Research report

Patterns of response to scrambled scenes reveal the importance of visual properties in the organization of scene-selective cortex

CrossMark

David M. Watson, Tom Hartley and Timothy J. Andrews[*]

Department of Psychology and York Neuroimaging Centre, University of York, York, United Kingdom

ARTICLE INFO

ABSTRACT

Neuroimaging studies have found distinct patterns of neural response to different categories of scene in scene-selective regions of the human brain. However, it is not clear how information about scene category is represented in these regions. Images from different categories vary systematically in their visual properties as well as their semantic category. So, it is possible that patterns of neural response could reflect variation in visual properties. To address this question, we used fMRI to measure patterns of neural response to intact and scrambled scene categories. Although scrambling preserved many of their visual characteristics, perception of scene categories was severely impaired. Nevertheless, we found distinct patterns of response to different scene categories in the parahippocampal place area (PPA) and the occipital place area (OPA) for both intact and scrambled scenes. Moreover, intact and scrambled scenes produced highly similar patterns of response. Our finding that reliable and distinct patterns of response in scene-selective regions are still evident when categorical perception is impaired suggests that visual properties play an important role in the topographic organization of these regions.

1. Introduction

The ability to perceive and recognize the spatial layout of visual scenes is essential for spatial navigation. Neuroimaging studies have identified a number of regions in the human brain that respond selectively to visual scenes (Epstein, 2008). For example, the parahippocampal place area (PPA) is a region on the ventral surface of the temporal lobe that displays preferential activity to images of scenes over and above images of objects and faces (Aguirre, Zarahn, & D'Esposito, 1998; Epstein & Kanwisher, 1998). Other place selective regions include the retrosplenial complex (RSC) located immediately superior to the PPA and the transverse occipital sulcus (TOS) or occipital place area (OPA) on the lateral surface of the occipital lobe (Dilks, Julian, Paunov, & Kanwisher, 2013). Damage to these regions leads to specific impairments in scene perception and spatial navigation (Aguirre & D'Esposito, 1999; Mendez & Cherrier, 2003).

Despite the importance of scene-selective regions for spatial navigation, the functional organisation of these regions remains unclear (Groen, Silson, & Baker, 2017; Lescroart, Stansbury, & Gallant, 2015). For example, although scene-selective regions show distinct patterns of response to images of different scene categories (Walther, Caddigan, Fei-Fei, & Beck, 2009; Watson, Hartley, & Andrews, 2014), the basic organizing principles are unresolved. Some studies have argued that scene-selective regions represent information about 'high-level' semantic properties of natural scenes (Huth, Nishimoto, Vu, & Gallant, 2012; Stansbury, Naselaris, & Gallant, 2013; Walther, Chai, Caddigan, Beck, & Fei-Fei, 2011; Walther et al., 2009). This conclusion has, however, been challenged by other studies that have suggested that the patterns of response in scene-selective regions are better explained by properties of the scene, such as openness (Kravitz, Peng, & Baker, 2011; Park, Brady, Greene, & Oliva, 2011) or distance (Amit, Mehoudar, Trope, & Yovel, 2012; Park, Konkle, & Oliva, 2015) rather than by semantic category.

Although concepts such as openness or distance provide plausible 'mid-level' dimensions with which to understand the organization of scene-selective regions, it is not clear whether they can be explained at an even more basic level in terms of low-level visual properties that co-vary with these properties (Oliva & Torralba, 2001). In recent studies, we have shown that variance in the patterns of response to different scene categories can be explained by corresponding variance in the image properties of the scenes (Andrews, Watson, Rice, & Hartley, 2015; Watson, Hymers, Hartley, & Andrews, 2016; Watson et al., 2014). These findings are consistent with previously reported biases in scene-selective regions for orientation (Nasr & Tootell, 2012; Nasr, Echavarria, & Tootell, 2014), spatial frequency (Musel et al., 2014; Rajimehr, Devaney, Bilenko, Young, & Tootell, 2011) and visual field location (Arcaro, McMains, Singer, & Kastner, 2009; Golomb & Kanwisher, 2012; Levy, Hasson, Avidan, Hendler, & Malach, 2001; Silson, Chan, Reynolds, Kravitz, & Baker, 2015) and provide further evidence for the role of image properties in the organization of scene-selective regions. However, a fundamental problem is that images drawn from the same scene category or with the same spatial layout are likely to have similar visual properties (Oliva & Torralba, 2001). So, reliable patterns of response are expected under high-level, mid-level and low-level accounts of scene perception.

The aim of this study was to directly determine the extent to which the patterns of neural response across scene-selective regions can be explained by selectivity to more basic properties of the stimulus. To address this question, we measured the neural response across scene-selective regions to intact images of different scene categories, as well as versions of these images that had been phase-scrambled at a global or local level. Our rationale for using scrambled images is that they have many of the visual properties found in intact images, but disrupt perception of categorical and semantic information (Andrews, Clarke, Pell, & Hartley, 2010; Coggan, Liu, Baker, & Andrews, 2016; Loschky, Hansen, Sethi, & Pydimarri, 2010; Loschky et al., 2007). Applying scrambling both locally and globally allowed us to further investigate the importance of the spatial properties of scenes to the neural response, as local scrambling better preserves the coarse-scale spatial arrangement of visual features in the original image. Our hypothesis was that, if scene-selective regions are sensitive to the visual differences between scene categories, then we would expect to find similar patterns of neural response to these categories even when images are scrambled.

2. Methods

2.1. Participants

20 participants (5 males; mean age: 25.85; age range: 19–34) took part in the experiment. All participants were neurologically healthy, right-handed, and had normal or corrected-to-normal vision. Written consent was obtained for all participants and the study was approved by the York Neuroimaging Centre Ethics Committee.

2.2. Stimuli

Participants viewed scene images in two independent runs, one to localize the scene-selective regions, the other to experimentally investigate the effects of local and global scrambling manipulations. Images presented in the experiment runs were taken from the LabelMe database (http://cvcl.mit.edu/database.htm; Oliva & Torralba, 2001). Images for the localiser run were taken from the SUN database (http://groups.csail.mit.edu/vision/SUN/; Xiao, Hays, Ehinger, Oliva, & Torralba, 2010). Stimuli were presented using PsychoPy (Peirce, 2007, 2009) and were back-projected onto a custom in-bore acrylic screen at a distance of approximately 57 cm from the participant, with all images presented at a resolution of 256×256 pixels subtending approximately $10.7°$ of visual angle.

The image set for the main experiment comprised 180 greyscale images from 5 scene categories: city, coast, forest, indoor, and mountain (36 images per category). Each image was shown at 3 levels of image scrambling: intact, locally scrambled, and globally scrambled. Globally scrambled images were created by randomising the phase of the 2D frequency components across the whole image while keeping the magnitude constant. Locally scrambled images were created by the same process, except that scrambling was applied independently within each of 64 windows of an 8×8 grid across the image. Luminance histograms across all images in all conditions were normalised using the SHINE toolbox (Willenbockel et al., 2010). Examples of the stimuli used in each condition are shown in Fig. 1. Corresponding Fourier amplitude spectra plots are shown in Supplementary Fig. 1. In order to assess the impact of the scrambling process on the visual similarity of the scene categories, we assessed the visual statistics of the images using the GIST descriptor (Oliva & Torralba, 2001). This generates a vector for each image describing the spectral energy at assorted spatial frequencies, orientations, and spatial positions within the image. We employed 32 filters spanning 8 orientations and 4 spatial frequencies, within 64 windows of an 8×8 spatial grid, yielding vectors of 2048 values. These vectors were then correlated within- and between-categories using a leave-one-