NeuroImage 125 (2016) 681–692

Contents lists available at ScienceDirect

NeuroImage

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

Feature diagnosticity and task context shape activity in human scene-selective cortex

Matthew X. Lowe [a,b,*], Jason P. Gallivan [c,d], Susanne Ferber [b,e], Jonathan S. Cant [a]

[a] Department of Psychology at Scarborough, University of Toronto, Toronto, ON M1C 1A4, Canada
[b] Department of Psychology at St. George, University of Toronto, Toronto, ON M5S 3G3, Canada
[c] Centre for Neuroscience Studies, Queen's University, Kingston, ON K7L 3N6, Canada
[d] Department of Psychology, Queen's University, Kingston, ON K7L 3N6, Canada
[e] Rotman Research Institute, Baycrest, Toronto, ON M6A 2E1, Canada

ARTICLE INFO

Article history:
Received 8 July 2015
Accepted 31 October 2015
Available online 2 November 2015

Keywords:
Scene recognition
Texture
Layout
Ventral visual cortex
Feature integration
Attention

ABSTRACT

Scenes are constructed from multiple visual features, yet previous research investigating scene processing has often focused on the contributions of single features in isolation. In the real world, features rarely exist independently of one another and likely converge to inform scene identity in unique ways. Here, we utilize fMRI and pattern classification techniques to examine the interactions between task context (i.e., attend to diagnostic global scene features; texture or layout) and high-level scene attributes (content and spatial boundary) to test the novel hypothesis that scene-selective cortex represents multiple visual features, the importance of which varies according to their diagnostic relevance across scene categories and task demands. Our results show for the first time that scene representations are driven by interactions between multiple visual features and high-level scene attributes. Specifically, univariate analysis of scene-selective cortex revealed that task context and feature diagnosticity shape activity differentially across scene categories. Examination using multivariate decoding methods revealed results consistent with univariate findings, but also evidence for an interaction between high-level scene attributes and diagnostic visual features within scene categories. Critically, these findings suggest visual feature representations are not distributed uniformly across scene categories but are shaped by task context and feature diagnosticity. Thus, we propose that scene-selective cortex constructs a flexible representation of the environment by integrating multiple diagnostically relevant visual features, the nature of which varies according to the particular scene being perceived and the goals of the observer.

© 2015 Elsevier Inc. All rights reserved.

Introduction

How does the brain process the environment around us? Since the initial description of the scene-selective parahippocampal place area (PPA; Epstein and Kanwisher, 1998), investigations have sought to answer this question by attempting to clarify the nature of the neural representations in this region. Much of this research has revealed a primary role for PPA in the encoding of spatial features within a scene, such as structural geometry or layout (Epstein and Kanwisher, 1998; Epstein et al., 2003), spatial boundary (Park et al., 2011), and spatial depth (Kravitz et al., 2011). Conversely, recent studies support the notion that its neural representations extend beyond spatial features and include the encoding of non-spatial contextual associations of objects (Bar et al., 2008), high-level conceptual scene categories (Walther et al., 2009; 2011; Dilks et al., 2011), and surface texture and material properties (Peuskens et al., 2004; Cant and Goodale, 2007; 2011). In order to better understand scene representation, however, it is not only necessary to understand the contributions of individual features, but also how these features converge to contribute to the formation of scene identity. Yet disentangling feature-specific modulation of scene-selective neural activity within global scene representations remains a challenge, as these features rarely exist in isolation, and may inform scene identity through complex interactions which vary according to scene category.

Early research exploring diagnostic visual features in the recognition of objects revealed a primary role for edge-based information (i.e., structure), suggesting surface characteristics such as color and texture play only a secondary role in object recognition (Biederman and Ju, 1988). Research has since extended support for edge-based determinants of visual object recognition to scene perception (Delorme et al., 2000; Walther et al., 2011; Walther and Shen, 2014), yet a growing body of work suggests diagnostic surface characteristics such as color and texture are instrumental in mediating early-stage scene gist processing that is responsible for successful scene recognition (Schyns and Oliva, 1994; Oliva and Schyns, 2000; Goffaux et al., 2005; Steeves et al., 2004; Castelhano and Henderson, 2008). Given these differences, the interplay between surface properties and structural features as determinants for

---

* Corresponding author at: Department of Psychology, University of Toronto Scarborough, 1265 Military Trail, Toronto, ON M1C 1A4, Canada.
E-mail address: matthew.lowe@mail.utoronto.ca (M.X. Lowe).

http://dx.doi.org/10.1016/j.neuroimage.2015.10.089
1053-8119/© 2015 Elsevier Inc. All rights reserved.

scene recognition is currently unclear. One framework for scene perception, which may reconcile these differences, proposes that the recognition of complex visual scenes can be understood through interactions between perceptually available information and categorization demands (Oliva and Schyns, 1997). This recognition framework centers on the notion of feature diagnosticity: the idea that specific visual cues are used for specific types of categorizations and an interaction between task demands and available visual information can explain how different cues are used to recognize scenes. In other words, diagnostic visual features may emerge as a function of their usefulness in defining the identity of a scene, and the task demands placed on the observer. Thus, the present study aims to investigate the influence of diagnostic surface- and edge-based visual features on neural scene processing across a range of scene categories.

Given the variability of visual information across scene categories, Oliva and Torralba (2006) proposed that the most effective global features for scene identification will be those capturing the global structure and meaning of the visual world. For example, manufactured environments (e.g., cities) are dominated by prominent edge-based information containing straight horizontal and vertical lines, while natural landscapes (e.g., deserts) tend to have zones of characteristic textures and undulating contours which may be meaningful for scene identification (Oliva and Torralba, 2001). Thus, structural information (e.g., layout and geometry) may be of greater diagnostic relevance when discriminating scenes within manufactured environments, whereas both distinctive textured zones and undulating spatial structures may be diagnostic for scene identification in natural environments. Indeed, behavioral research has revealed the importance of global texture cues in capturing the diagnostic structure of natural scenes (Oliva and Torralba, 2006). For instance, a forest can be described in terms of the roughness and homogeneity of its textural components, providing meaningful information to a human observer comparing two forest scenes (Rao and Lohse, 1993). The neural representations of texture perception in PPA, however, have been investigated using isolated objects, and not entire scenes (Cant and Goodale, 2007; 2011), and it is therefore unclear how texture contributes to scene representations in scene-selective cortex.

In light of the importance of layout and texture information in scene perception, and potential differences in the relevance of these features for categorizing different scenes, the present study examined neural activity in scene-selective cortex while observers attended to either the layout or texture of natural and manufactured scenes, either of which could change while the other was held constant. We hypothesized that PPA would show equal sensitivity to manipulations of both layout and texture in natural scenes, where textured zones and layout may be equally relevant for distinguishing scene identity. In manufactured scenes, however, we hypothesized that PPA would show less sensitivity to texture, relative to layout, as these scenes contain prominent horizontal and vertical structural components that can aid in the discrimination of scene identity. In order to isolate effects to PPA, we also examined the modulation of brain activity in other areas of scene- and object-processing networks, and additionally localized a region of early visual cortex to examine if activation patterns observed in PPA can be dissociated from activity in early visual areas. Building on previous research (Walther et al., 2009; 2011; Park et al., 2011; Kravitz et al., 2011), we took advantage of both univariate and multivariate analyses to investigate previously unexplored questions of how task-dependent global scene features (i.e., attend to texture or layout) interact with high-level conceptual scene attributes (i.e., content: natural vs. manufactured scenes; and spatial boundary: open vs. closed scenes) to shape scene representation in human visual cortex.

Materials and methods

Observers

Twelve paid observers (6 males; mean age 27.4 ± 3.8 years) with normal or corrected-to-normal visual acuity were recruited from the University of Toronto community. Observers gave informed consent in accordance with the University of Toronto Ethics Review Board.

Stimuli and procedure

Stimuli were grayscale photographs from four different scene categories devoid of foreground objects to avoid interference (see Davenport and Potter, 2004; Joubert et al., 2007) and created by varying features of spatial boundary (open vs. closed) and scene content (natural vs. manufactured; Fig. 1)(Oliva and Torralba, 2001). After selection of our four scene categories, twelve unique structural arrangements (i.e., layouts) were selected for each category, and twelve appropriate textures were applied to the dominant surface of each layout (mapped onto scene gradient and depth using Adobe Photoshop CS3), yielding 144 unique images per scene category (12 layouts/category × 12 textures/layout × 4 scene categories = 576 total images). E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA) was used to control stimulus presentation and collect behavioral responses. Images were rear-projected onto a screen in the MRI scanner at a resolution of 500 × 500 pixels (subtending 10.4° × 10.4° of visual angle), and observers viewed stimuli through a mirror mounted to the head coil directly above the eyes. We used a blocked fMRI experimental paradigm, wherein sixteen images from a single scene category were presented in blocks of 16-s each. Each block was preceded by a 12-s fixation period and a 4-s written instruction to attend to changes in either the texture or layout of the scenes in the ensuing block.

In each trial (8 per block, lasting 2 s each), two scenes were presented for 300 ms (separated by a 200-ms blank interval), and the task of the observers was to decide if the attended feature (i.e., layout or texture) was the same or different across the two images, responding during a 1.5-s period following the onset of the second image (via a response pad placed in the observer's right hand). Each block contained an equal number of "same" and "different" trials. Observers were instructed to maintain central fixation and respond as accurately as possible, placing no emphasis on fast response times to help encourage accurate performance. Images from a single scene category were presented randomly within each block, and each image could be repeated only once per observer. Each observer took part in 8 runs (4 min 28 s each). Each run contained a unique and counterbalanced order of 8 different stimulus blocks (i.e., 8 different conditions: attend to texture or layout in each of the four scene categories). Run order was randomized across observers, and scene category was held constant per block.

Localizer scan

Stimuli used to localize object-, scene-, and face-sensitive areas of cortex, as well as early visual cortex, were photographs of various scenes, faces, common objects, and tile-scrambled images. Stimuli were presented in 16-s blocks of 32 images at a resolution of 375 × 375 pixels (7.8° × 7.8°) and were displayed for 400 ms each, with an interstimulus interval of 50 ms. Observers fixated on a centrally presented black fixation cross and were instructed to respond with a button press when the fixation cross changed from black to red (randomly occurring once or twice per stimulus block). There were 4 blocks for each stimulus category within a run, and there were two unique run orders. Each observer took part in three localizer runs (6 min 40 s each).

MRI acquisition

Scanning was performed at the Center for Addiction and Mental Health using a 3-T GE Discovery MR750 whole-body MRI scanner equipped with an 8-channel head coil. T1-weighted anatomical images were acquired using a 3D SAG T1 BRAVO spiral pulse sequence [repetition time (TR), 6736 ms; echo time (TE), 3 ms; inversion time, 650 ms; flip angle 8°, 256 × 256 matrix size, 200 slices, 1 mm isovoxel]. For the functional runs, T2*-weighted images sensitive to blood oxygenation level-dependent (BOLD) contrasts were acquired using a spiral pulse

Download English Version:

https://daneshyari.com/en/article/6024159

Download Persian Version:

https://daneshyari.com/article/6024159

Daneshyari.com