Shared and Independent Neural Representation Between Visual Perception and Mental Imagery

Yingying Huang

University of Glasgow <https://orcid.org/0000-0001-5420-5122>

Frank Pollick

University of Glasgow

Ming Liu

South China Normal University

Delong Zhang (✉ delong.zhang@m.scnu.edu.cn)

South China Normal University

Research Article

Keywords: fMRI, visual perception, mental imagery, encoding model, representation

Posted Date: September 7th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-788978/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Shared and independent neural representation between visual perception and mental imagery

Yingying Huang^{1,2}, Frank Pollick², Ming Liu¹, Delong Zhang^{1*}

¹ Key Laboratory of Brain, Cognition and Education Sciences, Ministry of Education, China; School of Psychology, Center for Studies of Psychological Application, and Guangdong Key Laboratory of Mental Health and Cognitive Science, South China Normal University, China

² School of Psychology, University of Glasgow, Glasgow, UK

* Correspondence to: Delong Zhang, Key Laboratory of Brain, Cognition and Education Sciences, Ministry of Education, China; School of Psychology, Center for Studies of Psychological Application, and Guangdong Key Laboratory of Mental Health and Cognitive Science, South China Normal University, Guangzhou 510631, China, E-mail: delong.zhang@m.scnu.edu.cn (DZ)

(Note: Yingying Huang is now studying the University of Glasgow, the work was done in the South China Normal University.)

Declaration of Interest: none

Author contribution

Yingying Huang: Conceptualization, Methodology, Formal Analysis, Resources, Data Curation, Writing – Original draft; **Frank Pollick:** Writing – Reviewing and Editing; **Ming Liu:** Supervision; **Delong Zhang:** Conceptualization, Methodology, Writing – Reviewing and Editing, Project Administration, Funding Acquisition

Acknowledgements

Thanks for Professor Zhaoping Li, Yibo Wang, Junchao Li and Bingqing Jiao giving writing suggestions and assistances.

Funding

This work was supported by the National Natural Science Foundation of China (No. 31600907).

Abstract

Visual mental imagery and visual perception have been shown to share a hierarchical topological visual structure of neural representation. Meanwhile, many studies have reported a dissociation of neural substrate between mental imagery and perception in function and structure. However, we have limited knowledge about how the visual hierarchical cortex involved into internally generated mental imagery and perception with visual input. Here we used a dataset from previous fMRI research (Horikawa & Kamitani, 2017), which included a visual perception and an imagery experiment with human participants. We trained two types of voxel-wise encoding models, based on Gabor features and activity patterns of high visual areas, to predict activity in the early visual cortex (EVC, i.e., V1, V2, V3) during perception, and then evaluated the performance of these models during mental imagery. Our results showed that during perception and imagery, activities in the EVC could be independently predicted by the Gabor features and activity of high visual areas via encoding models, which suggested that perception and imagery might share neural representation in the EVC. We further found that there existed a Gabor-specific and a non-Gabor-specific neural response pattern to stimuli in the EVC, which were shared by perception and imagery. These findings provide insight into mechanisms of how visual perception and imagery shared representation in the EVC.

Key words: fMRI; visual perception; mental imagery; encoding model; representation

“Whilst part of what we perceive comes through our senses from the object before us, another part (and it may be the larger part) always comes out of our own head”.

“William James (1890), *The Principle of Psychology*”.

1 Introduction

Every day, we are bombarded with plenty of visual stimulation, such as colours, textures, objects, and our brain then selectively processes the information to generate visual experiences. Therefore, visual perception is a kind of reflection of the interaction between feedforward sensory input, externally driven by “bottom-up” pathways, and feedback signals, internally generated from “top-down” pathways (Hsieh et al., 2010; Kastner et al., 1998). Another form of mental process that can generate internally visual experience is mental imagery, which refers to generation and representation of a visual image without corresponding feedforward stimuli from the real world (Andersson et al., 2019; Kosslyn, 1996; Kosslyn & Thompson, 2003).

What is the relationship between visual perception and mental imagery generating visual experiences within our brain? To address this question, much empirical evidence from behavioral and neural aspects supports that visual perception and mental imagery have similar functions during sensory processing. For example, in a series of behavioral experiments, participants could reach similar performance between scanning mental images and visual perception (Borst & Kosslyn, 2008). Some researchers have proposed that mental imagery resembles weak perception (Pearson, 2019; Pearson & Kosslyn, 2015), based on the similar neural mechanism of sensory processing between them (Borst & Kosslyn, 2008; Cichy et al., 2012; Dijkstra et al., 2019; Ishai & Sagi, 1995; Maier et al., 2020; Reddy et al., 2010; Stokes et al., 2009; Xie et al., 2020), and according to perception anticipation theory (Aitken et al., 2020; Kok et al., 2012; Kosslyn & Thompson, 2003; Sohoglu et al., 2012) it may have some impact on the top-down modulation process of perception (Berger & Ehrsson, 2013; Diekhof et al., 2011; Pearson et al., 2008). Specifically, some functional magnetic resonance imaging (fMRI) studies have provided evidence that mental imagery and visual perception shared brain activity pattern in the early visual cortex (EVC, i.e., V1, V2, and V3) encoding low-level visual features (Kosslyn & Thompson, 2003; Maier et al., 2020) as well as the high visual areas encoding category information (S. H. Lee et al., 2012; O’Craven & Kanwisher, 2000; Stokes et al., 2009).

Recently, advances in machine learning approaches have been used in combination with fMRI data to provide a novel viewpoint for exploring the specific shared neural representation between visual perception and mental imagery (Albers et al., 2013; Horikawa & Kamitani, 2017; S. H. Lee et al., 2012; Pearson et al., 2008; Reddy et al., 2010). Gallant and colleagues firstly decoded visual perception with Gabor features (i.e., spatial frequency, orientation, position and phase) of natural images in the EVC (Kay et al., 2008), and after that they conducted the same decoding analysis to the content of imagining specific famous pictures (Naselaris et al., 2015), which further provided insight into the similarity between visual perception and mental imagery in the EVC. More recently, Horikawa and Kamitani (2017) used features extracted by deep convolutional neural network (CNN) to decode brain activity in the ventral visual stream when participants were seeing and imagining natural images, which considered to support the claim that visual perception and mental imagery share neural representation in the hierarchy of visual system. Overall, multiple studies have collectively shown that neural representation of the

ventral visual cortex during perception has a hierarchical topological structure, where high visual activity represents the semantic category of natural images (Naselaris et al., 2009) and the early visual activity represents low-level visual features (Albers et al., 2013; Cichy et al., 2012; S. H. Lee et al., 2012). The hierarchical topological structure was also observed during mental imagery (Reddy et al., 2010). Despite this, it should be attracted great attention that many previous studies have also reported the existence of a dissociation of function and structure of neural substrate between perception and mental imagery (Butter et al., 1997; Sirigu & Duhamel, 2001). However, it is still unclear how internally generated visual experience interacts with the externally visual stimuli to produce visual awareness during perception.

It is possible to consider that category information and visual features of a natural image may induce a sense of continuity, and thus the corresponding brain activities from high to the EVC might constitute a neural “signature” of the natural image. In the present study, we attempted to elaborate the internal structure of the neural “signature” during perception and mental imagery. To this end, we obtained the fMRI data from previous human research involving visual perception and mental imagery (Horikawa & Kamitani, 2017) and then trained two types of fMRI-based voxel-wise encoding models, i.e., an exogenous encoding model based on Gabor features of input stimuli, and an endogenous encoding model based on the neural connectivity between regions in the high visual areas (i.e., FFA, PPA, and LOC) and in the EVC (Fig 1.1, see more details in Methods section). Based on this, we further generalized these trained voxel-wise encoding models from perception to mental imagery. Next, we investigated the neural relationship between perception and imagery via combination and separation of the predicted neural representation patterns in the EVC with the two types of encoding models. Finally, we divided the voxels in the EVC into two groups according to the specificity and non-specificity of Gabor features, to refine the neural substrate between perception and imagery.

Fig 1.1 about here, please

Fig 1.1 Neural encoding through Gabor wavelet pyramid and brain areas’ connectivity model. When a person was seeing or imaging object image, the brain visual areas were processing neural information. We used the Gabor wavelet pyramid to filter images to get low visual features and then combined the voxel activity pattern in the early visual cortex (V1, V2, and V3) to train voxel-wise Gabor-encoding models. At the same time, we trained the analogical voxel-wise encoding models with neural connectivity between the early visual areas and the high visual areas (FFA, LOC, and PPA). Moreover, we further investigated the neural representation during perception and imagery according to the multi-encoding models we constructed.

2 Material and Methods

2.1 Data acquisition

The fMRI datasets were provided by Kamitani Lab at Kyoto University and ATR (<https://github.com/KamitaniLab>). Here, we gave a brief description of stimuli, datasets, and experimental design.

There were 1250 natural images ($j = 1, 2, 3, 4, \dots, 1250$), which were selected from 200 representative object categories ($c = 1, 2, 3, \dots, 200$), used in the whole study. Five healthy human subjects (four males) joined in the whole experiments. During BOLD-fMRI scanning, each subject was instructed to view natural images at the center of the projection screen (visual perception experiment, two sessions) and to imagine corresponding objects according to the word cues presented on the screen (visual imagery experiment, one test session). There was a

training image session and a test image session in the perception experiment with the same procedures. In each session, there were 24 and 35 separate runs, each of which had a 9 min 54 s duration, and consisted of 50 trials with different images and 5 randomly interspersed repetition trials. These repetition trials were used to keep subjects' attention and subjects were instructed to perform a one-back task and to press a button if trial was repeated. Each image was presented at the center of the screen with 12×12 degrees of visual angle and flashed at 2 Hz for 9 s. In sum, 1200 different images ($j = 1, 2, 3, \dots, 1200$) from 150 object categories (8 images used in each category, $c = 1, 2, 3, \dots, 150$) were presented only once in the training image session. 50 different images ($j = 1201, 1202, \dots, 1250$) from new 50 object categories ($c = 151, 152, \dots, 200$) were presented 35 times each in the test image session. In the imagery experiment, each subject was instructed to read the red word cue presented on the screen and then visually imagine objects with closed eyes after hearing a beep sound. The word cues were names of 50 object categories ($c = 151, 152, \dots, 200$) that were matched to the images presented in the perception test session, but subjects were asked to freely imagine objects from the same category as many as they could during each trial. The imagery session consisted of 20 runs, and each contained 25 imagery trials and each duration was 10 min 39 s.

The detailed procedures of MRI data preprocessing and functional brain region localization can be found in the original paper (Horikawa & Kamitani, 2017). The previous study was supported by the Ethics Committee of ATR, and the data re-analysis was approved by authors of the original paper.

2.2 Gabor feature

We used Gabor features extracted from the 1250 natural images by a Gabor wavelet pyramid (GWP) model to encode the activity in the EVC (Jones & Palmer, 1987; Ringach, 2002). Prior studies demonstrated that the GWP model could be viewed as an appropriate method to describe voxel activity in the EVC (Daugman, 1985; Jones & Palmer, 1987; T. S. Lee, 1996; Rainer et al., 2001). The GWP model has been used to describe the dimensions of space (Deyoe et al., 1996; Engel et al., 1994; Sereno et al., 1995), orientation (Haynes & Rees, 2005; Kamitani & Tong, 2005; Sasaki et al., 2006), and spatial frequency (Olman et al., 2004; Singh et al., 2000) of natural images. According to a prior study (Kay et al., 2008), these Gabor features extracted from natural images by a GWP model could be fitted to voxel responses in the EVC, especially in the primary visual cortex (V1).

In the present study, at first, we constructed the GWP model with a series of Gabor filters. Based on Kay's study (2008), we defined the same wavelets with six spatial frequencies: 1, 2, 4, 8, 16 and 32 cycles per field of view ($\text{FOV} = 192 \times 192 \text{ mm}^2$). At each frequency, f cycles per FOV, wavelets were positioned on an $f \times f$ grid. At each grid position, wavelets occurred at eight orientations: $0^\circ, 22.5^\circ, 45^\circ, 67.5^\circ, 90^\circ, 112.5^\circ, 135^\circ$ and 157.5° , and two quadrature phases were: 0° , and 90° . So, there were 10920 ($(1^2 + 2^2 + 4^2 + 8^2 + 16^2 + 32^2) \times 8 = 10920$) Gabor filters in the GWP model.

We then applied the constructed GWP model to extract Gabor features of the stimuli images. The original $500 \text{ px} \times 500 \text{ px}$ images were down-sampled to $128 \text{ px} \times 128 \text{ px}$ image resolution for the analysis. Accordingly, the features were defined as

$$f(\mathbf{s}) = \log(|W^T \mathbf{s}| + 1) \quad (1)$$

where the f was an $F \times 1$ vector containing the features ($F = 10920$, the number of filters used for the model), and W denoted a matrix of complex Gabor filters. The variable $\mathbf{s}$ corresponds to the matrix of each image used in

the whole experiment, W had as many rows as the pixels in s , and each column contained a different Gabor filter; thus, its dimension was $128^2 \times 10920$. The features were the log of the magnitudes derived from filtering the image by each filter. These parameters corresponded to those used by Gallant and colleagues (Kay et al., 2008; Naselaris et al., 2015).

2.3 Encoding models

The responses in the EVC when people see something could be explained by two sources, namely from the external stimulus and internally generated signal from high visual areas (i.e., FFA, PPA, and LOC). To simulate the two-source information processing, we constructed two types of voxel-wise encoding models, from the Gabor features aspect (exogenous encoding model) and the aspect of neural connectivity between the EVC and high visual areas (endogenous encoding model), to encode the neural activity of each voxel in the EVC.

For the exogenous encoding model, we extracted the BOLD-fMRI signals of each voxel in the EVC related to the 1200 training images ($j = 1, 2, \dots, 1200$) during perception. For each voxel, we regarded the Gabor features as the input variable and the fMRI signal activity as the output response to train the voxel-wise encoding models.

According to the prior study, p was defined as the number of training images and q as the number of input channels. The response of each voxel in the EVC could be modeled as,

$$\mathbf{y} = \mathbf{X}\mathbf{h} + c\mathbf{1} + n \quad (2)$$

Where $\mathbf{y}$ is the set of output responses (i.e., the response of each voxel in the EVC) with the dimension of ($p \times 1$), $\mathbf{X}$ is the set of input channels ($p \times q$), $\mathbf{h}$ is the kernel ($q \times 1$), c is the DC offset (1×1), $\mathbf{1}$ is the vector of constant ones ($p \times 1$), and n is the noise ($p \times 1$). We used the functions from STRflab toolbox (Version 1.45, <http://strflab.berkeley.edu/>) to automatically estimate the model parameters. The model parameters were estimated with gradient descent based on the early stopping algorithm to prevent parameters from over-fitting, and the stopping set consisted of 20% randomly selected responses (Kay et al., 2008; Tugnait, 1994). There was a bootstrap sampling approach for iterative analysis. This procedure was conducted independently for each voxel in the EVC; thus, we finally obtained an encoding model for each voxel in V1, V2, and V3 during perception.

To train the endogenous encoding model, we employed the voxel activity in each high visual area as the input features (each voxel as a feature) to replace the Gabor features in the exogenous encoding model and the fMRI signal activity as the output. The rest of the operations were consistent with those above used in training the exogenous encoding model.

2.4 Image identification analysis

The exogenous and endogenous encoding models trained with the perception training dataset were used in the identification of images, based on brain activities from the testing dataset during perception and mental imagery, respectively.

In the viewed image identification analysis (perception), we used the test session dataset of 50 new viewed images ($j = 1201, 1202, \dots, 1250$) to estimate the performance of the trained exogenous and endogenous encoding models, respectively. For the exogenous encoding model, we extracted Gabor features of each test image, and then put these low-level visual features into the model to predict the corresponding voxel response. In this way, each voxel-wise encoding model would produce a prediction value for each voxel in the EVC. Each region of the EVC (i.e., V1, V2, and V3) was viewed as a basic unit, and thus the predicted brain activity pattern in V1, V2,

and V3 could be used to compare with the real fMRI signal pattern to estimate the performance of the encoding model. To this end, a Pearson's correlation coefficient between each pair of the two sets (predicted set and real set) of the test images ($n = 50$) was conducted and then a 50-by-50 correlation matrix for each subject was obtained. If the diagonal value of the obtained matrix was the maximum in each column, it suggested that the predicted voxel activity pattern matched very well with the real one for the same image, and then we regarded it as the correct prediction. Finally, we calculated the ratio of the number of correct predictions to the total number of test images (correct predictions/50), which was regarded as predictive accuracy.

We used the same procedure to calculate accuracy with the trained endogenous encoding models. The only difference was input features. We regarded the brain activities of FFA, PPA and LOC as the input features and then predicted brain activity of each region of the EVC. At last, we compared the predicted responses with the real one to get prediction accuracies.

In the imagined image identification analysis (mental imagery), there was no stimulus input during mental imagery, so we adopted the same Gabor features of the viewed images to estimate the performance of the exogenous encoding model. We then compared the predicted activity in the EVC with the real one of the imagery test data. The specific calculation procedures of prediction accuracy were consistent with those described above in the perception experiment. The prediction of the endogenous encoding models in the imagery experiment were conducted in the same way to obtain corresponding accuracies.

2.5 Linear combination with two sources representations

The neural representation the EVC could be explained by two sources, stimuli relevance neural representation and corresponding inner consciousness representation. We, therefore, applied two approaches to explore the co-action of the two source representations in the EVC. Take V1 for example, the trained exogenous and endogenous encoding models could separately provide a predicted value for each voxel, and the values of these voxels in V1 could form a brain activity pattern; thus, we could obtain two types of predicted brain activity patterns corresponding to the types of encoding models. Namely, the linear combination was conducted by calculating the arithmetic mean of the predicted neural activity patterns from two types of voxel-wise encoding models. In addition, the nonlinear combination was obtained by calculating the geometric mean of the predicted neural activity patterns from two types of encoding models. The combined V1 activity pattern was further compared with the real V1 fMRI signal pattern to show the combined effect of the two types of encoding models. With the same approaches, we evaluated the combined effect within V2 and V3, respectively.

2.6 Separation of imagery from perception

Imagery is considered as a weak perception (Pearson & Kosslyn, 2015), which means that mental imagery shares a large part with visual perception. Based on this view, we further refined the relationship of the two mental processes in terms of the neural connectivity between high visual areas and the EVC. We repeated prediction procedures of the image identification analysis of the trained endogenous encoding model, and then regarded the predicted imagery activity pattern in the EVC as a covariate of the predicted perception activity pattern to remove. Finally, the remaining predicted activity pattern in perception was compared with the real fMRI signal pattern in the EVC to make a prediction. To illustrate the effect of this separation, we conducted another two prediction performances to make a comparison. One was the scattered predicted activity pattern of imagery and the other

was a fake activity pattern consisting of random white noise. We regarded them as covariates to remove from the predicted perception pattern and then calculated prediction accuracies, respectively.

2.7 Predictions of Gabor specificity and non-Gabor specificity

To further elaborate the represented information content in the EVC, we labelled the voxels in each region of the EVC with two opposite names. The names were defined by the weight values that obtained from the process of training exogenous encoding model: Gabor specificity which referred to the voxels with useful (non-zero) weight values, non-Gabor specificity which referred to the voxels with zero weight value. Here, the useful or zero weight value denoted whether the voxels could encode the Gabor features or not. Accordingly, we divided the voxels in each region of the EVC into two groups. After that, we conducted the same prediction procedures via the endogenous encoding model for each group in V1, V2, and V3, respectively.

2.8 Code and data accessibility

The code for generating the Gabor wavelet pyramid is freely available online (<https://www.mathworks.cn/matlabcentral/fileexchange/60088-gabor-wavelet-pyramid>). The code for the encoding models is based on the STRFlab toolbox, which is accessible at <http://strflab.berkeley.edu/>. The preprocessed data and relevant code are publicly available at the website of Kamitani Lab (<https://github.com/KamitaniLab/GenericObjectDecoding>) and the unprocessed fMRI data are available from OpenNeuro (<https://openneuro.org/datasets/ds001246>). All calculation processes were run on Matlab R2017a on a Windows operating system.

3 Results

3.1 Hierarchical structure exists in the EVC

3.1.1 Stimuli relevance representation during perception and imagery

During perception, performance of the trained exogenous encoding model showed that Gabor features could remarkably predict brain activity pattern in the EVC. The average accuracies of five subjects were $29\% \pm 0.13$ (Chi-square test, $\chi^2(1, N = 50) = 11.758, p = 0.0006$), $22\% \pm 0.17$ ($\chi^2(1, N = 50) = 8.333, p = 0.0039$), $18\% \pm 0.11$ ($\chi^2(1, N = 50) = 6.4, p = 0.0114$) in V1, V2, and V3, respectively (the green bar in Fig 3.1). Hence, Gabor features could successfully predict the neural activity in the EVC during perception because the predictions obtained significantly higher prediction accuracies than the chance level of 2% (1/50). Interestingly, there was a gradient descent of prediction accuracies in the EVC (i.e., $V1 > V2 > V3$), except for Subject 1 with the highest accuracy for V3 (10%, 5/50; V1 was 6%, V2 was 4%). The predicted accuracies attenuated from V1 to V3, showing a hierarchical organization of the EVC on representation of Gabor features.

During imagery, the average V1 accuracy of five subjects was $1.6\% \pm 0.01$ ($\chi^2(1, N = 50) = 0.02222, p = 0.8815$), V2 was $2\% \pm 0.01$ ($\chi^2(1, N = 50) = 0, p = 1.0000$), and V3 was $3.2\% \pm 0.02$ ($\chi^2(1, N = 50) = 0.1385, p = 0.7098$) (the red bar in Fig 3.1a). These results showed that almost all the prediction accuracies using the exogeneous encoding model with Gabor features in the EVC during mental imagery were around the chance level (2%), poor prediction performances. They were different from findings from prior studies (Cichy et al., 2012; Naselaris et al., 2015), which demonstrated that brain activity in the EVC generated during mental imagery could be explained by Gabor features. The results in the present study were caused largely by the task-dependent factors.

In the imagery experiment, subjects were required to imagine freely as many object images matched with the word cue as possible (Horikawa & Kamitani, 2017), but we conducted the prediction performance based on a specific image that was the same as the procedures in perception. Thus, there was no ground truth that the Gabor features of a specific image could successfully encode the brain activity triggered by the whole of object categories.

To further explore brain activity of imagery in the EVC that were explained by Gabor features, we extracted 1000 voxels in the whole EVC. At first, the values of Pearson's correlation coefficients between the predicted voxel activity and the real one of each subject were sorted in descending order and then we took the top 1000 values (except for the fifth subject who had only around 700 voxels conforming to the condition). Next, we depicted their spatial distribution pattern in the EVC, in which the Pearson's correlation coefficients were regarded as weights of nodes via the BrainNetViewer toolbox (<https://www.nitrc.org/projects/bnv/>). All above calculations were done both with perception and imagery data. We found that overlap distributions of the extracted voxels between perception and mental imagery were very large: 96% of subject 1, 95% of subject 2, 76.3% of subject 3, 84.5% of subject 4, 100% of subject 5 (the number of corresponding voxels in the EVC was 734), which suggested that spatial distributions were similar between the two mental processes for each subject. The median of the average weight of subjects during perception was 0.28 (ranging from -0.09 to 0.81), and the median during imagery was 0.02 (ranging from -0.21 to 0.44), which determined the size of the colour nodes in Fig 3.1b.

3.1.2 Visual cortex internal neural correlation during perception and imagery

From the prediction accuracies of the trained endogenous encoding models, we found that the brain activity patterns in the three high visual areas were able to successfully predict the brain activity pattern in the EVC (Fig 3.1a). Especially, FFA and LOC both obtained higher prediction accuracies more than 40% of V1 during perception (more than 10% during imagery). Critically, the endogenous encoding models were trained only with the perception training data, which could be generalized to predict the brain activity under the mental imagery condition (e.g., the average accuracies of FFA predictions were $12\% \pm 0.08$ ($\chi^2(1, N = 50) = 3.5714, p = 0.0588$) in V1, $21\% \pm 0.09$ ($\chi^2(1, N = 50) = 7.8478, p = 0.0051$) in V2, and $26\% \pm 0.10$ ($\chi^2(1, N = 50) = 10.2860, p = 0.0013$) in V3). Moreover, the tendency of prediction accuracies showed a gradient ascent of the EVC's prediction performance where higher-level visual areas predicted V1 with the lowest accuracy and V3 with the highest one, of which imagery prediction was the same as perception prediction (i.e., $V1 < V2 < V3$). This therefore provided further evidence for existence of a gradient hierarchical structure in the EVC, which was shared by perception and imagery (Fig 3.1a).

Fig 3.1 about here, please

Fig 3.1 Prediction performances with two types of voxel-wise encoding models in the EVC (a) Prediction accuracy with two types of voxel-wise encoding models during perception and imagery. The three different levels of green bar represented the prediction accuracy in the EVC during perception. And another three different levels of red bar represented the prediction accuracy in the EVC during imagery. The first attributes of the abscissa in each subject meant the prediction of Gabor features via the exogenous encoding model, and the other three attributes meant the predictions of three different high visual areas via the endogenous encoding models. The red dotted line referred the chance level (2%). (b) The spatial distribution of 1000 voxels with the order of Pearson's correlation coefficients in the EVC. The upward row showed the voxels distribution during perception of five subjects and the row beneath was the voxels distribution during imagery. And the size of the colorful dots was decided by the weight (Pearson's correlation coefficients). The red dots were voxels in V1; the orange dots were voxels

in V2; the green dots were voxels in V3. The voxels overlap proportions between perception and imagery for five subjects were: 96 %, 95 %, 76.3 %, 84.5 %, and 100%.

3.2 Dual representations from Gabor features and cortical internal correlations

We explored the combined effect of the exogenous and endogenous encoding models for explaining brain activity in the EVC during perception and mental imagery. For perception, results showed that linear combination of the two types of encoding models could effectively improve the prediction power, the average accuracies were increased by more than 15%. But the average accuracies were mostly reduced to lower than 10% under the nonlinear combination condition (the green lines in Fig 3.2a). There was not any evident linear or nonlinear combination effect in the mental imagery experiment (the red lines in Fig 3.2a).

When removing the neural activity of mental imagery from that of perception with the help of the two types of encoding models, we found that the average prediction accuracies fell to a range from 10% to 50%, which were still more significant than chance level (2%). However, after we removed the scattered imagery pattern and random white noise from perception part (Fig 3.2b), the prediction accuracies of the three high visual areas predicting the EVC were close to the original prediction performances of inner cortical connectivity. So, these observations showed that imagery shared plenty of information representation with perception in a fashion of visual neural connectivity.

Fig 3.2 about here, please

Fig 3.2 Neural representation prediction performance during perception and mental imagery (a) The combination effect of the exogenous encoding model and the endogenous encoding model via linear combination and nonlinear combination. The left three attributes of the abscissa in each subject line chart represented the linear combination and the right counterpart represented the nonlinear combination, respectively. The green line represented the prediction performance during perception in the EVC and the red one represented the counterpart during imagery in the EVC. (b) The performance in the separation of the predicted imagery pattern from the predicted perception pattern. The blue line represented the performance of FFA predicting the EVC; the green line represented the performance of LOC predicting the EVC; the red line represented the performance of PPA predicting the EVC. We conducted four comparisons: *original*, the prediction of the original high visual areas predicted the EVC; *imagery*, the prediction of high visual areas removed the predicted imagery pattern by a way of covariate; *noise*, the prediction of high visual areas removed the random white noise in a way of covariate; *scatter*, the prediction of high visual areas removed the scattered imagery pattern in a way of covariate.

3.3 Shared and independent representation in the EVC

To further elaborate the information content of the representation of voxel activity in the EVC, we divided these voxels into two groups with Gabor specificity and with non-Gabor specificity (for more details, see the Methods). We then tested performance of the exogenous and endogenous encoding models specially related to the two groups of voxels.

Interestingly, during the procedure of labelling the voxels in the EVC, we found that there was a stable topological distribution of the rates between Gabor specificity and non-Gabor specificity. Because whether the total number of selected voxels was 1000, 800, 600, 400, or 200, the rate of the two labelled groups were almost to 1:2 (see the Supplement Fig 3.4).

Comparing the prediction performances of the two groups, we found that accuracies obtained via the endogenous encoding model were almost equal in each area of the EVC, both in the perception and the mental

imagery conditions. Besides, the prediction performances of the two groups were similar to the original performance conducted by the endogenous encoding model. Here, to show the results clearly, we selected the results of Subject 3 in Fig 3.3 (for other subjects' results, see Supplemental Fig 3.3.1, 3.3.2, 3.3.3, and 3.3.4). Likewise, we explored the combined effect of the two types of the encoding models related to the two groups of voxels, respectively. The results showed that there were similar prediction performances between the two groups of voxels within a hierarchical fashion, which further proved the parallel representation contents in the EVC.

Fig 3.3 about here, please

Fig 3.3 Shared and independent representation in the EVC. Here, we selected the results of subject 3 to show three high visual areas prediction in the Gabor specificity and non-Gabor specificity voxel groups via the endogenous encoding models. *GP*, the Gabor specificity prediction during perception; *NP*, the non-Gabor specificity prediction during perception; *OP*, the original high visual areas predictions during perception; the red one that **I** represented the same prediction during imagery. The bar in the right lower part showed the superimposed effect of three high visual areas, which compared the original superimposed effect with the combination of high visual areas prediction based on the Gabor non-specificity voxels and Gabor features prediction in the EVC.

4 Discussion

Using the trained exogenous and endogenous encoding models, we quantified the elaboration of a visual neural “signature” and its connectivity to visual features, which would provide an avenue to explore how internally generated representations and externally stimulated representations interact within our brain. Especially, the application of voxel-wise encoding models allowed us to understand the encoded information content related to each voxel within the EVC. Accordingly, we found the existence of a shared and independent neural representation between perception and visual imagery, which helps us to understand the neural subtract between the two similar mental processes.

Prior to this current work, previous research has applied MVPC to investigate the relationship between perception and imagery (Albers et al., 2013; Harrison & Tong, 2009; S. H. Lee et al., 2012; Tong, 2013) and to show the complexity of interaction within the visual cortex. Because MVPC are limited in their ability to decompose brain activity patterns into distinct sources of variation, how to measure the represent content is still largely unclear. This study used voxel-wise encoding models (Naselaris et al., 2015; Thirion et al., 2006) to refine the shared neural representation between visual perception and mental imagery. The result provided excellent prediction performance of the internal neural correlation with high visual areas (especially in FFA and LOC) both in perception and imagery conditions. Besides, we found a linear combination effect of interaction between visual input and internal neural association between high level visual areas and EVC, which was consistent with a previous finding (Zhang et al., 2014). Furthermore, we removed the imagery activity, in a fashion of covariate, from the internal activation part of perception to make an in-depth exploration of the representation mechanism of perception. The result showed that internal neural association in the ventral visual pathway during perception could be partly explained by that of mental imagery. More importantly, we could easily divide the voxels in the EVC into the Gabor specificity group and non-Gabor specificity group based on the method of voxel-wise encoding model. These results indicated that the shared and independent.

Combining with the voxel-wise encoding models, we found perceptual representation in a fashion of linear rather than nonlinear between Gabor features and cortical internal connectivity in the EVC, which was different from the modular representation of the brain (Op de Beeck et al., 2008). In addition, the majority of representation

in the EVC could be explained by the internal neural connectivity, top-down expectation, which was consistent with assertions that the cortical feedback is an essential part of brain processing (Morgan, 2018). The event of the linear representation could be regarded as independent representations of the two sources of information, i.e., stimulus relevant information and internally generated information, which might provide a new evidence for the “conscious copy” model of working memory (Jacobs & Silvanto, 2015). The same prediction performances between Gabor specificity and non-Gabor specificity voxels via endogenous encoding model provided effective evidence for independent representation. The analysis based on voxel-wise level and associated the EVC with high level visual areas provides us a viewpoint to refine the shared neural representation mechanism between visual perception and mental imagery.

Additionally, Albert and colleagues found empirical evidence to support the common internal representation between imagery and working memory (Albers et al., 2013). In history, there has been a longstanding debate about whether the content of working memory can be regarded as its conscious experience (Atkinson & Shiffrin, 1968; Baars, 2002; Baars & Franklin, 2003; Cowan, 1988; Oberauer, 2002, 2009). The classical working memory model proposed that all the contents of working memory could be accessible to consciousness (A. Baddeley, 1992; A. D. Baddeley & Hitch, 1974). However, subsequent studies showed that the contents of working memory accessible to conscious awareness was selective (Baars, 1993) and that attention had a role in allowing some part of the contents to reach consciousness (Baars, 2002; Cowan, 1988; Oberauer, 2002, 2009). Further research revealed that there was not an identical representation between working memory contents and corresponding consciousness, and that there may be a kind of independent and parallel representation between subjective conscious awareness and the memory trace. First, studies showed that working memory performance accuracy was independent of visual working memory contents accessible to conscious awareness, as shown by performance of the vividness of mental imagery and task confidence (Hassin et al., 2009; Soto et al., 2011). Second, changing the memory trace could influence the performance of working memory, but not the corresponding consciousness (Bona et al., 2013; Bona & Silvanto, 2014). Third, the contents of working memory whether accessing to consciousness or not, has distinct influences on visual input processing (Craver-Lemley & Reeves, 1992; Pan et al., 2014; Perky, 1910). Specifically, if visual input matches with the content of working memory, it can get access to consciousness easily and the mismatched part would be restrained or discarded. Empirical evidence demonstrated that the representation of working memory was internally unconscious, while the corresponding conscious awareness may create a new representation accessible to consciousness, whose function was different from the one of working memory.

Mental imagery, an internal conscious percept process (Kosslyn, 1996), can be cast as the conscious experience of working memory content (Jacobs & Silvanto, 2015) and may directly address the separate representations between external stimulus and internal consciousness. Herein, we confirmed that mental imagery shared an underlying neural representation with visual perception in the cortical connectivity, which might be considered as the introspection of working memory content. Strikingly, the prediction accuracies between Gabor specificity and non-Gabor specificity of imagery were almost equal, the same as the ones of perception. This may imply that imagery, internally generated consciousness, not only creates corresponding conscious experience but also internally represents the relevance of an external stimulus. Jacobs and Silvanto (2015) has proposed that the conscious awareness of working memory does not engage in changing the original memory trace, but rather creates a new representation that is co-existing with the original.

Mental imagery, regarded as an ability to generate visual experience without triggering by corresponding sensory stimulations (Kosslyn, 1996), provides an important means to study the complex topic of brain representation. Many neuroscience studies tried to demonstrate similarity and discrepancy between perception and imagery (Cichy et al., 2012; S. H. Lee et al., 2012; Schaefer et al., 2013), and have furnished much evidence for the neural representation of the world. Especially, an increasing number of studies have confirmed that neurons in EVC do not process in a fashion of feature detectors (Kayser et al., 2004; T. S. Lee & Nguyen, 2001; Sugita, 1999). With ground truth measurements, our results demonstrated that neurons in the EVC were active both during imagery and perception, and activation of neurons in the EVC could be effectively predicted by the high level visual neural activity. Moreover, according to our results, which showed the same prediction performance between Gabor specificity and non-Gabor specificity of labelled voxels in the EVC, suggesting that the brain might simultaneously generate an internal copy representation of the stimulation when making responses to perturbations of the stimulation from the real world.

There were several issues which should be considered in future work. First, application of the voxel-wise encoding models gave us an opportunity to explore the shared and independent representation mechanism between visual perception and mental imagery, and thus, in the future development of technique and methods, the questions of dynamic neural representation (Dijkstra et al., 2018) and the visualization of the independent representation need to be addressed. Second, the clear and elaborate representation mechanism offers an alternative possibility concerning applying and updating the artificial intelligence system with consciousness. Third, the shared and independent neural representation mechanism, in a more active insight, may lead us towards a deeper understanding and explanation of abnormal phenomenon of brain information processing, such as found in hallucination and schizophrenia. These results could then in turn lead to more effective and reasonable treatments.

In conclusion, the present study applied the voxel-wise encoding models with representing two sources of information, stimuli relevance and the internal neural connectivity. We firstly showed that during perception there is a linear combination of Gabor features and internal cortical connectivity, which was shared by mental imagery. We further demonstrated that the two sources of information were represented independently in the hierarchy of the visual system during visual perception and mental imagery. These observations provided new insights into the underlying neural subtract between perception and imagery.

References

- Aitken, F., Turner, G., & Kok, P. (2020). Prior Expectations of Motion Direction Modulate Early Sensory Processing. *J Neurosci*, 40(33), 6389-6397. <https://doi.org/10.1523/JNEUROSCI.0537-20.2020>
- Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology*, 23(15), 1427-1431. <https://doi.org/http://doi.org/10.1016/j.cub.2013.05.065>
- Andersson, P., Ragni, F., & Lingnau, A. (2019). Visual imagery during real-time fMRI neurofeedback from occipital and superior parietal cortex. *Neuroimage*, 200, 332-343. <https://doi.org/10.1016/j.neuroimage.2019.06.057>
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of learning and motivation*, 2, 89-195. [https://doi.org/10.1016/S0079-7421\(08\)60422-3](https://doi.org/10.1016/S0079-7421(08)60422-3)
- Baars, B. J. (1993). *A cognitive theory of consciousness*: Cambridge University Press.
- Baars, B. J. (2002). The conscious access hypothesis: origins and recent evidence. *Trends Cogn Sci*, 6(1), 47-52. [https://doi.org/10.1016/s1364-6613\(00\)01819-2](https://doi.org/10.1016/s1364-6613(00)01819-2)
- Baars, B. J., & Franklin, S. (2003). How conscious experience and working memory interact. *Trends in cognitive sciences*, 7(4), 166-172. [https://doi.org/10.1016/S1364-6613\(03\)00056-1](https://doi.org/10.1016/S1364-6613(03)00056-1)
- Baddeley, A. (1992). Working memory. *Science*, 255(5044), 556-559. <https://doi.org/10.1126/science.1736359>
- Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation*, 8, 47-89. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- Berger, C. C., & Ehrsson, H. H. (2013). Mental imagery changes multisensory perception. *Curr Biol*, 23(14), 1367-1372. <https://doi.org/10.1016/j.cub.2013.06.012>
- Bona, S., Cattaneo, Z., Vecchi, T., Soto, D., & Silvanto, J. (2013). Metacognition of visual short-term memory: dissociation between objective and subjective components of VSTM. *Frontiers in psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00062>
- Bona, S., & Silvanto, J. (2014). Accuracy and confidence of visual short-term memory do not go hand-in-hand: behavioral and neural dissociations. *PLoS One*, 9(3), e90808. <https://doi.org/10.1371/journal.pone.0090808>
- Borst, G., & Kosslyn, S. M. (2008). Visual mental imagery and visual perception: structural equivalence revealed by scanning processes. *Mem Cognit*, 36(4), 849-862. <https://doi.org/10.3758/mc.36.4.849>
- Butter, C. M., Kosslyn, S., MijovicPrelec, D., & Riffle, A. (1997). Field-specific deficits in visual imagery following hemianopia due to unilateral occipital infarcts. *Brain*, 120, 217-228. <https://doi.org/DOI 10.1093/brain/120.2.217>
- Cichy, R. M., Heinze, J., & Haynes, J. D. (2012). Imagery and perception share cortical representations of content and location. *Cereb Cortex*, 22(2), 372-380. <https://doi.org/10.1093/cercor/bhr106>
- Cowan, N. (1988). Evolving Conceptions of Memory Storage, Selective Attention, and Their Mutual Constraints within the Human Information-Processing System. *Psychological bulletin*, 104(2), 163-191. <https://doi.org/10.1037/0033-2909.104.2.163>
- Craver-Lemley, C., & Reeves, A. (1992). How visual imagery interferes with vision. *Psychological Review*, 99(4), 633. <https://doi.org/10.1037/0033-295x.99.4.633>
- Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J Opt Soc Am A*, 2(7), 1160-1169. <https://doi.org/10.1364/josaa.2.001160>
- Deyoe, E. A., Carman, G. J., Bandetini, P., Glickman, S., Wieser, J., Cox, R., . . . Neitz, J. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 93(6), 2382-2386. <https://doi.org/10.1073/pnas.93.6.2382>
- Diekhof, E. K., Kipshagen, H. E., Falkai, P., Dechent, P., Baudewig, J., & Gruber, O. (2011). The power of imagination--how anticipatory mental imagery alters perceptual processing of fearful facial expressions. *Neuroimage*, 54(2), 1703-1714. <https://doi.org/10.1016/j.neuroimage.2010.08.034>

- Dijkstra, N., Bosch, S. E., & van Gerven, M. A. J. (2019). Shared Neural Mechanisms of Visual Perception and Imagery. *Trends Cogn Sci*, 23(5), 423-434. <https://doi.org/10.1016/j.tics.2019.02.004>
- Dijkstra, N., Mostert, P., Lange, F. P., Bosch, S., & van Gerven, M. A. (2018). Differential temporal dynamics during visual imagery and perception. *Elife*, 7. <https://doi.org/10.7554/eLife.33904>
- Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E. J., & Shadlen, M. N. (1994). fMRI of human visual cortex. *Nature*, 369(6481), 525. <https://doi.org/10.1038/369525a0>
- Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, 458(7238), 632-635. <https://doi.org/10.1038/nature07832>
- Hassin, R. R., Bargh, J. A., Engell, A. D., & McCulloch, K. C. (2009). Implicit working memory. *Conscious Cogn*, 18(3), 665-678. <https://doi.org/10.1016/j.concog.2009.04.003>
- Haynes, J. D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci*, 8(5), 686-691. <https://doi.org/10.1038/nn1445>
- Horikawa, T., & Kamitani, Y. (2017). Generic decoding of seen and imagined objects using hierarchical visual features. *Nature communications*, 8, 15037.
- Hsieh, P. J., Vul, E., & Kanwisher, N. (2010). Recognition alters the spatial pattern of FMRI activation in early retinotopic cortex. *J Neurophysiol*, 103(3), 1501-1507. <https://doi.org/10.1152/jn.00812.2009>
- Ishai, A., & Sagi, D. (1995). Common mechanisms of visual imagery and perception. *Science*, 268(5218), 1772-1774. <https://doi.org/10.1126/science.7792605>
- Jacobs, C., & Silvano, J. (2015). How is working memory content consciously experienced? The ‘conscious copy’ model of WM introspection. *Neuroscience & Biobehavioral Reviews*, 55, 510-519. <https://doi.org/10.1016/j.neubiorev.2015.06.003>
- James, W. (1890). *The principles of psychology* (Vol. 1): Macmillan London.
- Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J Neurophysiol*, 58(6), 1233-1258. <https://doi.org/10.1152/jn.1987.58.6.1233>
- Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nat Neurosci*, 8(5), 679-685. <https://doi.org/10.1038/nn1444>
- Kastner, S., De Weerd, P., Elizondo, M., Desimone, R., & Ungerleider, L. (1998). Attentional response modulation in human extrastriate cortex as revealed by functional MRI. *Neuroimage*, 7(4 PART II), S44. <https://doi.org/10.1073/pnas.96.4.1663>
- Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, 452(7185), 352-355. <https://doi.org/10.1038/nature06713>
- Kayser, C., Kording, K. P., & Konig, P. (2004). Processing of complex stimuli and natural scenes in the visual cortex. *Curr Opin Neurobiol*, 14(4), 468-473. <https://doi.org/10.1016/j.conb.2004.06.002>
- Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron*, 75(2), 265-270. <https://doi.org/10.1016/j.neuron.2012.04.034>
- Kosslyn, S. M. (1996). *Image and brain: The resolution of the imagery debate*: MIT press.
- Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological bulletin*, 129(5), 723. <https://doi.org/10.1037/0033-2909.129.5.723>
- Lee, S. H., Kravitz, D. J., & Baker, C. I. (2012). Disentangling visual imagery and perception of real-world objects. *Neuroimage*, 59(4), 4064-4073. <https://doi.org/10.1016/j.neuroimage.2011.10.055>
- Lee, T. S. (1996). Image representation using 2D gabor wavelets. *Ieee Transactions on Pattern Analysis and Machine Intelligence*, 18(10), 959-971. <https://doi.org/10.1109/34.541406>
- Lee, T. S., & Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 98(4), 1907-1911. <https://doi.org/10.1073/pnas.031579998>

- Maier, M., Frömer, R., Rost, J., Sommer, W., & Rahman, R. A. (2020). Time course and shared neurocognitive mechanisms of mental imagery and visual perception. *bioRxiv*. <https://doi.org/10.1101/2020.01.14.905885>
- Morgan, A. T. (2018). *Encoding and decoding of cortical feedback to human early visual cortex*. University of Glasgow,
- Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., & Gallant, J. L. (2015). A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *Neuroimage*, 105, 215-228. <https://doi.org/10.1016/j.neuroimage.2014.10.018>
- Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron*, 63(6), 902-915. <https://doi.org/10.1016/j.neuron.2009.09.006>
- O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of cognitive neuroscience*, 12(6), 1013-1023. <https://doi.org/10.1162/08989290051137549>
- Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *J Exp Psychol Learn Mem Cogn*, 28(3), 411-421. <https://doi.org/10.1037/0278-7393.28.3.411>
- Oberauer, K. (2009). Interference between storage and processing in working memory: Feature overwriting, not similarity-based competition. *Memory & cognition*, 37(3), 346-357. <https://doi.org/10.3758/MC.37.3.346>
- Olman, C. A., Ugurbil, K., Schrater, P., & Kersten, D. (2004). BOLD fMRI and psychophysical measurements of contrast response to broadband images. *Vision Res*, 44(7), 669-683. <https://doi.org/10.1016/j.visres.2003.10.022>
- Op de Beeck, H. P., Haushofer, J., & Kanwisher, N. G. (2008). Interpreting fMRI data: maps, modules and dimensions. *Nat Rev Neurosci*, 9(2), 123-135. <https://doi.org/10.1038/nrn2314>
- Pan, Y., Lin, B., Zhao, Y., & Soto, D. (2014). Working memory biasing of visual perception without awareness. *Attention, Perception, & Psychophysics*, 76(7), 2051-2062. <https://doi.org/10.3758/s13414-013-0566-2>
- Pearson, J. (2019). The human imagination: the cognitive neuroscience of visual mental imagery. *Nat Rev Neurosci*, 20(10), 624-634. <https://doi.org/10.1038/s41583-019-0202-9>
- Pearson, J., Clifford, C. W., & Tong, F. (2008). The functional impact of mental imagery on conscious perception. *Curr Biol*, 18(13), 982-986. <https://doi.org/10.1016/j.cub.2008.05.048>
- Pearson, J., & Kosslyn, S. M. (2015). The heterogeneity of mental representation: Ending the imagery debate. *Proc Natl Acad Sci U S A*, 112(33), 10089-10092. <https://doi.org/10.1073/pnas.1504933112>
- Perky, C. W. (1910). An experimental study of imagination. *The American Journal of Psychology*, 21(3), 422-452. <https://doi.org/10.2307/1413350>
- Rainer, G., Augath, M., Trinath, T., & Logothetis, N. K. (2001). Nonmonotonic noise tuning of BOLD fMRI signal to natural images in the visual cortex of the anesthetized monkey. *Curr Biol*, 11(11), 846-854. [https://doi.org/10.1016/s0960-9822\(01\)00242-1](https://doi.org/10.1016/s0960-9822(01)00242-1)
- Reddy, L., Tsuchiya, N., & Serre, T. (2010). Reading the mind's eye: decoding category information during mental imagery. *Neuroimage*, 50(2), 818-825. <https://doi.org/10.1016/j.neuroimage.2009.11.084>
- Ringach, D. L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *J Neurophysiol*, 88(1), 455-463. <https://doi.org/10.1152/jn.2002.88.1.455>
- Sasaki, Y., Rajimehr, R., Kim, B. W., Ekstrom, L. B., Vanduffel, W., & Tootell, R. B. (2006). The radial bias: a different slant on visual orientation sensitivity in human and nonhuman primates. *Neuron*, 51(5), 661-670. <https://doi.org/10.1016/j.neuron.2006.07.021>
- Schaefer, R. S., Desain, P., & Farquhar, J. (2013). Shared processing of perception and imagery of music in decomposed EEG. *Neuroimage*, 70, 317-326. <https://doi.org/10.1016/j.neuroimage.2012.12.064>
- Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., . . . Tootell, R. B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, 268(5212), 889-893. <https://doi.org/10.1126/science.7754376>
- Singh, K. D., Smith, A., & Greenlee, M. W. (2000). Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. *Neuroimage*, 12(5), 550-564. <https://doi.org/10.1006/nimg.2000.0642>

- Sirigu, A., & Duhamel, J. R. (2001). Motor and visual imagery as two complementary but neurally dissociable mental processes. *Journal of cognitive neuroscience*, 13(7), 910-919. <https://doi.org/10.1162/089892901753165827>
- Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *J Neurosci*, 32(25), 8443-8453. <https://doi.org/10.1523/JNEUROSCI.5069-11.2012>
- Soto, D., Mäntylä, T., & Silvanto, J. (2011). Working memory without consciousness. *Current Biology*, 21(22), R912-R913. <https://doi.org/10.1016/j.cub.2011.09.049>
- Stokes, M., Thompson, R., Cusack, R., & Duncan, J. (2009). Top-down activation of shape-specific population codes in visual cortex during mental imagery. *J Neurosci*, 29(5), 1565-1572. <https://doi.org/10.1523/JNEUROSCI.4657-08.2009>
- Sugita, Y. (1999). Grouping of image fragments in primary visual cortex. *Nature*, 401(6750), 269-272. <https://doi.org/10.1038/45785>
- Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J. B., Lebihan, D., & Dehaene, S. (2006). Inverse retinotopy: inferring the visual content of images from brain activation patterns. *Neuroimage*, 33(4), 1104-1116. <https://doi.org/10.1016/j.neuroimage.2006.06.062>
- Tong, F. (2013). Imagery and visual working memory: one and the same? *Trends Cogn Sci*, 17(10), 489-490. <https://doi.org/10.1016/j.tics.2013.08.005>
- Tugnait, J. K. (1994). Estimation of linear parametric models of nonGaussian discrete random fields with application to texture synthesis. *IEEE Trans Image Process*, 3(2), 109-127. <https://doi.org/10.1109/83.277894>
- Xie, S., Kaiser, D., & Cichy, R. M. (2020). Visual Imagery and Perception Share Neural Representations in the Alpha Frequency Band. *Curr Biol*, 30(13), 2621-2627 e2625. <https://doi.org/10.1016/j.cub.2020.04.074>
- Zhang, D., Wen, X., Liang, B., Liu, B., Liu, M., & Huang, R. (2014). Neural associations of the early retinotopic cortex with the lateral occipital complex during visual perception. *PLoS One*, 9(9), e108557. <https://doi.org/10.1371/journal.pone.0108557>

Figures

Figure 1

Neural encoding through Gabor wavelet pyramid and brain areas' connectivity model. When a person was seeing or imaging object image, the brain visual areas were processing neural information. We used the Gabor wavelet pyramid to filter images to get low visual features and then combined the voxel activity pattern in the early visual cortex (V1, V2, and V3) to train voxel-wise Gabor-encoding models. At the same time, we trained the analogical voxel-wise encoding models with neural connectivity between the early visual areas and the high visual areas (FFA, LOC, and PPA). Moreover, we further investigated the neural representation during perception and imagery according to the multi-encoding models we constructed.

exogenous encoding model, and the other three attributes meant the predictions of three different high visual areas via the endogenous encoding models. The red dotted line referred the chance level (2%). (b) The spatial distribution of 1000 voxels with the order of Pearson's correlation coefficients in the EVC. The upward row showed the voxels distribution during perception of five subjects and the row beneath was the voxels distribution during imagery. And the size of the colorful dots was decided by the weight (Pearson's correlation coefficients). The red dots were voxels in V1; the orange dots were voxels in V2; the green dots were voxels in V3. The voxels overlap proportions between perception and imagery for five subjects were: 96 %, 95 %, 76.3 %, 84.5 %, and 100%.

Figure 3

Neural representation prediction performance during perception and mental imagery (a) The combination effect of the exogenous encoding model and the endogenous encoding model via linear combination and

nonlinear combination. The left three attributes of the abscissa in each subject line chart represented the linear combination and the right counterpart represented the nonlinear combination, respectively. The green line represented the prediction performance during perception in the EVC and the red one represented the counterpart during imagery in the EVC. (b) The performance in the separation of the predicted imagery pattern from the predicted perception pattern. The blue line represented the performance of FFA predicting the EVC; the green line represented the performance of LOC predicting the EVC; the red line represented the performance of PPA predicting the EVC. We conducted four comparisons: original, the prediction of the original high visual areas predicted the EVC; imagery, the prediction of high visual areas removed the predicted imagery pattern by a way of covariate; noise, the prediction of high visual areas removed the random white noise in a way of covariate; scatter, the prediction of high visual areas removed the scattered imagery pattern in a way of covariate.

Figure 4

Shared and independent representation in the EVC. Here, we selected the results of subject 3 to show three high visual areas prediction in the Gabor specificity and non-Gabor specificity voxel groups via the endogenous encoding models. GP, the Gabor specificity prediction during perception; NP, the non-Gabor

specificity prediction during perception; OP, the original high visual areas predictions during perception; the red one that I represented the same prediction during imagery. The bar in the right lower part showed the superimposed effect of three high visual areas, which compared the original superimposed effect with the combination of high visual areas prediction based on the Gabor non-specificity voxels and Gabor features prediction in the EVC.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigures.pdf](#)