT* and *wFGT*) in the final training examination. The results showed that the behavior performances were not significant between the two categories in the two testing tasks (*vFMT: Z* = –0.11, *wFGT: Z* = –0.58; *Wilcoxon* signed rank test *ps* > 0.05). We further included a categorization task during post-training fMRI scanning to investigate participants' responses to the two training categories (Fig. 2B). No statistically significant difference was found in the inverse efficiency (IE) score (i.e., reaction time/accuracy) between the wordlike and non-wordlike stimuli (wordlike vs. non-wordlike: 996.13 ± 40.36 vs. 981.21 ± 43.14, $t_{15} = 1.35$, $p = 0.20$). These results revealed that participant's familiarity with the wordlike and non-wordlike conditions did not significantly differ.

3.2. Category-related effects in the VOTC

Activation location. To explore whether learning the features of wordlike and non-wordlike categories resulted in different category-specific locations within the VOTC, we conducted two complementary analyses. First, at the group level, we examined the degree of overlap between two regions of interest (ROIs) in the VOTC: the wordlike and non-wordlike activation clusters derived from the group-averaged

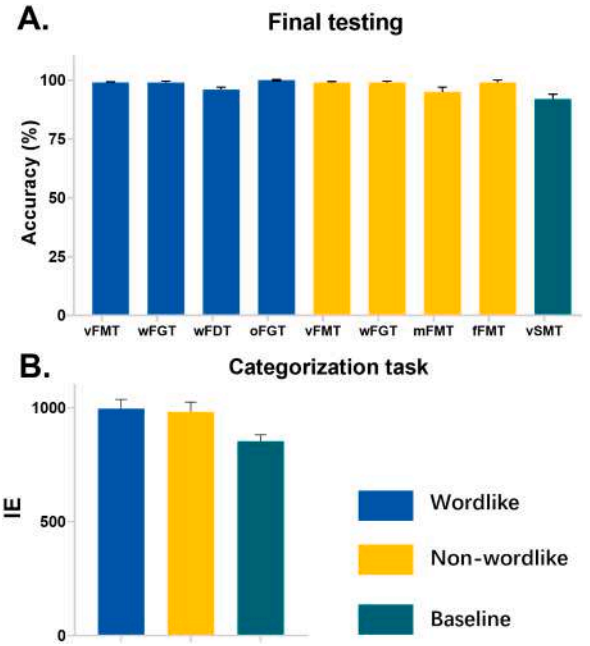

Fig. 2. The average behavioral performance for the three training conditions in the final examination and fMRI scan. (A) The wordlike (blue) and non-wordlike (yellow) conditions both had four different testing tasks, and the baseline condition (green) had one testing task. The y-axis represents the accuracy (%) of the testing task in the final examination. The full names of the tasks in the training stages are given in Fig. 1C. (B) Performance on the categorization task during post-training scanning. The y-axis represents the inverse efficiency score (IE) as reaction time (RT) divided by the accuracy (ACC) in the categorization task. Error bars represent the standard errors of the mean (SEMs).

activation map. The wordlike cluster (wordlike condition vs. baseline condition) and the non-wordlike cluster (non-wordlike condition vs. baseline condition) contained 217 voxels and 51 voxels within the VOTC mask, respectively (*GRF* corrected, voxel-level $p < 0.001$; cluster-level $p < 0.025$, one-tailed). Notably, a common region of interest (i.e., $\text{ROI}_{\text{common}}$), consisting of 46 voxels, was identified. Therefore, the overlap between $\text{ROI}_{\text{common}}$ and non-wordlike regions was as high as 90.20 % (46/51) (Fig. 3A). Second, at the individual level, we extracted the activation peak of interest (POI) within the VOTC for both the wordlike (wordlike condition vs. baseline condition) and non-wordlike (non-wordlike condition vs. baseline condition) conditions from each subject and examined potential differences. The results revealed no differences in the coordinates of each of the three axes of the peak between the two categories (*x-axis*: –51.41 ± 1.33 vs. –49.65 ± 1.90; *y-axis*: –52.59 ± 1.11 vs. –50.47 ± 1.68; *z-axis*: –13.76 ± 1.42 vs. –14.94 ± 1.59; Wilcoxon signed-rank test, *ps* > 0.10) (Fig. 3B). These results indicate that wordlike and non-wordlike training activated highly overlapping locations in the VOTC. We also performed the same overlapping analysis of the whole brain and found that wordlike and non-wordlike patterns also overlapped in the left inferior temporal cortex, insula, inferior parietal cortex, inferior frontal cortex, supplementary motor area, and precentral cortex (Supplementary Fig. 2 and Supplementary Table 1).

Activation intensity. We examined whether there were differences in activation intensity between the wordlike and non-wordlike conditions within the same location of the VOTC (i.e., $\text{ROI}_{\text{common}}$ and $\text{POI}_{\text{common}}$). While the primary focus of the analyses was to examine differences between the wordlike and non-wordlike categorical conditions, we also explored the contrasts between these conditions and the baseline condition to measure the extent of neural signal changes induced by categorical training. These analyses were applied to both the one-back and categorization tasks. The results of the activation intensity analysis are illustrated in Fig. 4.

For the categorization task, the mean activation intensities (beta values) of the $\text{ROI}_{\text{common}}$ in the wordlike and non-wordlike conditions were significantly greater than those in the baseline condition (wordlike vs. baseline: 3.16 ± 0.50 vs. 2.12 ± 0.49; $t_{18} = 6.63$, *FDR*-corrected, $p < 0.001$; non-wordlike vs. baseline: 2.94 ± 0.48 vs. 2.12 ± 0.49; $t_{18} = 5.02$, *FDR*-corrected, $p < 0.001$). There was a tendency for the intensities in the wordlike condition to be greater than those in the non-wordlike condition, but the difference was not significant ($t_{18} = 1.35$, $p = 0.19$). Furthermore, the intensities of the $\text{POI}_{\text{common}}$ in the wordlike and non-wordlike conditions were significantly greater than those in the baseline condition (wordlike vs. baseline: 2.59 ± 0.40 vs. 1.11 ± 0.32; $t_{18} = 7.79$, *FDR*-corrected, $p < 0.001$; non-wordlike vs. baseline: 2.32 ± 0.38 vs. 1.11 ± 0.32; $t_{18} = 7.45$, *FDR*-corrected, $p < 0.001$). Most importantly, the intensities in the wordlike condition were significantly greater than those in the non-wordlike condition ($t_{18} = 2.43$, *FDR*-corrected, $p = 0.03$) (Fig. 4). These results demonstrated that wordlike and non-wordlike training induced different activity intensities at the same location in the VOTC.

For the one-back task during the post-training scan, the activation intensities of the $\text{ROI}_{\text{common}}$ in the wordlike and non-wordlike conditions were marginally significantly greater than those in the baseline condition (wordlike vs. baseline: 0.45 ± 0.15 vs. 0.29 ± 0.16; $t_{18} = 2.55$, *FDR*-corrected, $p = 0.02$; non-wordlike vs. baseline: 0.43 ± 0.16 vs. 0.29 ± 0.16; $t_{18} = 1.96$, uncorrected, $p = 0.07$). There were no significant differences in the activation intensities between the wordlike and non-wordlike conditions ($t_{18} = 0.34$, $p = 0.74$). Similarly, the activation intensity values of the $\text{POI}_{\text{common}}$ in the wordlike and non-wordlike

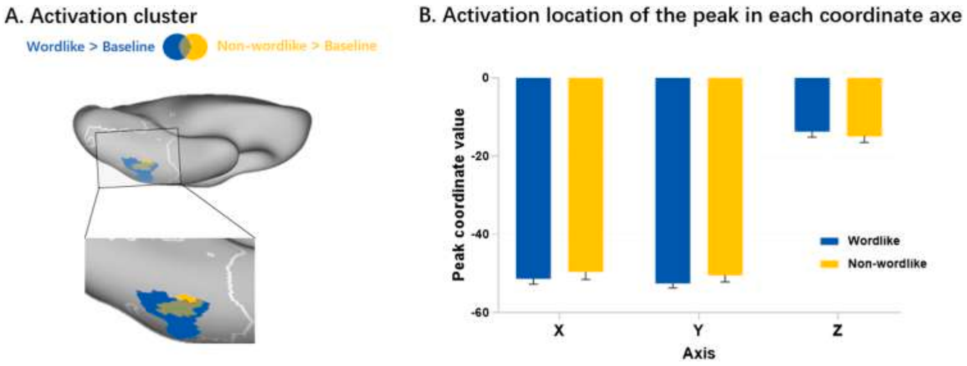

Fig. 3. Activation location of the two trained categories within the VOTC during the categorization task. (A) At the group level, the activation locations of the wordlike (blue) and non-wordlike (yellow) conditions are shown alongside their overlap regions (green). The gray boundary line marks the VOTC region, encompassing voxels in the inferior occipital gyrus, fusiform gyrus, parahippocampal gyrus, and inferior temporal cortex of the left hemisphere (axis y ranged from -90 to -30) defined by WFU_PickAtlas's IBASPM 116 atlas. (B) A comparison of individual peak voxel coordinates along the X, Y, and Z axes between the wordlike (blue) and non-wordlike (yellow) conditions.

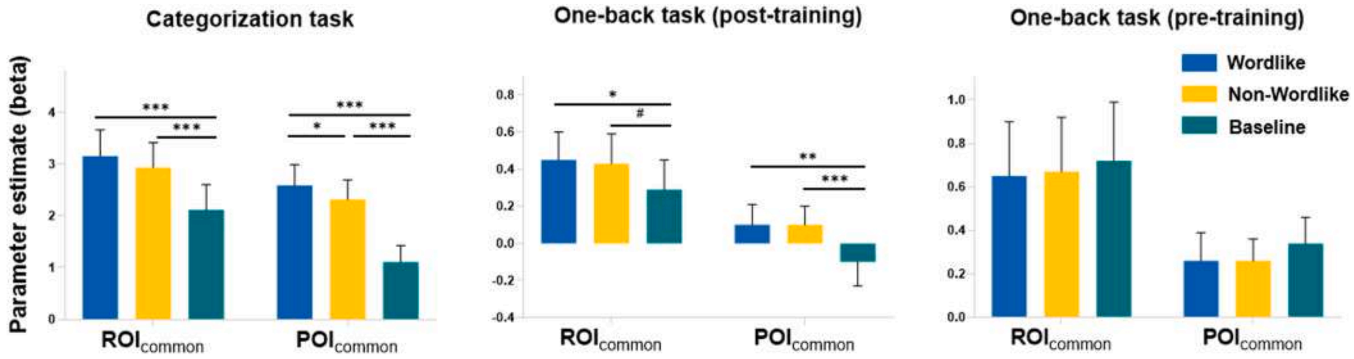

Fig. 4. Activation intensities in the ROI$_{common}$ and POI$_{common}$ of three training conditions among different tasks (i.e., the categorization task, post-training one-back task, and pre-training one-back task). ROI$_{common}$ and POI$_{common}$ refer to the common activation locations for wordlike and non-wordlike conditions at the group and individual levels, respectively. Error bars represent the standard errors of the mean (SEMs). POI = peak of interest, ROI = region of interest. #: $p \leq 0.10$; *: *FDR* corrected, $p \leq 0.05$; **: *FDR* corrected, $p \leq 0.01$; ***: *FDR* corrected, $p \leq 0.005$.

conditions were both significantly greater than those in the baseline condition (wordlike vs. baseline: 0.10 ± 0.11 vs. –0.10 ± 0.13; t_{18} = 2.74, *FDR*-corrected, $p = 0.01$; non-wordlike vs. baseline: 0.10 ± 0.10 vs. –0.10 ± 0.13; t_{18} = 3.25, *FDR*-corrected, $p = 0.004$). No significant differences were observed between the wordlike and non-wordlike conditions (t_{18} = 0.01, $p = 0.99$). The lack of significant differences in categorical comparisons may be attributed to the low cognitive demand in this task.

For the one-back task during the pre-training scan, the intensities between the three conditions showed no differences regardless of whether the ROI$_{common}$ (0.65 ± 0.25 vs. 0.67 ± 0.25 vs. 0.72 ± 0.27; *ps* > 0.30) or POI$_{common}$ (0.26 ± 0.13 vs. 0.26 ± 0.10 vs. 0.34 ± 0.12; *ps* > 0.30) was considered. These results suggest that the activation intensities of the three conditions were comparable before the categorical features were learned.

The above results demonstrated that the two categorical conditions had activation intensities comparable to the baseline condition before the subjects were trained to associate the stimuli. However, training in these two categories generated stronger activation compared to training at the baseline. Importantly, word training induced stronger activation than nonword training. In other words, learning each of the two categorical features activated the same location but at different intensities within the VOTC.

Activation pattern. We utilized three distinct multivariate pattern analyses (MVPAs) to examine differences in brain activation patterns between the wordlike and non-wordlike conditions in the VOTC. These analyses included split-half correlation analysis in the ROI$_{common}$ and POI$_{common}$ to assess the stability of activation patterns, leave-one-out SVM analysis in the ROI$_{common}$ and POI$_{common}$ to determine classification accuracy between conditions, and leave-one-out SVM searchlight analysis across the entire VOTC to identify regions capable of distinguishing different conditions based on activation patterns. These analyses were applied to both the one-back and categorization tasks.

For the categorization task, the split-half correlation analysis revealed that the difference (D value) in the activation pattern in the ROI$_{common}$ between each pair of conditions was significant (wordlike vs. non-wordlike: 0.12 ± 0.04, t_{18} = 2.74, *FDR*-corrected, $p = 0.01$; wordlike vs. baseline: 0.36 ± 0.06, t_{18} = 6.30, *FDR*-corrected, $p < 0.001$; non-wordlike vs. baseline: 0.31 ± 0.06, t_{18} = 5.08, *FDR*-corrected, $p < 0.001$). Similar significant results were also observed for POI$_{common}$ (wordlike vs. baseline: 0.43 ± 0.08, t_{17} = 5.17, *FDR*-corrected, $p < 0.001$; non-wordlike vs. baseline: 0.29 ± 0.06, t_{18} = 5.19, *FDR*-corrected, $p < 0.001$), except for the comparison of the activation patterns between wordlike and non-wordlike stimuli (0.02 ± 0.05, t_{17} = 0.53, $p = 0.60$). The results with no category-specific beta step and no Fisher z step also showed a similar pattern of results (Supplementary Fig. 3). Moreover, the leave-one-out SVM analysis also showed that each pair of conditions had significantly different classification accuracies for the activation patterns in the ROI$_{common}$ (wordlike vs. non-wordlike: 7.28 ± 3.24, t_{18} = 2.25, *FDR*-corrected, $p = 0.04$; wordlike vs. baseline: 16.58 ± 3.68, t_{18} = 4.50, *FDR*-corrected, $p < 0.001$; non-wordlike vs. baseline: 14.30 ± 2.99, t_{18} = 4.78, *FDR*-corrected, $p < 0.001$) and in the POI$_{common}$ (wordlike vs. baseline: 22.72 ± 3.10, t_{18} =7.33, *FDR*-corrected, $p < 0.001$; non-wordlike vs. baseline: 18.33 ± 3.45, t_{18} = 5.32, *FDR*-corrected, $p < 0.001$), except for in the comparison of the activation pattern between wordlike and non-wordlike (4.48 ± 3.39, t_{18} = 1.32, $p = 0.20$) (Fig. 5). In addition, the leave-one-out SVM searchlight analysis across the entire VOTC also yielded a cluster with a significant difference in the activation pattern in the VOTC for each pair of the three conditions (*GRF* corrected, voxel-level $p < 0.001$, cluster-level $p < 0.025$, one-tailed). The cluster representing the pattern difference between the wordlike and non-wordlike conditions included 85 voxels (peak coordinates: –48, –50, and –16), the cluster between the wordlike condition and baseline included 632 voxels (peak coordinates: –56, –50, and –10), and the cluster between the non-wordlike condition and baseline included 494 voxels (peak coordinates: –50, –54, and –12). More relevantly, the three clusters strongly overlapped with or contained the ROI$_{common}$ (Fig. 6). These findings further validate that, compared with other areas within the VOTC, the ROI$_{common}$ with diverse activation patterns uniquely supports distinct representations between the training conditions.

For the one-back task during the post-training scan, the split-half correlation analysis revealed that the *D* values of the activation pattern between each pair of conditions were not significant in the ROI$_{common}$ (wordlike vs. non-wordlike: 0.03 ± 0.08, t_{16} = 0.43, $p = 0.68$; wordlike vs. baseline: –0.08 ± 0.06, $t_{16\,=}$ –1.31, $p = 0.21$; non-wordlike vs. baseline: 0.13 ± 0.07, t_{16} = 1.90, $p = 0.08$); or POI$_{common}$ (wordlike vs. non-wordlike: –0.02 ± 0.06, t_{17} = –0.30, $p = 0.76$; wordlike vs. baseline: –0.01 ± 0.08, t_{17} = –0.12, $p = 0.91$; non-wordlike vs. baseline: 0.01 ± 0.05, t_{17} = 0.20, $p = 0.84$). The results obtained without implementing a category-specific beta step and Fisher z step similarly exhibited a consistent pattern (Supplementary Fig. 3). Similarly, the leave-one-out SVM analysis in the ROI$_{common}$ did not reveal distinct activation patterns between each pair of conditions in either the ROI$_{common}$ (wordlike vs. non-wordlike: 0.23 ± 2.57, t_{17} = 0.09, $p = 0.93$; wordlike vs. baseline: 2.26 ± 2.86, t_{17} = 0.79, $p = 0.44$; non-wordlike vs. baseline: 3.13 ± 1.80, t_{17} = 1.74, $p = 0.10$) or POI$_{common}$ (wordlike vs. non-wordlike: 1.26 ± 2.75, t_{18} = 0.46, $p = 0.65$; wordlike vs. baseline: 0.49 ± 2.41, t_{18} = 0.21, $p = 0.84$; non-wordlike vs. baseline: –1.81 ± 2.69, t_{18} = –0.67, $p = 0.51$) (Fig. 5). Moreover, the leave-one-out SVM analysis in the VOTC did not indicate significant clusters with distinct activation patterns between each pair of conditions. These null results might be due to the low cognitive demand of this task.

Similarly, for the one-back task during the pre-training scan, the difference values of the activation patterns of both the ROI$_{common}$ and

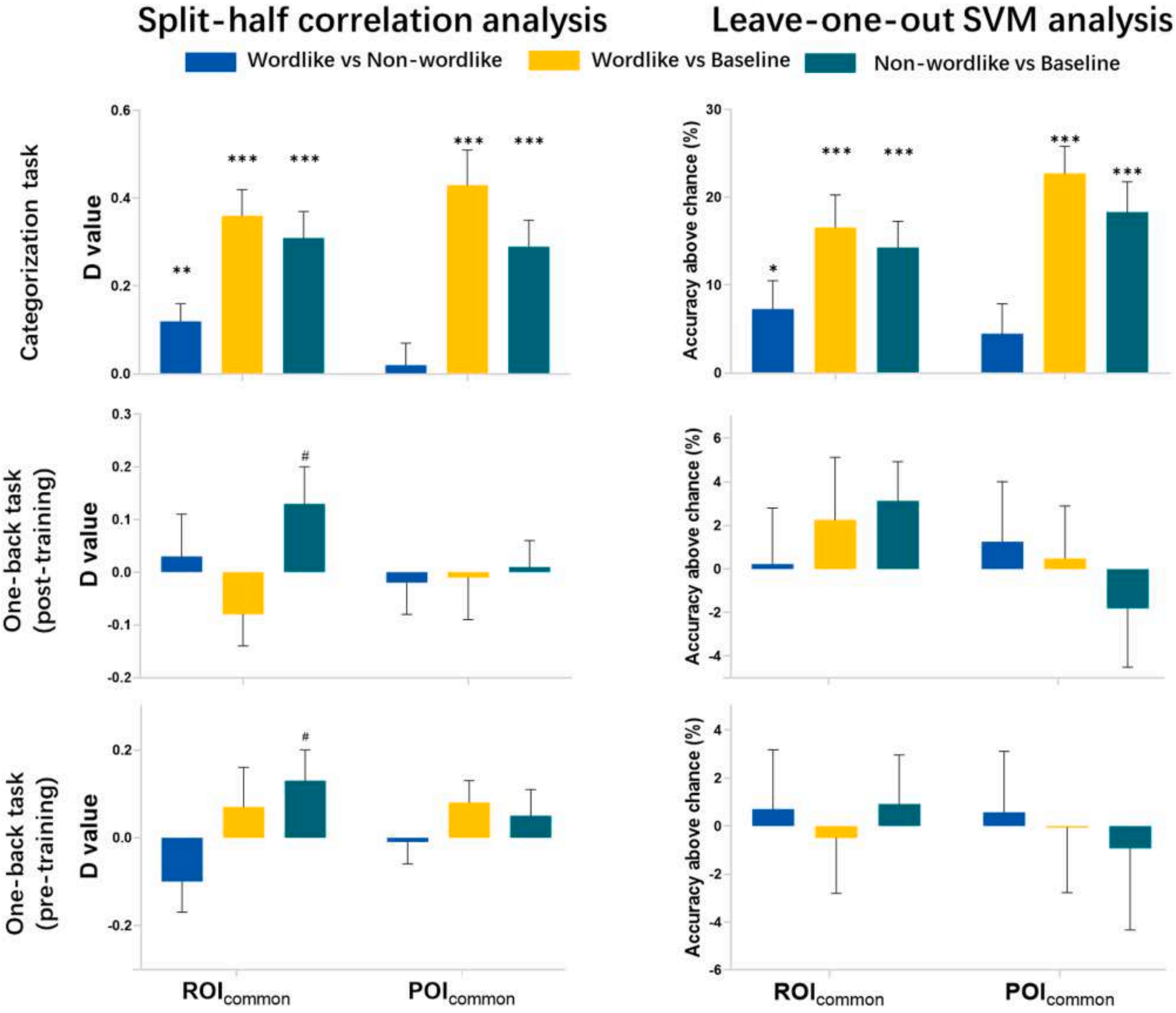

Fig. 5. Results of split-half correlation analysis and leave-one-out SVM analysis for ROI$_{common}$ and POI$_{common}$. The categorization, post-training, and pre-training one-back tasks are shown from top to bottom. The left side shows the split-half correlation analysis, where the D value represents the difference between within-condition and between-condition correlation coefficients. The right side displays the leave-one-out SVM analysis, using the classification accuracy above chance (50 %) to detect differences in activation patterns between the two conditions. SVM = support vector machine. The error bars indicate the SEMs. #: $p \leq 0.10$; *: *FDR* corrected, $p \leq 0.05$; **: *FDR* corrected, $p \leq 0.01$; ***: *FDR* corrected, $p \leq 0.005$.

POI$_{common}$ between each pair of conditions were not significant based on the split-half correlation analysis in both areas (ROI$_{common}$: wordlike vs. non-wordlike: –0.1 ± 0.07, t_{18} = –1.55, p = 0.14; wordlike vs. baseline: 0.07 ± 0.09, t_{18} = 0.85, p = 0.41; non-wordlike vs. baseline: 0.13 ± 0.07, t_{18} = 1.89, p = 0.07; POI$_{common}$: wordlike vs. non-wordlike: –0.01 ± 0.05, t_{17} = –0.14, p = 0.89; wordlike vs. baseline: 0.08 ± 0.05, t_{1_7} = 1.42, p = 0.17; non-wordlike vs. baseline: 0.05 ± 0.06, t_{1_7} = 0.86, p = 0.40). The results obtained without utilizing a category-specific beta step and Fisher z step consistently demonstrated a similar pattern (Supplementary Fig. 3). Moreover, the leave-one-out SVM analysis in the ROI$_{common}$ (wordlike vs. non-wordlike: 0.71 ± 2.47, t_{18} = 0.29, p = 0.78; wordlike vs. baseline: –0.49 ± 2.31, t_{18} = –0.21, p = 0.83; non-wordlike vs. baseline: 0.93 ± 2.03, t_{18} = 0.46, p = 0.65) and the leave-one-out SVM analysis in the POI$_{common}$ (wordlike vs. non-wordlike: 0.58 ± 2.53, t_{17} = 0.23, p = 0.82; wordlike vs. baseline: –0.06 ± 2.71, t_{17} = –0.02, p = 0.98; non-wordlike vs. baseline: –0.93 ± 3.39, Z = 0.07, p = 0.95) (Fig. 5). The leave-one-out SVM analysis in the VOTC also did not reveal any cluster with significant differences in activation patterns in the VOTC between the conditions. These results suggest that the subjects presented similar activation patterns for the three types of meaningless figures before they learned the categorical features.

The above results indicate that the three conditions induced similar activation patterns in the VOTC before training. However, after approximately 13 training sessions for learning word and nonword features, different activation patterns emerged between the conditions, indicating an obvious categorical dissociation in the activation pattern within the VOTC.

4. Discussion

Using meaningless novel figures with different nonvisual features and controlled visual features, we investigated how learning experiences from top-down processing modulate category-specific regions in the VOTC. Our findings indicate that learning experiences from top-down processing predominantly impact activation intensity and pattern rather than activation location.

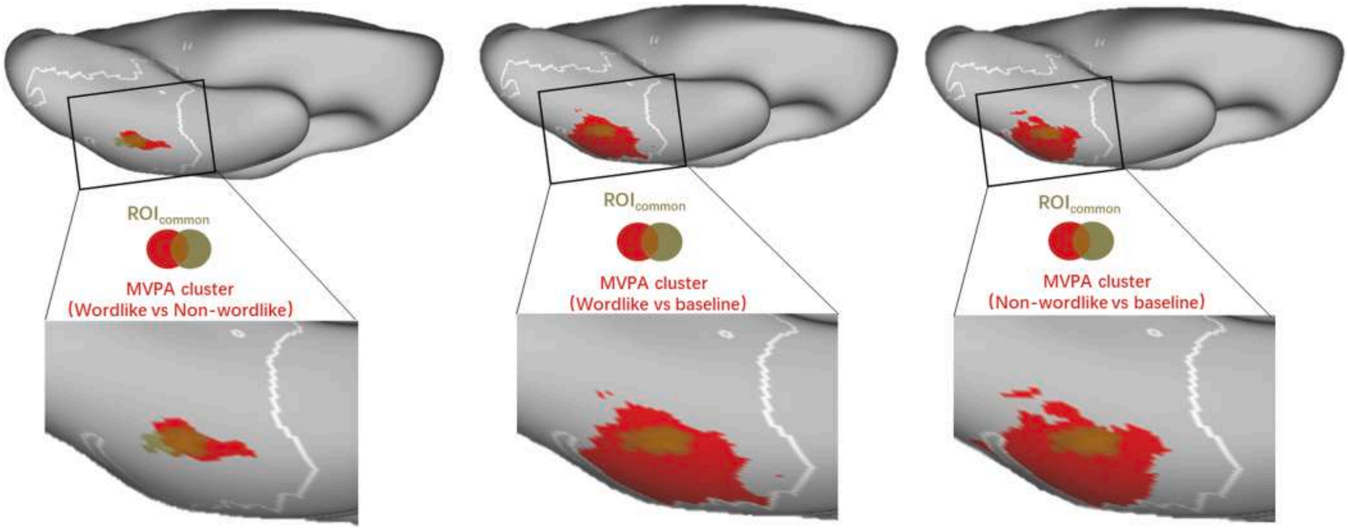

Fig. 6. Results of leave-one-out SVM searchlight analysis across the entire VOTC. From left to right, the red cluster represents the significant searchlight cluster (i.e., MVPA cluster) for the contrast wordlike vs. non-wordlike, wordlike vs. baseline, and non-wordlike vs. baseline in the categorization task. The green cluster corresponds to the location of ROI$_{common}$, and the brown cluster indicates the overlap between searchlight clusters and ROI$_{common}$.

4.1. How top-down nonvisual experience modulates the VOTC

In the present study, we found that the locations of category-specific regions for wordlike and non-wordlike conditions highly overlap. This suggested that stimuli with similar visual features but different top-down nonvisual features may share the same spatial location in the VOTC, which indicated that the location in the VOTC activated by visual stimuli might not be determined by the top-down experience but mainly by bottom-up visual features.

Furthermore, our results replicated our recent finding that learning the top-down nonvisual features of words (e.g., pronunciation) changes the activity intensity in the VOTC's word-related region (Li et al., 2020). This finding is also consistent with the finding that iconic objects (e.g., the Eiffel Tower) elicit stronger activation in the VWFA than noniconic objects (e.g., towers) (Song et al., 2012). Additionally, we observed a significant disparity in activation intensity between the wordlike and non-wordlike conditions, suggesting that top-down nonvisual features of visual stimuli might contribute to the strength of activation in the VOTC (Gauthier et al., 1999; Op de Beeck et al., 2006; Jiang et al., 2007). Notably, this difference in activation cannot be solely attributed to the familiarity effect, as the final test of behavior training and the categorization decision task during fMRI scanning showed no significant difference between wordlike and non-wordlike stimuli.

The results also showed that in the same location of the VOTC (i.e., ROI$_{common}$ or POI$_{common}$), the activation pattern for the wordlike condition was distinct from that for the non-wordlike condition. The literature has shown that the activation of neural populations in the VOTC can represent higher-level features beyond visual information, such as orthography (Zhao et al., 2017; Taylor et al., 2019), pronunciation (Zhao et al., 2017) and semantics (Martin et al., 2018; Taylor et al., 2019). These results indicate that the activation pattern of the same neural population could also encode different top-down nonvisual features (Chen et al., 2017; Amaral et al., 2021).

Overall, our findings support the notion that top-down nonvisual features influence the intensity and pattern of activation within the VOTC for different categories. The results suggested how top-down information shapes the representation of categories in the human brain and provided insights into the complex processes involved in category learning and visual object recognition.

4.2. Possible effects of bottom-up visual experience on the VOTC

Our experiments revealed that top-down information is essential for VOTC's response. However, we cannot neglect the possible impact of bottom-up visual features on VOTC regions. On the contrary, our findings may imply a role for visual features in VOTC organization. This is evident from our observation that the locations of regions specific to stimuli with different nonvisual features but homogenous visual features (i.e., wordlike or non-wordlike conditions) exhibited a high degree of overlap. Indeed, some training studies have demonstrated that novel figures with distinct low-level visual features but the same nonvisual features can elicit disparate activation locations in the VOTC, indicating the important role of low-level features in determining the topography of the VOTC (Moore et al., 2014; Srihasam et al., 2014). These low-level visual features might include the stimuli's curvature (Nasr et al., 2014), eccentricity (Levy et al., 2001; Hasson et al., 2002), real-world size (Konkle and Oliva, 2011, 2012), shape (Bao et al., 2020), and so on. Therefore, if we used another type of visual feature, the location of the category-specific region might differ. Additional research is required to ascertain whether this category-specific region also exerts discriminative effects on stimuli with other types of visual features beyond those explored in our current study.

4.3. The growth and maturity of category-specific regions in the VOTC

Our experiments unveiled a category-related effect for the novel stimuli with controlled visual features after learning their distinct nonvisual category features. However, we cannot disregard the impact of inherent factors, such as the connectivity fingerprint of the VOTC, which might serve as the structural foundation of the growth of category-related regions (Wang et al., 2017; Li et al., 2018). Supportive evidence comes from previous studies that have found the white matter connectivity pattern between the VOTC and other brain areas can successfully predict the functional activity intensity of visual stimuli in the category-specific regions of the VOTC in adults (Saygin et al., 2011; Osher et al., 2016; Ekstrand et al., 2020), even can predict the location of the later-appeared VWFA in young children (Saygin et al., 2016). Therefore, both the innate characteristics and the acquired experience within the VOTC might affect the formation of category-related regions in the VOTC. The innate part (e.g., the early presence of domain-specific connectivity from the VOTC to other brain regions) might constrain where category-related areas will emerge, and experience-driven learning is similar to a lighter way to let the category-related area finally come into existence (Srihasam et al., 2012; Op de Beeck et al., 2019; Arcaro and Livingstone, 2021).

The maturity of category-specific regions is evident in their ability to automate the processing of various categories of information, regardless of the task demands. For instance, basic visual tasks (e.g., one-back and color dot detection tasks) typically yield significant differences between categories in the well-known category-specific areas in VOTC (e.g., VWFA, FFA) (Stigliani et al., 2015; Coggan et al., 2016). However, the categorical distinction between the wordlike and non-wordlike conditions disappeared in our one-back experiments. This may be attributed to the fact that training in two categories over a short period may not lead to the same level of automated processing in the VOTC as observed in classic categories like words and faces. In such cases, the VOTC categorizes information differently only when the task involves categorical information, necessitating a deeper level of task demand (Ju and Bassett, 2020; White et al., 2023). Hence, achieving the maturation of category-specific regions may require a considerable amount of time and effort (Nordt et al., 2023).

5. Limitations

This study has at least the following limitations. 1) Training effects in the postscanning one-back task were weaker than those in the categorization task, possibly due to the different cognitive demands of processing learned information between the two tasks. 2) Although the participants performed perfectly for all the items, the limited number of trained items in each condition, the constrained trained features for each item, and the incomparability of these items to real objects in terms of familiarity and processing depth may have contributed to a weaker training effect in our study. 3) Due to possible limitations in how well the training process reflects real-world language acquisition (e.g., the grammatical training task may not perfectly mimic the learning of actual words), the knowledge gained in the wordlike condition might not entirely capture specific word-related information. However, we believe it does involve some word-related knowledge, as the pronunciation training imitates natural language acquisition, and we observed a significant overlap between the wordlike clusters and real word regions obtained from a meta-analysis using the Neurosynth database (Yarkoni et al., 2011) (see Supplementary Fig. 4). Future research calls for the incorporation of training tasks with higher ecological validity to obtain language representations that closely resemble real language acquisition. 4) We found that the wordlike cluster size in VOTC was larger than the non-wordlike cluster size, and the difference in the training size of the parameter feature space may impact the cluster size of these category-specific regions. 5) During training, subjects might have been aware that the visual familiarity condition served as the baseline, potentially introducing response bias. However, our study primarily aimed to investigate neural differences between the wordlike and non-wordlike conditions rather than comparing them to the visual familiarity condition. Therefore, any potential response bias from participants is expected to have minimal impact on our results.

6. Conclusion

Utilizing the feature-based associative learning paradigm for meaningless novel figures, we found that learning nonvisual features of different categories for the visually controlled novel stimulus affects the intensities and patterns of activity in the category-specific area of the VOTC but does not influence the activation location. These findings offer critical insights into how top-down information about objects is encoded into neural representations in the VOTC.

CRediT authorship contribution statement

Xiangqi Luo: Writing – original draft, Visualization, Investigation, Formal analysis, Data curation, Conceptualization, Writing – review & editing. **Mingyang Li:** Writing – original draft, Visualization, Investigation, Formal analysis, Data curation, Conceptualization. **Jiahong Zeng:** Writing – review & editing, Investigation, Data curation, Conceptualization. **Zhiyun Dai:** Writing – review & editing, Investigation. **Zhenjiang Cui:** Writing – review & editing. **Minhong Zhu:** Writing – review & editing. **Mengxin Tian:** Writing – review & editing. **Jiahao Wu:** Writing – review & editing. **Zaizhu Han:** Writing – review & editing, Supervision, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (32271091, 81972144, 81870833, and 82372555). We extend our heartfelt thanks to all research participants for their kindness, patience, and conscientious cooperation throughout the study.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi 10.1016/j.neuroimage.2024.120520.

References

Amaral, L., Bergstrom, F., Almeida, J., 2021. Overlapping but distinct: Distal connectivity dissociates hand and tool processing networks. Cortex 140, 1–13. https://doi.org/10.1016/j.cortex.2021.03.011.

Arcaro, M.J., Livingstone, M.S., 2021. On the relationship between maps and domains in inferotemporal cortex. Nat. Rev. Neurosci. 22 (9), 573–583. https://doi.org/10.1038/s41583-021-00490-4.

Arcaro, M.J., Schade, P.F., Vincent, J.L., Ponce, C.R., Livingstone, M.S., 2017. Seeing faces is necessary for face-domain formation. Nat. Neurosci. 20 (10), 1404–1412. https://doi.org/10.1038/nn.4635.

Bao, P., She, L., McGill, M., Tsao, D.Y., 2020. A map of object space in primate inferotemporal cortex. Nature 583 (7814), 103–108. https://doi.org/10.1038/s41586-020-2350-5.

Bi, Y., Wang, X., Caramazza, A., 2016. Object domain and modality in the ventral visual pathway. Trend. Cogn. Sci. 20 (4), 282–290. https://doi.org/10.1016/j.tics.2016.02.002.

Bracci, S., Op de Beeck, H., 2016. Dissociations and associations between shape and category representations in the two visual pathways. J. Neurosci. 36 (2), 432–444. https://doi.org/10.1523/JNEUROSCI.2314-15.2016.

Carreiras, M., Armstrong, B.C., Perea, M., Frost, R., 2014. The what, when, where, and how of visual word recognition. Trend. Cogn. Sci. 18 (2), 90–98. https://doi.org/10.1016/j.tics.2013.11.005.

Chang, C.C., Lin, C.J., 2011. LIBSVM: A Library for Support Vector Machines. ACM Trans. Intell. Syst. Technol. 2 (3), 1–27. https://doi.org/10.1145/1961189.1961199.

Chen, L., Wassermann, D., Abrams, D.A., Kochalka, J., Gallardo-Diez, G., Menon, V., 2019. The visual word form area (VWFA) is part of both language and attention circuitry. Nat. Commun. 10 (1), 5601. https://doi.org/10.1038/s41467-019-13634-z.

Chen, Q., Garcea, F.E., Almeida, J., Mahon, B.Z., 2017. Connectivity-based constraints on category-specificity in the ventral object processing pathway. Neuropsychologia 105, 184–196. https://doi.org/10.1016/j.neuropsychologia.2016.11.014.

Coggan, D.D., Liu, W., Baker, D.H., Andrews, T.J., 2016. Category-selective patterns of neural response in the ventral visual pathway in the absence of categorical information. Neuroimage 135, 107–114. https://doi.org/10.1016/j.neuroimage.2016.04.060.

Dehaene-Lambertz, G., Monzalvo, K., Dehaene, S., 2018. The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. PLoS Biol. 16 (3), e2004103 https://doi.org/10.1371/journal.pbio.2004103.

Dehaene, S., Cohen, L., Morais, J., Kolinsky, R., 2015. Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. Nat. Rev. Neurosci. 16 (4), 234–244. https://doi.org/10.1038/nrn3924.

Dehaene, S., Pegado, F., Braga, L.W., 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 (6009), 1359–1364. https://doi.org/10.1126/science.1194140.

Ekstrand, C., Neudorf, J., Kress, S., Borowsky, R., 2020. Structural connectivity predicts functional activation during lexical and sublexical reading. Neuroimage 218, 117008. https://doi.org/10.1016/j.neuroimage.2020.117008.

Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarski, P., Gore, J.C., 1999. Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nat. Neurosci. 2 (6), 568–573. https://doi.org/10.1038/9224.

Grill-Spector, K., Weiner, K.S., 2014. The functional architecture of the ventral temporal cortex and its role in categorization. Nat. Rev. Neurosci. 15 (8), 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. Trend. Cogn. Sci. 19 (7), 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 (3), 479–490. https://doi.org/10.1016/S0896-6273(02)00662-1.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293 (5539), 2425–2430. https://doi.org/10.1126/science.1063736.

Hebart, M.N., Gorgen, K., Haynes, J.D., 2015. The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. Front. Neuroinform. 8, 88. https://doi.org/10.3389/fninf.2014.00088.

Ishai, A., Ungerleider, L.G., Martin, A., Schouten, J.L., Haxby, J.V., 1999. Distributed representation of objects in the human ventral visual pathway. Proc. Natl. Acad. Sci. U.S.A. 96 (16), 9379–9384. https://doi.org/10.1073/pnas.96.16.9379.

Jiang, X., Bradley, E., Rini, R.A., Zeffiro, T., Vanmeter, J., Riesenhuber, M., 2007. Categorization training results in shape- and category-selective human neural plasticity. Neuron 53 (6), 891–903. https://doi.org/10.1016/j.neuron.2007.02.015.

Ju, H., Bassett, D.S., 2020. Dynamic representations in networked neural systems. Nat. Neurosci. 23 (8), 908–917. https://doi.org/10.1038/s41593-020-0653-3.

Kang, J., Kim, H., Hwang, S.H., Han, M., Lee, S.H., Kim, H.F., 2021. Primate ventral striatum maintains neural representations of the value of previously rewarded objects for habitual seeking. Nat. Commun. 12 (1), 2100. https://doi.org/10.1038/s41467-021-22335-5.

Kim, J.S., Kanjlia, S., Merabet, L.B., Bedny, M., 2017. Development of the visual word form area requires visual experience: evidence from blind Braille readers. J. Neurosci. 37 (47), 11495–11504. https://doi.org/10.1523/JNEUROSCI.0997-17.2017.

Knecht, S., Deppe, M., Dräger, B., Bobe, L., Lohmann, H., Ringelstein, E.-B., Henningsen, H., 2000. Language lateralization in healthy right-handers. Brain 123 (1), 74–81. https://doi.org/10.1093/brain/123.1.74.

Konkle, T., Oliva, A., 2011. Canonical visual size for real-world objects. J. Exp. Psychol. Hum. Percept. Perform. 37 (1), 23–37. https://doi.org/10.1037/a0020413.

Konkle, T., Oliva, A., 2012. A real-world size organization of object responses in occipitotemporal cortex. Neuron 74 (6), 1114–1124. https://doi.org/10.1016/j.neuron.2012.04.036.

Levy, I., Hasson, U., Avidan, G., Hendler, T., Malach, R., 2001. Center–periphery organization of human object areas. Nature 4, 533–539. https://doi.org/10.1038/87490.

Li, M., Xu, Y., Luo, X., Zeng, J., Han, Z., 2020. Linguistic experience acquisition for novel stimuli selectively activates the neural network of the visual word form area. Neuroimage 215, 116838. https://doi.org/10.1016/j.neuroimage.2020.116838.

Li, Y., Fang, Y., Wang, X., Song, L., Huang, R., Han, Z., Gong, G., Bi, Y., 2018. Connectivity of the ventral visual cortex is necessary for object recognition in patients. Hum. Brain Mapp. 39 (7), 2786–2799. https://doi.org/10.1002/hbm.24040.

Liu, Y., Shi, G., Li, M., Xing, H., Song, Y., Xiao, L., Guan, Y., Han, Z., 2021. Early top-down modulation in visual word form processing: evidence from an intracranial SEEG study. J. Neurosci. 41 (28), 6102–6115. https://doi.org/10.1523/JNEUROSCI.2288-20.2021.

Mahon, B.Z., Caramazza, A., 2009. Concepts and categories: a cognitive neuropsychological perspective. Annu. Rev. Psychol. 60 (1), 27–51. https://doi.org/10.1146/annurev.psych.60.110707.163532.

Mahon, B.Z., Caramazza, A., 2011. What drives the organization of object knowledge in the brain? Trend. Cogn. Sci. 15 (3), 97–103. https://doi.org/10.1016/j.tics.2011.01.004.

Malach, R., Levy, I., Hasson, U., 2002. The topography of high-order human object areas. Trend. Cogn. Sci. 6 (4), 176–184. https://doi.org/10.1016/s1364-6613(02)01870-3.

Maldjian, J.A., Laurienti, P.J., Kraft, R.A., Burdette, J.H., 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 19 (3), 1233–1239. https://doi.org/10.1016/s1053-8119(03)00169-1.

Mars, R.B., Passingham, R.E., Jbabdi, S., 2018. Connectivity fingerprints: From areal descriptions to abstract spaces. Trend. Cogn. Sci. 22 (11), 1026–1037. https://doi.org/10.1016/j.tics.2018.08.009.

Martin, C.B., Douglas, D., Newsome, R.N., Man, L.L., Barense, M.D., 2018. Integrative and distinctive coding of visual and conceptual object features in the ventral visual stream. Elife 7. https://doi.org/10.7554/eLife.31873.

Mattioni, S., Rezk, M., Battal, C., Bottini, R., Cuculiza Mendoza, K.E., Oosterhof, N.N., Collignon, O., 2020. Categorical representation from sound and sight in the ventral occipito-temporal cortex of sighted and blind. Elife 9. https://doi.org/10.7554/eLife.50732.

Moore, M.W., Durisko, C., Perfetti, C.A., Fiez, J.A., 2014. Learning to read an alphabet of human faces produces left-lateralized training effects in the fusiform gyrus. J Cogn Neurosci 26 (4), 896–913. https://doi.org/10.1162/jocn_a_00506.

Morgenstern, Y., Hartmann, F., Schmidt, F., Tiedemann, H., Prokott, E., Maiello, G., Fleming, R.W., 2021. An image-computable model of human visual shape similarity. PLoS Comput. Biol. 17 (6), e1008981 https://doi.org/10.1371/journal.pcbi.1008981.

Nasr, S., Echavarria, C.E., Tootell, R.B., 2014. Thinking outside the box: rectilinear shapes selectively activate scene-selective cortex. J. Neurosci. 34 (20), 6721–6735. https://doi.org/10.1523/JNEUROSCI.4802-13.2014.

Nordt, M., Gomez, J., Natu, V.S., Rezai, A.A., Finzi, D., Kular, H., Grill-Spector, K., 2023. Longitudinal development of category representations in ventral temporal cortex predicts word and face recognition. Nat. Commun. 14 (1), 8010. https://doi.org/10.1038/s41467-023-43146-w.

Ojemann, J.G., Akbudak, E., Snyder, A.Z., McKinstry, R.C., Raichle, M.E., Conturo, T.E., 1997. Anatomic localization and quantitative analysis of gradient refocused echo-planar fMRI susceptibility artifacts. Neuroimage 6 (3), 156–167. https://doi.org/10.1006/nimg.1997.0289.

Op de Beeck, H.P., Pillet, I., Ritchie, J.B., 2019. Factors determining where category-selective areas emerge in visual cortex. Trend. Cogn. Sci. 23 (9), 784–797. https://doi.org/10.1016/j.tics.2019.06.006.

Op de Beeck, H.P., Baker, C.I., DiCarlo, J.J., Kanwisher, N.G., 2006. Discrimination training alters object representations in human extrastriate cortex. J. Neurosci. 26 (50), 13025–13036. https://doi.org/10.1523/JNEUROSCI.2481-06.2006.

Osher, D.E., Saxe, R.R., Koldewyn, K., Gabrieli, J.D., Kanwisher, N., Saygin, Z.M., 2016. Structural connectivity fingerprints predict cortical selectivity for multiple visual categories across cortex. Cereb. Cortex 26 (4), 1668–1683. https://doi.org/10.1093/cercor/bhu303.

Price, C.J., Devlin, J.T., 2011. The interactive account of ventral occipitotemporal contributions to reading. Trend. Cogn. Sci. 15 (6), 246–253. https://doi.org/10.1016/j.tics.2011.04.001.

Rauschecker, A.M., Bowen, R.F., Perry, L.M., Kevan, A.M., Dougherty, R.F., Wandell, B. A., 2011. Visual feature-tolerance in the reading network. Neuron 71 (5), 941–953. https://doi.org/10.1016/j.neuron.2011.06.036.

Reich, L., Szwed, M., Cohen, L., Amedi, A., 2011. A ventral visual stream reading center independent of visual experience. Curr. Biol. 21 (5), 363–368. https://doi.org/10.1016/j.cub.2011.01.040.

Saygin, Z.M., Osher, D.E., Koldewyn, K., Reynolds, G., Gabrieli, J.D., Saxe, R.R., 2011. Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. Nat. Neurosci. 15 (2), 321–327. https://doi.org/10.1038/nn.3001.

Saygin, Z.M., Osher, D.E., Norton, E.S., Youssoufian, D.A., Beach, S.D., Feather, J., Gaab, N., Gabrieli, J.D., Kanwisher, N., 2016. Connectivity precedes function in the development of the visual word form area. Nat. Neurosci. 19 (9), 1250–1255. https://doi.org/10.1038/nn.4354.

Seghier, M.L., Price, C.J., 2011. Explaining left lateralization for words in the ventral occipitotemporal cortex. J. Neurosci. 31 (41), 14745–14753. https://doi.org/10.1523/JNEUROSCI.2238-11.2011.

Song, Y., Tian, M., Liu, J., 2012. Top-down processing of symbolic meanings modulates the visual word form area. J. Neurosci. 32 (35), 12277–12283. https://doi.org/10.1523/JNEUROSCI.1874-12.2012.

Srihasam, K., Vincent, J.L., Livingstone, M.S., 2014. Novel domain formation reveals proto-architecture in inferotemporal cortex. Nat. Neurosci. 17 (12), 1776–1783. https://doi.org/10.1038/nn.3855.

Srihasam, K., Mandeville, J.B., Morocz, I.A., Sullivan, K.J., Livingstone, M.S., 2012. Behavioral and anatomical consequences of early versus late symbol training in macaques. Neuron 73 (3), 608–619. https://doi.org/10.1016/j.neuron.2011.12.022.

Stigliani, A., Weiner, K.S., Grill-Spector, K., 2015. Temporal processing capacity in high-level visual cortex is domain specific. J. Neurosci. 35 (36), 12412–12424. https://doi.org/10.1523/jneurosci.4822-14.2015.

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. U.S.A. 116 (36), 17723–17728. https://doi.org/10.1073/pnas.1818575116.

Thorpe, S., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. Nature 381 (6582), 520–522. https://doi.org/10.1038/381520a0.

van den Hurk, J., Van Baelen, M., Op de Beeck, H.P., 2017. Development of visual category selectivity in ventral visual cortex does not require visual experience. Proc. Natl. Acad. Sci. U.S.A. 114 (22), E4501–E4510. https://doi.org/10.1073/pnas.1612862114.

Wang, X., He, C., Peelen, M.V., Zhong, S., Gong, G., Caramazza, A., Bi, Y., 2017. Domain selectivity in the parahippocampal gyrus is predicted by the same structural connectivity patterns in blind and sighted individuals. J. Neurosci. 37 (18), 4705–4716. https://doi.org/10.1523/JNEUROSCI.3622-16.2017.

White, A.L., Kay, K.N., Tang, K.A., Yeatman, J.D., 2023. Engaging in word recognition elicits highly specific modulations in visual cortex. Curr. Biol. 33 (7), 1308–1320. https://doi.org/10.1016/j.cub.2023.02.042 e1305.

Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. Nat. Method. 8 (8), 665–670. https://doi.org/10.1038/nmeth.1635.

Zhao, L., Chen, C., Shao, L., Wang, Y., Xiao, X., Chen, C., Yang, J., Zevin, J., Xue, G., 2017. Orthographic and phonological representations in the fusiform cortex. Cereb. Cortex 27 (11), 5197–5210. https://doi.org/10.1093/cercor/bhw300.