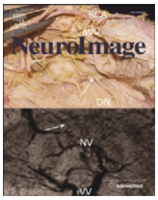
NeuroImage 56 (2011) 1540–1545

Contents lists available at ScienceDirect

NeuroImage

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



Imagery for shapes activates position-invariant representations in human visual cortex

Mark Stokes *, Ana Saraiva, Gustavo Rohenkohl, Anna C. Nobre

Oxford University, Department of Experimental Psychology, South Parks Road, Oxford, OX1 3UD, UK

ARTICLE INFO

Article history:
Received 25 October 2010
Revised 23 February 2011
Accepted 26 February 2011
Available online 3 March 2011

Keywords:
Visual imagery
fMRI
MVPA
Position-invariance
Top-down control

ABSTRACT

In this study, we show that top-down control mechanisms engaged during visual imagery of simple shapes (letters X and O) can selectively activate position-invariant perceptual codes in visual areas specialised for shape processing, including lateral occipital complex (LOC). First, we used multivoxel pattern analysis (MVPA) to identify visual cortical areas that code for shape within a position-invariant reference frame. Next, we examined the similarity between these high-level visual codes and patterns elicited while participants imagined the corresponding stimulus at central fixation. Our results demonstrate that imagery engages object-centred codes in higher-level visual areas. More generally, our results also demonstrate that top-down control mechanisms are able to generate highly specific patterns of visual activity in the absence of corresponding sensory input. We argue that a general model of top-down control must account for dynamic modulation of functional connectivity between high-level control centres and overlapping neural codes in visual cortex.

Introduction

Visual imagery illustrates how top-down control mechanisms can modulate visual cortex independently of direct sensory input (Kosslyn et al., 2001). For example, to conjure up an image of a face in the mind's eye increases neural activity in cortical areas that are most important for processing face stimuli (Ishai et al., 2000). Presumably, top-down visual activity furnishes imagery with the perceptual detail that characterises the mental picture (Mechelli et al., 2004). Similar feedback dynamics may also underlie other high-level visual processes, such as VSTM (Pasternak and Greenlee, 2005) and attention (Desimone and Duncan, 1995).

Distributed neuronal populations coding different visual items can be highly overlapping within brain regions (deCharms and Zador, 2000), therefore top-down control signals need to be precise. For example, if we want to imagine Marilyn Monroe's face, we must be careful to activate only neurons that represent Monroe, and not the neural population that codes for the face of Margret Thatcher. However, neural types distinguishing such high-level visual characteristics are not topographically organised (Kriegeskorte et al., 2007; Rolls and Tovee, 1995; Young and Yamane, 1992). Rather, the neural population coding Monroe is most likely to be spatially overlapping and intermingled with the population coding Thatcher (Rotshtein et al., 2005) or indeed population codes for Halle Berry or Jennifer Aniston (Quiroga et al., 2005), thereby complicating the task for top-down mechanisms to access just the right representation.

Recently, we demonstrated that top-down mechanisms can activate precise perceptual codes in visual cortex, even when these are distributed across overlapping neural populations that code functionally distinct representations (Stokes et al., 2009a). We further speculated that these perceptual codes represent high-level characteristics of visual imagery, that is, they are insensitive to low-level attributes such as retinal position. Although this hypothesis accords with previous evidence for position-invariant perceptual representations in human lateral occipital complex (LOC) (Grill-Spector et al., 1999), and monkey inferotemporal cortex (e.g., Logothetis et al., 1995), top-down activation of such position-invariant representations has not yet been demonstrated in human visual cortex.

In this experiment, we extend upon our previous study to determine more accurately the coding properties of neurons that are activated in a top-down manner during imagery. Previously, we used a cross-comparison approach to MVPA to compare activation patterns in visual cortex elicited by imagery and perception (Stokes et al., 2009a). However, because we instructed participants to imagine stimuli with the same low-level characteristics to the stimuli presented during perception, we were not able to determine the extent to which the neural similarity we observed in visual cortex reflected similarity between low-level perceptual features, such as retinotopically-specific activation profiles, or higher-level features, such as position-invariant shape codes. Because our previous MVPA results were specific to higher-level visual areas, including LOC (Spiridon et al., 2006), we favoured the latter interpretation as the

---

* Corresponding author. Fax: +44 1865 310 447.
E-mail address: mark.stokes@sjc.ox.ac.uk (M. Stokes).

1053-8119/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.02.071

most likely account. Nevertheless, we could not rule out contribution from retinotopically-specific activation patterns. In early visual cortex, the contents of imagery can be decoded from pattern analyses that characterise retinotopic maps (Thirion et al., 2006). Although retinotopy is more well defined in earlier visual areas, careful analysis also reveals spatiotopic organisation in high-level visual cortex traditionally associated with position-invariant coding for shapes (Cichy et al., 2011; Kravitz et al., 2010; Sayres and Grill-Spector, 2008). Therefore, it remains possible that neural similarity between perception and imagery observed in our previous study could have been due to the similarity in retinal-coordinates associated with the corresponding representations, rather than to more abstract representations of shape.

Here we explicitly tested for top-down activation of position-invariant perceptual codes during visual imagery. As in our previous study, participants imagined either the letter X or O at the centre of the screen. Patterns of visual activity elicited during these imagery trials were then compared to corresponding patterns for X or O driven by direct stimulation observed during a separate pattern localiser task. However, unlike in our previous study, the pattern localiser was extended to include stimuli presented at random positions within the visual field. We also extended the duration of visual imagery to explore the neural time-course of pattern-similar activity. Exploiting these changes, we can now confirm that top-down control mechanisms access neurons in visual cortex that code the contents of imagery within a spatially invariant reference frame.

Materials and methods

Participants and procedure

Participants (12; 5 females; aged between 20 and 33; all right-handed) were screened for MR contraindications, and all had normal, or correct-to-normal vision. The experimental protocols were approved by Hertfordshire Local Research Ethics Committee, and all participants provided written informed consent. One participant was excluded due to a self-reported inability to generate the appropriate visual images.

The experiment was conducted over two consecutive scanning runs (2×~28 minutes), each consisting of 4 blocks of the imagery task (4×~290 seconds) and 4 blocks of the pattern localiser task (4×128 seconds). Each scanning run began with a block of pattern localiser, and then alternated between the imagery and localiser task. Visual stimuli were presented against a black background, rear-projected via a Hitachi CP-X80 LCD projector at 1024×768 resolution and 60 Hz refresh rate. Presentation v10 (Neurobehavioral Systems, Albany, NY) was used for all aspects of experimental control, including stimulus presentation, recording behavioural responses (via an MR-compatible button-box during fMRI) and synchronizing experimental timing with scanner-pulse timing. Participants were instructed to minimise head movements during scanning, and to maintain central fixation during both imagery and localiser tasks. Eye position was monitored during scanning by the experimenters using an MR-compatible eye-tracking system (ASL 504 LRO Eye Tracking System, Applied Sciences Laboratory, Bedford, MA), and participants performed a practice session prior to scanning to ensure that they could perform the task as instructed.

Imagery task

See Fig. 2A for the task schematic. Each experimental trial commenced with an auditory tone (200 ms; high or low frequency, 500 Hz and 220 Hz respectively) that cued participants to imagine either the letter X or O. Prior to scanning, participants were instructed to imagine the appropriate letter as vividly as possible for the duration of the trial, randomised between 14 and 16 seconds. As a guide, participants were also instructed to imagine the letter at the centre of the screen, and as similarly as possible to the letters used in the pattern localiser (upper case, courier font, 1.8° size). At the end of each trial, participants were asked to indicate with a button-press response whether they were able to generate the appropriate image. Only data associated with positive imagery ratings were included for MVPA (79% of trials). The next trial began after a 4-second inter-trial interval, and there were 16 trials per block.

Half the participants were initially instructed to imagine X in response to the high frequency tone, and to imagine O in response to the low tone, whilst the other half were provided the converse instruction. Importantly, after the first scanning run (i.e., half way through the experiment), all participants were instructed to reverse the initial association between the cue stimulus and imagery condition. Because cue-imagery mapping was counter-balanced within participants, neural differences between imagery states that are consistent across the scanning session cannot be attributed to differences associated with the auditory stimulus used to direct the contents of visual imagery.

Pattern localiser

Participants also performed a perceptual task to allow us to define the visual cortical areas that code the critical visual stimuli (letters X and O) in a position-invariant reference frame (see Fig. 1A). Each block of the localiser task consisted of 4 mini-blocks of 16 seconds, each separated by a 16-second baseline rest period. During X_{centre} (or O_{centre}) blocks, the letter X (or O) was presented for 500 ms at the centre of the screen for 16 seconds as a rate of 1 Hz (500 ms on, 500 ms off). During $X_{\text{off-centre}}$ (or $O_{\text{off-centre}}$) blocks, the letter X (or O) was presented according to the same timing parameters, but the position was randomised to occur within a 10° radius of central fixation. To ensure that letter stimuli were attended to, participants were instructed to respond with a button press to occasional letters (5% probability) that were presented at a relatively lower contrast.

fMRI acquisition

Functional data (T2*-weighted echo planar images) were collected using a Siemens 3T Tim Trio scanner with a 12-channel head coil. Each volume consisted of 32 contiguous slices (3 mm thick) acquired in descending order (TR = 2 seconds, TE = 30 ms). The near-axial acquisition matrix (FOV = 192 mm × 192 mm, 64 × 64 voxels per slice; 3 × 3 × 3 mm voxel resolution) was angled to capture the whole brain. High resolution structural images (1 × 1 × 1 mm) were acquired for each participant using a T1-weighted MP-RAGE sequence (FOV = 192 × 192 mm, 176 sagittal slices) to enable co-registration and grey matter segmentation. In one participant, we collected additional functional data to estimate eccentricity maps in visual cortex using standard retinotopic mapping procedures (Engel et al., 1997).

Multi-voxel pattern analyses

All pattern analyses were performed in Matlab (The MathWorks), using customised procedures and SPM 5 (Wellcome Department of Cognitive Neurology, London) for image handling. The FreeSurfer (http://surfer.nmr.mgh.harvard.edu/) software package was used for projecting results onto flattened cortical surfaces.

Pre-processing

To preserve subtle differences between neighbouring voxels, pattern analyses were performed on minimally pre-processed functional data in native space. Images were spatially aligned/unwarped and corrected for differences in slice-timing, but not spatially normalised or smoothed. The time-series data were extracted for each cortical voxel, defined according to a grey matter mask generated by the segment function in SPM 5, and high-pass filtered (cut-off = 128 seconds).