

# Visual information representation and rapid-scene categorization are simultaneous across cortex: An MEG study



Pavan Ramkumar<sup>a,\*</sup>, Bruce C. Hansen<sup>b</sup>, Sebastian Pannasch<sup>a,c</sup>, Lester C. Loschky<sup>d</sup>

<sup>a</sup> Brain Research Unit, O.V. Lounasmaa Laboratory, Aalto University School of Science, Espoo, Finland

<sup>b</sup> Department of Psychology and Neuroscience Program, Colgate University, Hamilton, NY, USA

<sup>c</sup> Department of Psychology, Technische Universität Dresden, Dresden, Germany

<sup>d</sup> Department of Psychological Sciences, Kansas State University, Manhattan, KS, USA

## ARTICLE INFO

### Article history:

Received 18 November 2015

Accepted 13 March 2016

Available online 18 March 2016

### Keywords:

Scene gist

Timing of visual perception

MEG

Multivariate decoding

Spatial envelope

Confusion matrices

Multiple linear regression

## ABSTRACT

Perceiving the visual world around us requires the brain to represent the features of stimuli and to categorize the stimulus based on these features. Incorrect categorization can result either from errors in visual representation or from errors in processes that lead to categorical choice. To understand the temporal relationship between the neural signatures of such systematic errors, we recorded whole-scalp magnetoencephalography (MEG) data from human subjects performing a rapid-scene categorization task. We built scene category decoders based on (1) spatiotemporally resolved neural activity, (2) spatial envelope (SpEn) image features, and (3) behavioral responses. Using confusion matrices, we tracked how well the pattern of errors from neural decoders could be explained by SpEn decoders and behavioral errors, over time and across cortical areas. Across the visual cortex and the medial temporal lobe, we found that both SpEn and behavioral errors explained unique variance in the errors of neural decoders. Critically, these effects were nearly simultaneous, and most prominent between 100 and 250 ms after stimulus onset. Thus, during rapid-scene categorization, neural processes that ultimately result in behavioral categorization are simultaneous and co-localized with neural processes underlying visual information representation.

© 2016 Elsevier Inc. All rights reserved.

## Introduction

To make effective decisions within the environment, our brains must be able to quickly recognize and comprehend real-world scenes. Humans are remarkable at rapidly recognizing scene categories from extremely brief exposure times (<20 ms; Potter, 1976; Fei-Fei et al., 2002; Greene and Oliva, 2009a; Loschky et al., 2010; Potter et al., 2014). The holistic semantic information that we can extract from a scene in such short durations—typically the category label—has come to be known as scene gist (Oliva and Torralba, 2001; Fei-Fei et al., 2007), and the behavioral phenomenon is known as scene gist perception or rapid-scene categorization. However, the computational and neural basis of rapid-scene categorization is not yet fully understood.

Neuroimaging studies are beginning to uncover the functional roles of brain areas involved in rapid-scene categorization (Epstein and Kanwisher, 1998; Epstein et al., 1999; Epstein et al., 2003; Epstein and Higgins, 2007; Epstein, 2008; Linsley and MacEvoy, 2014). Important

advances in the analysis of brain imaging data—encoding and decoding models of brain activity—have helped tease apart specific roles of cortical areas, resulting in a move from strictly activation-based to information-based brain mapping (Kriegeskorte et al., 2006; Kay et al., 2008; Ramkumar et al., 2013; Cecotti et al., 2015). On the one hand, functional magnetic resonance imaging (fMRI)-based decoding studies have provided novel insights into scene categorization (MacEvoy and Epstein, 2009, 2011; Walther et al., 2009, 2011) by elucidating the distributed patterns of activation of various scene-related regions. On the other hand, electroencephalography (EEG; e.g., Thorpe et al., 1996; Vanrullen and Thorpe, 2001; Goffaux et al., 2005; Hansen et al., 2011, 2012; Groen et al., 2013), electrocorticography (ECOG; e.g., Bastin et al., 2013) and sensor-level magnetoencephalography (MEG; e.g., Cichy et al., 2014) studies have revealed the rapid time course of visual information processing.

However, the temporal relationship between when scene-related information becomes available and when it becomes useful for behavioral categorization remains elusive. Corroborating human performance against scene categorization based on visual features of natural scenes may help to provide insight into that relationship. In particular, the spatial envelope model (SpEn; Oliva and Torralba, 2001), which has been shown to be diagnostic of global scene properties, such as naturalness, openness, and depth (Greene and Oliva, 2009b), and also of basic-level scene categories (Oliva and Torralba, 2006) is a promising

\* Corresponding author at: Sensorimotor Performance Program, Rehabilitation Institute of Chicago and Departments of Physical Medicine & Rehabilitation and Neurobiology, Northwestern University, Chicago, IL, USA.

E-mail addresses: [pavan.ramkumar@northwestern.edu](mailto:pavan.ramkumar@northwestern.edu) (P. Ramkumar), [bc Hansen@colgate.edu](mailto:bc Hansen@colgate.edu) (B.C. Hansen), [Sebastian.Pannasch@tu-dresden.de](mailto:Sebastian.Pannasch@tu-dresden.de) (S. Pannasch), [loschky@ksu.edu](mailto:loschky@ksu.edu) (L.C. Loschky).

candidate. Although these studies are informative about the nature of visual features that the brain might use to represent natural scenes, they do not interrogate when and where in the brain such information becomes useful for behavioral rapid-scene categorization (but see Park et al., 2011 for a study of where in the brain some spatial envelope features may be processed).

Temporally decoupling computational steps might be a good first step to understand this relationship. For instance, representing visual features, and making categorical judgments based on these representations have been studied as separate computational phenomena, but how are they executed temporally? The evidence from behavioral studies is mixed. Some studies suggest a serial process of gist perception. For instance, some global scene properties (such as naturalness) are inferred earlier than others (such as depth) (Greene and Oliva, 2009b), the distinction between natural and man-made scenes is made earlier than the basic-level scene category (Loschky and Larson, 2010; Joubert et al., 2007), and increasingly complex percepts (such as describing the relationships between objects) require increasingly longer exposure times to scenes (Fei-Fei et al., 2007). Other behavioral studies suggest a considerable overlap between visual representation and category judgment processes due to rapid reaction times to indicate percepts. For instance, saccades made to indicate decisions about objects or faces are made within 100–130 ms (e.g., Kirchner and Thorpe, 2006; Crouzet et al., 2010). These psychophysical and eye movement studies provide upper and lower bounds on the time course of visual processing, but they do not directly quantify visual information processing in the brain.

Here, we investigated the brain–behavior relationship underlying rapid-scene categorization by asking whether the early visual representation of scenes and category judgment are serial processes, or whether they occur at the same time. The whole-scalp coverage of magnetoencephalography (MEG), its good cortical resolution, and millisecond-scale temporal resolution make it particularly attractive for describing rapid-scene categorization phenomena with high spatial and temporal precision. To track the neural correlates of rapid-scene categorization over the cortex and in time, we recorded MEG data when subjects performed a rapid-scene categorization task. First, we show that it is possible to track visual scene information flow in the ventral visual stream using spatiotemporal maps of decoding accuracy. Next, and crucially, we describe a network of brain areas whose decoders' errors can be explained by errors in behavioral responses despite accounting for errors in image-feature-based decoders. Specifically, we show that the representation of low-level visual features and processing that informs behavioral choice are not sequential, but co-occur at the same time and within the same cortical networks.

## Materials and methods

### Apparatus

Visual stimuli were presented with a Panasonic D7700 DLP projector on a back-projection screen placed at 120-cm distance in front of the subject. The screen resolution was set to  $1024 \times 768$  pixels and images were presented at the center of the screen with a size of  $672 \times 672$  pixels, resulting in viewing angle of  $23.6 \times 23.6^\circ$ .

To track brain function at high temporal resolution, we recorded MEG data using a Vectorview system (Elekta Oy, Helsinki, Finland) comprising 306 channels (204 planar gradiometers, 102 magnetometers) providing whole-head coverage. Data were sampled at 1000 Hz and filtered at 0.03–330 Hz. We also recorded signals from four coils positioned around the head to continuously track head position.

Concurrently with MEG data, we acquired eye movement data using an SR Eyelink 1000 (SR Research Ltd., Mississauga, Ontario, Canada) infrared eye-tracking system (sampled at 1000 Hz) to ensure that our subjects fixated on the center of the screen.

### Scene categorization experiments

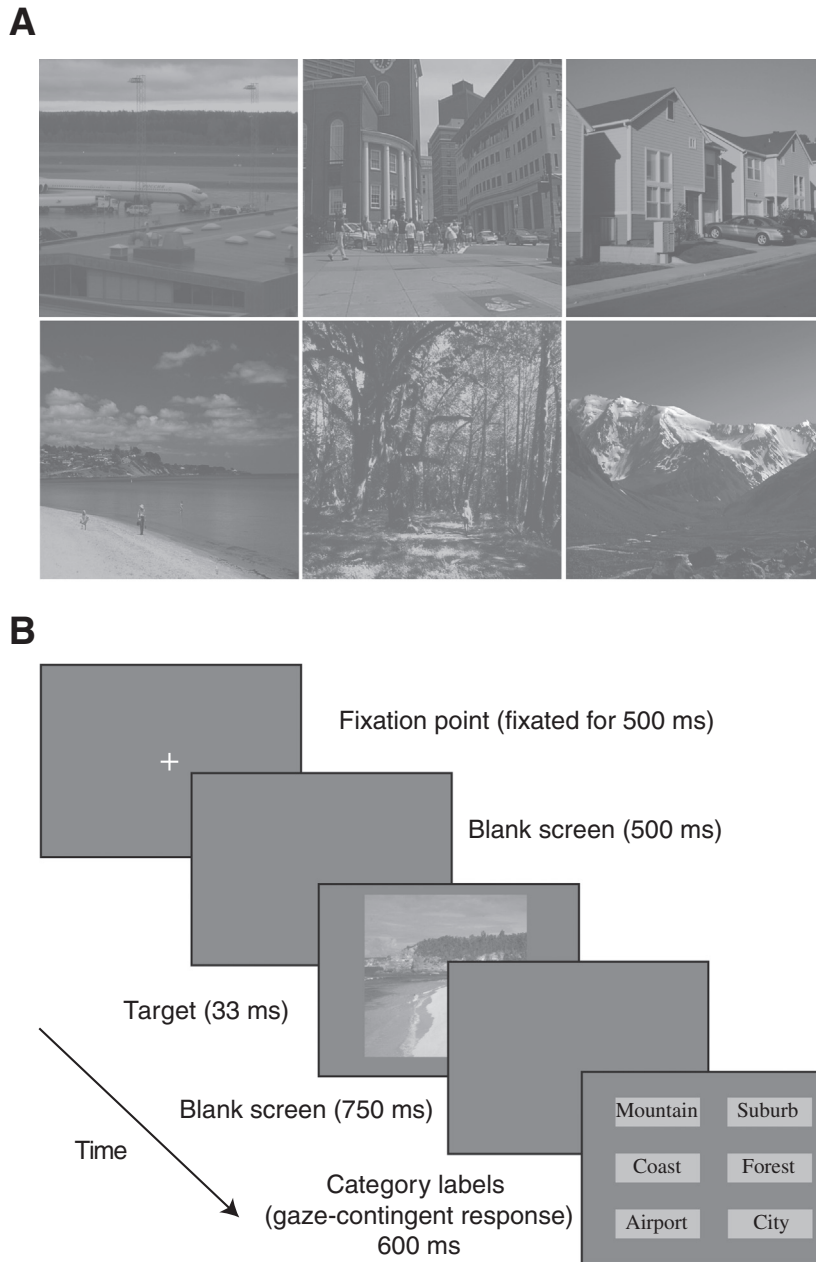
We designed a scene categorization experiment in which subjects had to view a briefly presented grayscale image and report one of six possible natural scene categories. Three categories were man-made scenes (airports, cities, or suburbs), and the other three were natural scenes (coasts, forests, or mountains); see Fig. 1A. Each category comprised 30 unique scenes. Although behavioral experiments of rapid-scene categorization often use visual masking to limit the duration of information extraction from the retinal image, we chose to implement an unmasked task for the MEG scanner because we wanted to measure the visual response to natural scenes without any influence from the masking images. In order to ensure that the low task difficulty did not produce noisy estimates of behavioral performance, we also designed a masked version of the task to be performed outside the scanner (see Results, Supplementary Text and Supplementary Figs. S1, S2).

A trial schematic is illustrated in Fig. 1B for the rapid-scene categorization task. Eight healthy volunteers (2 females, mean age 32 years) were asked to perform a basic-level categorization task using an eye-gaze-based response interface. Prior written informed consent was obtained from all subjects. MEG and eye movement recordings were approved by the Ethics Committee of the Helsinki and Uusimaa Hospital District. Each trial started with the presentation of a fixation point. After fixating the point for 500 ms, a blank screen appeared for 500 ms, followed by the target image presentation for 33 ms. After the target image, a blank screen was shown for 750 ms. Subjects then had to select the category of the previous target image by making an eye movement to the respective category label and fixating it for 600 ms. After indicating their choice, the experiment progressed to the next trial. In order to ensure central fixation for each trial, stimulus presentation was only initiated when fixation was within  $\sim 1^\circ$  of the fixation point.

### Preprocessing and source modeling

The MEG signal is contaminated by several types of artifacts arising from eye blinks, facial muscle contractions, head movements, vibrations, and sources of electromagnetic noise in the environment (Hämäläinen et al., 1993). Temporal signal space separation (tSSS) is a state-of-the-art technique that separates out contributions to the measurements arising from inside and outside the sphere circumscribed by the sensors using spherical harmonic functions to describe multipolar contributions to the magnetic field in space (Taulu and Simola, 2006). We used tSSS for artifact rejection and head movement compensation. We then down-sampled data to 500 Hz and low-pass filtered at 45 Hz to focus on stimulus-evoked responses. We extracted evoked responses from the continuous data stream around the stimulus onset events and applied a baseline correction using a time window of 150 ms preceding the stimulus onset. We then used data from a post-stimulus window of 600 ms for analysis. This interval preceded the display screen for indicating their category choice, and therefore, eye movement signals or artifacts did not confound the MEG responses to scenes.

We projected every single time bin (300 time bins for a 500 Hz sampling rate) for each single trial from MEG sensor space to the cortical surface using a standard inverse modeling technique, namely minimum-norm estimation (Hämäläinen and Ilmoniemi, 1994) as follows. First, we reconstructed the cortical surface from single-subject MR images using a suite of tools that segments gray matter tissue and models the shape of the cortical surface (<http://brainvisa.info/toolboxes.html>). Second, to model each source current as a magnetic current dipole, we defined a cortically constrained source space with an average spacing of 5 mm between source vertices, with a current dipole at each vertex. Third, a boundary-element model was used to compute the forward model from current sources to measured fields. Finally, we computed an inverse solution using a standard whitened, depth-weighted linear



**Fig. 1.** (A) Examples of the stimuli of all categories. Top panel: man-made categories, viz. airport, city, residential. Bottom panel: natural categories, viz. beach, forest, and mountain. (B) Trial schematic for the rapid-scene categorization task.

minimum-norm estimate. The entire inverse modeling was accomplished using the Brainstorm software package (Mosher et al., 2005).

#### Spatiotemporally resolved decoding

Inverse modeling provides us a trial-by-trial estimate of cortical activity at each vertex and each time point. From these estimates, we built decoders of scene category using data from spatiotemporally resolved windows. The spatial extent of each window was defined as a neighborhood of 25 nearest neighbors around each source-space vertex, defined on the tessellated cortical surface. The temporal extent of each window was 20 ms trailing each time point. For each window (10 timestamps at 500 Hz sampling rate  $\times$  25 vertices = 250 features per sample), we built linear support vector machine (SVM) decoders using the LIBSVM package for Matlab (Fan et al., 2008; Chang and Lin, 2011). Decoding accuracies were estimated using 5-fold cross-

validation (train on 144 trials, test on 36 trials for each fold), and 95% confidence intervals of decoding accuracies were estimated by bootstrapping over 1000 repeated samples with replacement on the test set (see e.g., Ramkumar et al. (2013) for methodological details). Group-level decoding accuracies at each vertex and time point were obtained by averaging the single-subject median accuracies, and group-level CIs were obtained by bootstrapping these median accuracies across vertices and subjects. We used the bias-corrected approximate percentile method, implemented in the Matlab function `bootci.m`, to estimate uncorrected 99.995% confidence intervals (resulting in an effective 95% CI after correcting for 1000 comparisons). Given an average of  $\sim$ 100 vertices for each region of interest and 8 subjects, this is a conservative correction for multiple comparisons. Decoding accuracies were considered significant if the lower bound CIs exceeded the chance level of one sixth. From the group-level decoding accuracies, we also computed confusion matrices: a matrix whose  $(i, j)^{\text{th}}$  entry represents

the fraction of trials for which category  $i$  was predicted as category  $j$  by the neural decoders (neural confusion matrices). We then visualized the time series of decoding accuracies averaged over selected anatomical regions of interest (ROIs), based on a recent atlas by Klein and Tourville (2012); see Table 1 for names of our regions of interest.

#### Decoding scene categories from spatial envelope features

The spatial envelope (SpEn) is a popular candidate model for how the brain might represent scenes during rapid categorization (Oliva and Torralba, 2001; Oliva and Torralba, 2006). To study the neural representation of low-level visual features, we computed SpEn features from each image and, using SVMs, decoded scene categories on their basis. For each stimulus image, we normalized local contrast and computed the SpEn features using an open source Matlab package provided by the authors (<http://people.csail.mit.edu/torralba/code/spatialenvelope/>). The SpEn features are localized energy spectra obtained by computing the energies of the input image convolved with Gabor filters at multiple scales and orientations. We precomputed Gabor filters at 8 orientations and 6 scales in the Fourier domain, multiplied each filter with the Fourier transform of the input image, and subsequently inverted the Fourier transform. We divided each filtered image into a coarse  $4 \times 4$  grid and averaged the Fourier energies across the pixels in each block of the coarse grid, resulting in  $8 \times 6 \times 4 \times 4 = 768$  image features. Using these features, we decoded scene categories using SVMs and computed decoding accuracies and the corresponding confusion matrices (SpEn confusion matrices).

#### Regression of confusion matrices

To understand when and where low-level visual features and behavioral categorization contribute to the pattern of errors in neural decoders, we applied a multiple linear regression approach. In particular, we explain the off-diagonal entries in neural confusion matrices as a linear combination of the corresponding entries in the SpEn confusion matrices, obtained from SpEn-based decoders and across-subject-averaged behavioral confusion matrices, obtained from behavioral categorization. For each cortical vertex and temporal window, we thus obtain two regression coefficients:  $\beta_v$  and  $\beta_b$ , corresponding to visual and behavioral covariates (see Fig. 2).

By regressing a feature of interest against a variable, we can measure the extent to which the feature can describe the variance using the  $R^2$  statistic, given by  $R^2 = 1 - SS_{\text{res}} / SS_{\text{tot}}$ , where  $SS_{\text{res}}$  is the sum of squares of the residual, and  $SS_{\text{tot}}$  is the sample data variance. However, these traditional regression models do not uniquely quantify the variance explained by a given feature with respect to other potential features of interest. For our problem, non-unique variance is the variance of neural confusions explained by either the SpEn confusions or the behavioral confusions alone.

Multiple linear regression allows us to selectively measure uniquely explained variance of a candidate feature by comparing the full model

comprising all features against a partial model with the feature of interest left out. In our case, unique variance is the variance of neural confusions explained by the SpEn or behavioral confusions after regressing out the behavioral or SpEn confusions respectively. This can be quantified using the relative  $R^2$  statistic, given by  $R^2_{\text{relative}} = R^2_{\text{full}} - R^2_{\text{partial}}$ .

We computed both  $R^2$ s and relative  $R^2$ s for SpEn and behavioral confusions for each subject using the median values of the confusion matrices. To visualize the time series of these relative  $R^2$ s in ROIs, we computed the mean and multiple-comparison corrected bootstrapped CIs (99.995%) across subjects and vertices within selected ROIs, using the same bias-corrected approximate percentile method described above. Note that although we implicitly treat subjects as fixed effects, the multiple-comparison correction of the mean of these fixed effects across subjects allows us to compare the distribution of the mean of the relative  $R^2$  statistic against zero, in order to quantify unique explained variance.

## Results

Our goal in this study was to investigate the temporal relationship between when low-level scene-related information becomes available and when it becomes useful for behavioral categorization. We approached this by mapping the cortical processing of scene gist in a spatially and temporally resolved manner. To this end, we recorded brain activity when subjects categorized natural scenes after brief presentations. Using cortically constrained current sources inferred from MEG recordings, we decoded scene categories. Once we established that brain activity contained scene category-selective information, we asked how well this information could be explained by low-level visual descriptors of scenes. Thus, we modeled the pattern of errors in our neural decoders using decoders built from visual features and behavioral reports of scene categories.

#### Behavioral categorization

We found that subjects were extremely accurate in reporting categories of natural scenes within the scanner ( $92.5 \pm 4.9\%$ , SD,  $N = 8$  subjects). To ensure that this high performance did not prevent us from estimating the pattern of errors accurately, we repeated a masked variant of the categorization task outside the scanner in order to increase the level of difficulty, on a different set of subjects (see Supplementary Text, Supplementary Figs. S1, S2). In this version of the task, subjects were less accurate ( $64.1 \pm 10.7\%$  SD,  $N = 20$  subjects). The correlation between the off-diagonal entries of the behavioral confusion matrices (Supplementary Fig. S2) between the masked and unmasked tasks was  $\rho = 0.51$  ( $p < 0.05$ ). This high degree of correlation between the confusion matrices suggested that even though accuracies were high within the scanner, the confusion matrices were good estimates of the underlying pattern of errors across scene categories.

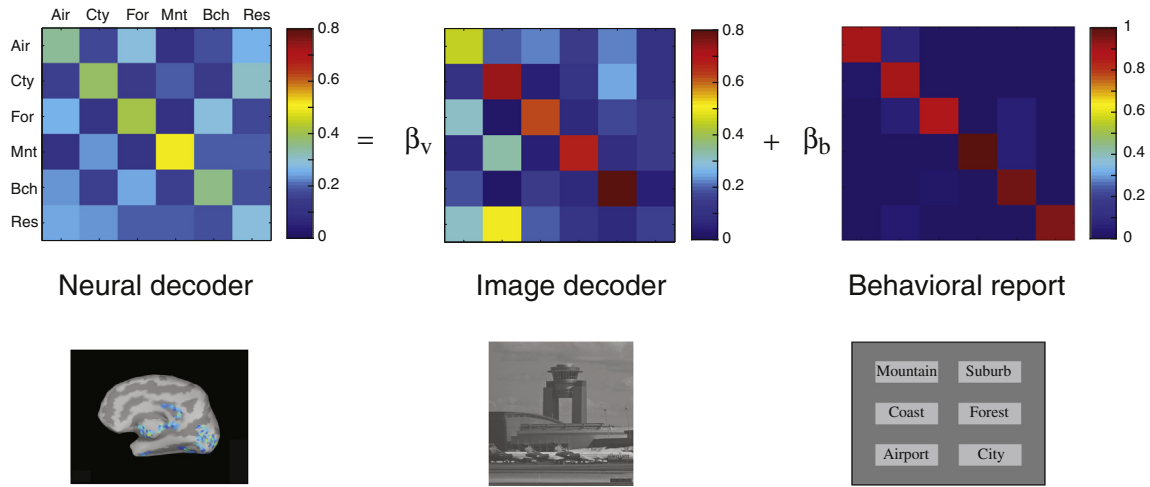
#### Decoding scene categories from source-localized MEG responses

We built decoders of scene category labels from small cortical neighborhoods spanning 25 vertices in short 20-ms time windows. Although the decoding accuracies were low ( $< 25\%$ ), they were significantly above the chance level of one sixth across several brain regions (Figs. 3 and 4).<sup>1</sup> As expected from previous EEG reports of early visual and categorical processing (e.g., Thorpe et al., 1996; Vanrullen and Thorpe, 2001; Goffaux et al., 2005), our accuracies rose above chance level as early as 100 ms, and peaked at about 150 ms, primarily in the medial occipital and temporal cortices of both hemispheres (Fig. 3). Thus, our finding of early rapid-scene information processing in the brain complements EEG studies of gist perception from extremely brief exposures.

<sup>1</sup> Since we decoded brain data at the single trial level, the low accuracies were most likely due to the limited number of trials for each scene category.

**Table 1**  
List of regions of interest (ROIs; adapted from Klein and Tourville (2012)) examined for time-resolved decoding and regression analysis of confusion matrices.

ROI	Expansion	ROI	Expansion
V1	Primary visual cortex	aFFG	Anterior fusiform gyrus
OP	Occipital pole	FFS/COS	Fusiform/Collateral sulcus
LING	Lingual gyrus	pFFG	Posterior fusiform gyrus
POS	Parieto-occipital sulcus	aPC	Anterior precuneus
CUN	Cuneus	pPC	Posterior precuneus
RSC	Retrosplenial cortex	TOS	Transverse occipital sulcus
PHC	Parahippocampal cortex	PIT	Posterior inferior parietal
PRC	Perirhinal cortex	VSP	Ventral superior parietal
ERC	Entorhinal cortex	VIP	Ventral inferior parietal
OFC	Orbitofrontal cortex		



**Fig. 2.** Schematic figure to explain CM regression analysis. Category labels are given abbreviated as follows: Air: airport; Cty: city; For: forest; Mnt: mountain; Bch: beach; Res: residential. The patterns of errors in the neural CM (off-diagonal entries) were explained as a linear combination of the pattern of errors in the image CM and the behavioral errors.

Next, we examined specific regions of interest informed by Fig. 3 based on anatomical parcellations of lateral occipital, and medial occipital and temporal cortices. As shown in Fig. 4, on the lateral surface, we found that scene category information could be decoded from a region of interest encompassing the transverse occipital sulcus (TOS) that contains the functionally defined occipital place area (OPA) (Nasr et al., 2011; Epstein and Julian, 2013; Dilks et al., 2013) as well as the ventral part of the superior parietal cortex (VSP), the posterior part of the inferior parietal cortex (PIP), and the posterior part of the inferior temporal cortex (PIT) in both hemispheres. Also shown in Fig. 4, on the medial surfaces of both hemispheres, the following areas were similarly informative: the primary visual cortex (V1), the occipital pole (OP), the lingual gyrus (LING), the parieto-occipital sulcus (POS), the cuneus (CUN), the posterior precuneus (pPC), the fusiform or collateral sulcus (FFS/COS), and the posterior fusiform gyrus (pFFG). Given the selectivity of the fusiform cortex to faces (Epstein and Kanwisher, 1998), their involvement in scene category encoding is potentially surprising. However, a recent meta-analysis of scene-selective cortical areas (Nasr et al., 2011) revealed that the parahippocampal place area is localized to parts of the fusiform gyrus. Perhaps more surprisingly, regions typically implicated in fMRI studies of scene perception (Kravitz et al., 2011; Nasr et al., 2011; Ranganath and Ritchey, 2012; Park et al., 2014), such as parahippocampal cortex (PHC), perirhinal cortex (PRC), entorhinal cortex (ERC), and orbitofrontal cortex (OFC) were not found to encode scene category-selective information, although the retrosplenial cortex (RSC) in the right hemisphere did seem to encode scene categories. To summarize, we could decode scene categories from brain areas all over the dorsal and ventral visual stream, as well as the fusiform cortex, but not from several medial temporal areas commonly implicated in scene-selective processing.

Decoding a variable of interest, in this case, categorization of briefly flashed scenes, from brain activity can selectively inform us that the brain regions and time windows in question are involved in processing information related to that variable. To examine how information related to the categorical variable is encoded while avoiding the high dimensionality of encoding models (e.g., Kay et al., 2008), we studied the pattern of errors in decoding models. A convenient advantage of analyzing errors is that even if decoding accuracies are weak, the pattern of miscategorizations can be used as a window into brain function.

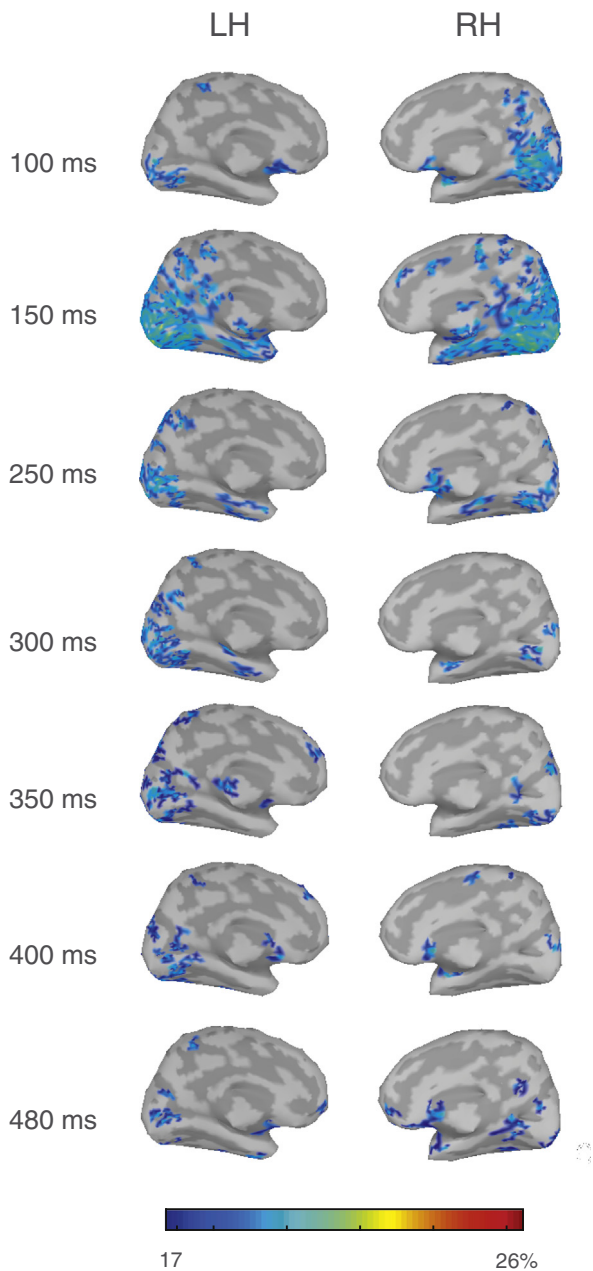
*Simultaneity of visual representation and behavioral categorization*

To obtain an understanding of the information in the neural code that enables scene category decoding, we aimed to distinguish between

brain regions that are involved in primarily representing low-level visual information that is captured by the spatial envelope and brain regions that are involved in representing alternate visual features, or potentially in making judgments about the scene category, as represented by behavior. We thus examined the pattern of errors in brain decoders in relation to the pattern of errors in SpEn decoders and behavioral reports. Brain areas which are better explained by behavioral confusion matrices than SpEn confusion matrices *may* represent the active synthesis involved in recognizing the scene category from visual representations. To this end, we attempted to explain the pattern of errors in spatiotemporally resolved neural decoders by a linear combination of the pattern of errors in SpEn decoders and the pattern of errors in behavioral reports of our subjects using multiple linear regression of confusion matrices (see [Materials and methods](#) for details).

We found that for a majority of the regions from which we were able to decode scene categories, both SpEn and behavioral confusions were able to explain some variance in the neural confusions (Figs. 5 and 6). These variances can each be broken down into non-unique and unique variance (Fig. 5A; red and blue traces show an example ROI). The variance explained by SpEn confusion matrices in general exceeds the variance explained by behavioral confusion matrices (Fig. 5A; blue bars). However, the behavioral confusion matrices explain unique variance that is independent of the variance explained by SpEn confusion matrices (Fig. 5B; all dots lie above the  $y = x$  line, indicating that the joint model does significantly better than the SpEn model). Thus, adding the behavioral CM as a covariate increases the explained variance, suggesting that the errors in the neural decoder result from both the representation of natural scene statistics as modeled by SpEn, and the behavioral judgment of scene category.

If the behavioral report of a scene category is the outcome of a sequential decision-making process, we might expect the neural encoding of visual features to precede decision-making based on those features. If on the other hand, decision-making happens at the same time as visual encoding, we might expect the respective neural correlates to be more temporally overlapping. One notable feature of the relative  $R^2$ s for the SpEn and behavioral confusions, which speaks to the different predictions of these alternate hypotheses, is their near simultaneous timing (Fig. 6). Specifically, both account for variance in neural confusions very early and around the same time range (~100–250 ms after the stimulus was flashed). We also repeated the regression analyses with behavioral confusion matrices estimated from the masked task and found that the results were indistinguishable from the outcome of analysis with behavioral confusion matrices within the scanner (Supplementary Figure S3). Thus, any inferences about unique variance



**Fig. 3.** Scene categories can be decoded in source space. Spatial maps of decoding accuracies are overlaid on the inflated and flattened cortical surface at different time points after scene onset; medial views of the left and right hemispheres are shown. Maps are thresholded to show regions from which scene categories can be decoded above chance level (16.67%).

explained by behavioral and SpEn error patterns cannot be attributed to biased or inaccurate estimates of the behavioral confusion matrices. Thus, the neural correlates of behavioral errors appear as early as the neural correlates of spatial envelope errors across categories, suggesting that rapid-scene categorization is comprised of highly parallel sets of sub-processes. Therefore, neural activity that is relevant to behavioral choice is temporally simultaneous with the encoding of low-level visual features.

## Discussion

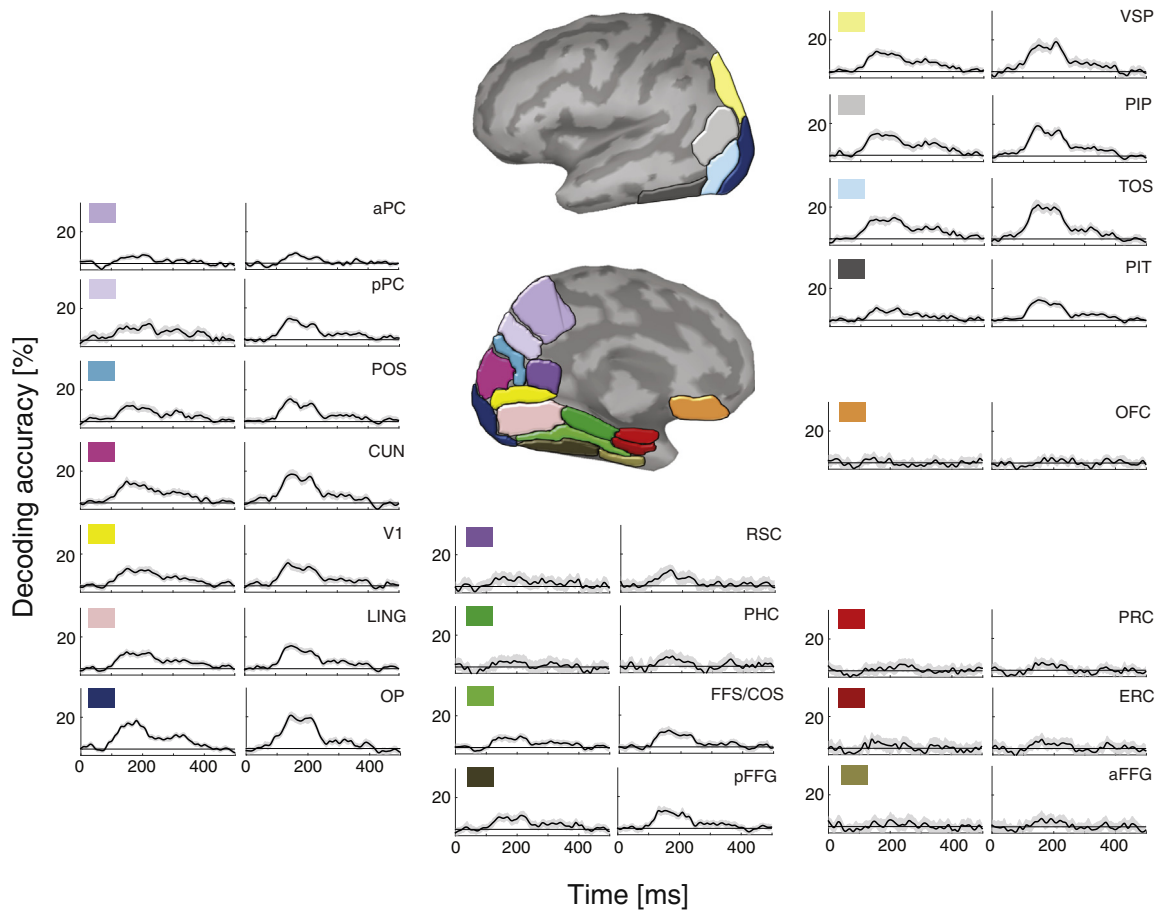
We studied the neural basis of rapid-scene categorization in a spatio-temporally resolved manner. By decoding scene categories from MEG signals generated when subjects categorized scenes, we showed that

the lateral and medial occipital, parietal, and temporal cortices contain information about scene gist from 100 to 250 ms after stimulus onset. To further tease apart whether the encoded information primarily represented low-level visual information or higher-order features that influenced behavioral errors, we applied multiple linear regression on neural confusions using visual (image statistical) and behavioral confusions. The very same regions that represented low-level features could also explain unique variance in neural confusions that were directly related to behavioral confusions. Crucially, both visual and behavioral confusions predicted neural confusions nearly simultaneously, suggesting a temporal overlap in encoding of visual features and processing that influences behavioral choice.

Earlier studies have shown that category-diagnostic information for object recognition tasks (e.g., animal or vehicle detection; e.g., Thorpe et al., 1996; Vanrullen and Thorpe, 2001; Goffaux et al., 2005) is available in the evoked-EEG response from around 150 ms. Our findings largely replicate these reports showing that category-selective decoding activity occurs largely between 100 and 250 ms post-stimulus. Relative to these studies, our results are novel in two different ways. First, we explicitly confirm using multiple linear regression that MEG signals contain category-diagnostic information over and above low-level visual features, which significantly influence behavioral choice. Second, in addition to identifying the time course of this scene-categorization-dependent activity, we specifically tie it to several anatomical areas in the lateral and medial occipital, parietal, and temporal lobes that are informative of scene category and behavioral choice.

Recent fMRI studies of scene perception have elucidated the functional roles of scene-selective areas such as the retrosplenial cortex (RSC), the parahippocampal place area (PPA), and the occipital place area (OPA) (Epstein et al., 2003; Epstein and Higgins, 2007; Epstein, 2008; Kravitz et al., 2011; Walther et al., 2009). Although we did not functionally localize the OPA and the PPA using a canonical contrast task such as faces vs. houses or places vs. non-places, the anatomical location of the PPA has been reported to lie at the intersection of the medial fusiform gyrus (mFFG) and the collateral sulcus (COS) (Nasr et al., 2011; Kanwisher and Dilks, 2013). Here, we found category-specific information in the fusiform sulcus (FFS), as well as the medial fusiform gyrus (mFFG) and COS. In addition, we found that the transverse occipital sulcus (TOS) (which anatomically overlaps with the OPA; see Nasr et al., 2011; Dilks et al., 2013), and its neighboring lateral surface ROIs, SVP, PIP, and PIT, encoded scene category information. Finally, we found scene category-specific coding to a somewhat lesser extent in the right RSC. Thus, our results are actually consistent with previous studies that have identified scene-specific areas. The weaker effect of scene category coding in RSC with respect to PPA and OPA could potentially be explained by imperfections in source localization to the RSC. In particular, because of its location right next to the calcarine sulcus (V1) and the parieto-occipital sulcus, depth-weighting in MEG source modeling (Lin et al., 2006) might attribute the RSC activations to these neighboring sulci. Although absence of clear evidence of scene category encoding in the RSC cannot be taken as evidence of absence, it is useful to consider emerging data on the difference between the functional roles of PPA, OPA, and RSC (Epstein, 2008; Cant and Goodale, 2011; Epstein and Julian, 2013). Whereas PPA is thought to represent textural properties, and OPA, contour and shape (Cant and Goodale, 2011), RSC is thought to be involved in higher cognitive functions such as contextualizing the scene in a broader environment (Epstein, 2008). Such broader scene contextualization might be outside the scope of gist perception involved in rapid-scene categorization.

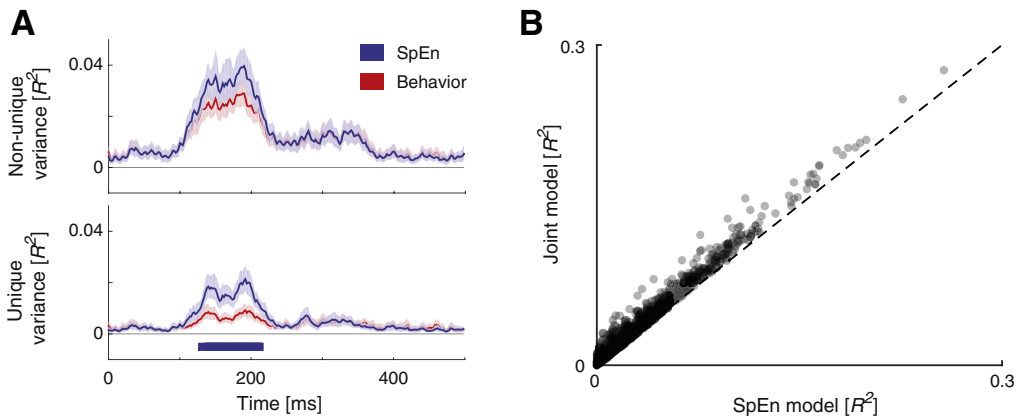
Advances in data analyses of neuroimaging experiments have enabled us to go beyond activation-based mapping, which simply implicates a certain region in a certain task. Information-based brain mapping (Kriegeskorte et al., 2006) using multivariate decoding models as well as high-dimensional encoding models enable us to specifically ask what information certain regions are representing at time epochs of interest that are relevant to the task. Here, we used decoding to



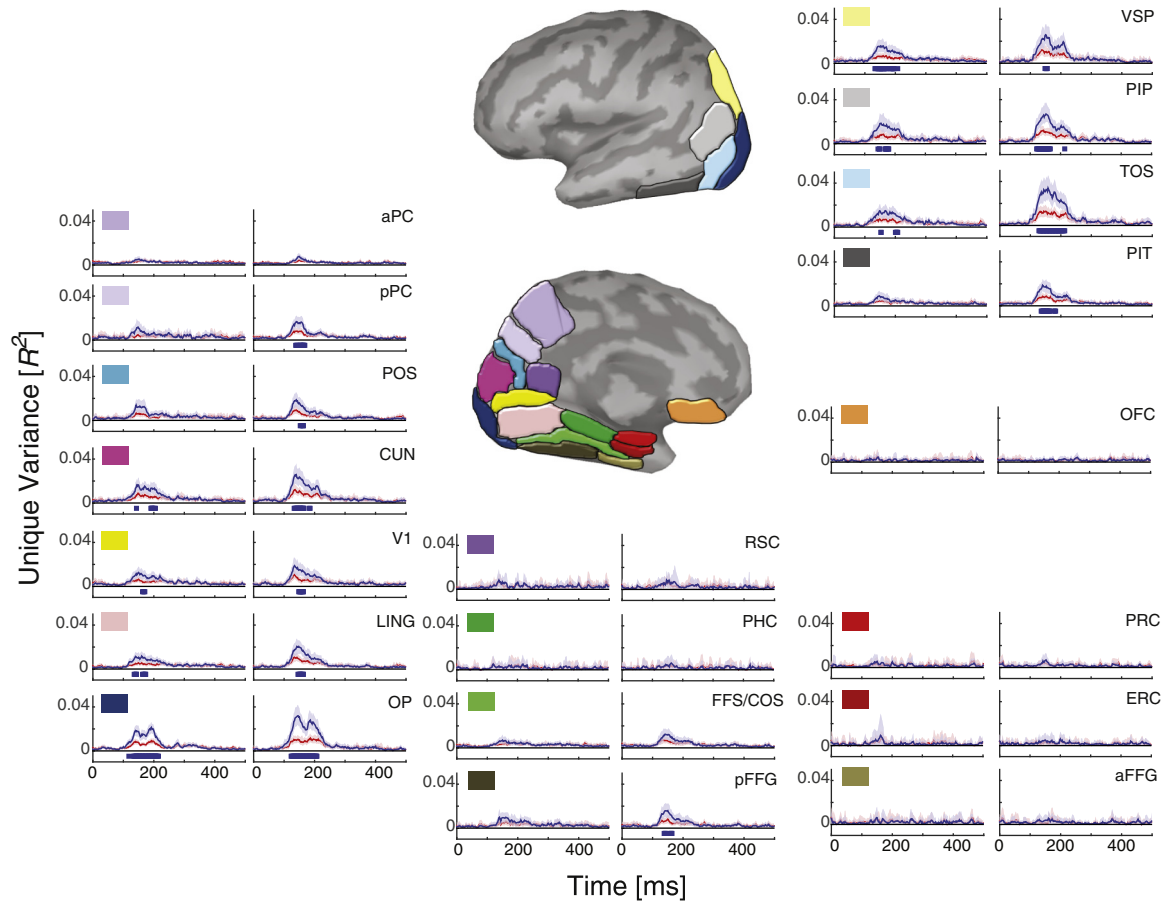
**Fig. 4.** When can scene gist be decoded? Time series of decoding accuracies averaged across selected regions of interest (ROI) and subjects. Error bars show multiple-comparison corrected, 95% bootstrapped confidence intervals of cross-validated median decoding accuracies. The thin horizontal line just above the x-axis is at 16.67% accuracy, the chance level. In the two time series plots for each ROI, the left and right plots represent data from the left and right hemispheres respectively. The legends show the ROIs on the lateral (above) and medial (below) surfaces of the left hemisphere, but corresponding ROIs were also analyzed from the right hemisphere.

investigate the spatiotemporal relationship between low-level image representations and processes related to behavioral categorization. Instead of using high-dimensional encoding models that suffer from combinatorial explosions, we used multiple linear regression of decoding errors to investigate the specific nature of the information that enables the decoding of scene categories. This analysis showed that the lower-

level image statistical information captured by the SpEn explains a large proportion of the decoding variance, but there is also a smaller independent portion of the decoding variance that explains human behavioral responses. Although neural encoding and decoding models are now well established, when combined with MEG's temporal and spatial resolution, and cortically constrained inverse modeling, they



**Fig. 5.** (A) Non-unique vs. unique variance explained for the left occipital pole (OP; an example ROI).  $R^2$  estimates are averaged all vertices within the ROI and again across subjects. Error bars give multiple-comparison corrected, bootstrapped 95% CIs. (B)  $R^2$  of a joint model comprising behavioral and SpEn confusions vs. the  $R^2$  of a partial model comprising SpEn confusions alone. Each dot represents the mean  $R^2$  of a unique vertex within area OP from one subject; vertices from all subjects are overlaid on the same plot.  $R^2$ s were averaged within a 100–250 ms window. The joint model performs only slightly better than the partial model, suggesting that adding behavioral confusions improves the fit to neural confusions only modestly.



**Fig. 6.** Median cross-validated relative  $R^2$ s for each ROI, averaged across vertices within the ROI and subjects. Error bars show multiple-comparison corrected bootstrapped 95% confidence intervals. Relative  $R^2$ s are a measure of the extent to which the pattern of errors of SpEn or behavioral decoders explains *unique* variance in the neural decoder's pattern of errors. Red: behavioral CM. Blue: SpEn CM. Blue bands at the bottom of each trace indicate the time durations for which unique variance in neural decoder-based errors explained by SpEn-decoder-based errors exceeds that of behavioral errors. In the two time series plots for each ROI, the left and right plots represent data from the left and right hemispheres respectively. The legends show the ROIs on the lateral (above) and medial (below) surfaces of the left hemisphere.

allow us to paint a portrait of the spatiotemporal neural code of scene categorization processes that is more detailed than any before it.

Although we bring together errors both in behavioral responses and in visual feature decoders to explain neural confusions, we only consider the spatial envelope feature space. Our primary motivation for using the SpEn model is the extensive literature showing that it is diagnostic of global scene properties (Greene and Oliva, 2009b), scene categories (Oliva and Torralba, 2001), errors (Greene and Oliva, 2009b), and scene typicality (Ehinger et al., 2011). SpEn features are closely related to multi-scale Gabor pyramids, which resemble receptive fields in primary visual cortex (Ringach, 2002) and are therefore better poised to explain neural confusions in early visual areas. Thus, any variance that remains unexplained by the SpEn confusion matrices could potentially be explained by visual decoders that use features encoded by the ventromedial areas (such as V4 and inferior temporal cortices). Indeed, other studies argue for local phase alignment (Hansen and Loschky, 2013; Loschky et al., 2007; Loschky et al., 2010) and color (Oliva and Schyns, 2000; Goffaux et al., 2005) as playing important roles in gist perception. Furthermore, recent studies based on line drawings and textural manipulations have challenged the view that spatial envelope statistics are necessary for scene categorization (Walther et al., 2011; Walther and Shen, 2013; Choo et al., 2014). Therefore, the question of what visual features best inform rapid-scene categorization remains open.

When the visual system arrives at a preconscious perceptual decision about the scene category, it combines developmental priors with

incoming natural image statistics (Bar, 2004; Kersten et al., 2004; Hegd , 2008). These priors result in behavioral biases; for instance, animals are recognized faster than non-living objects (Thorpe et al., 1996; Thurgood et al., 2011) and natural versus man-made category judgments occur faster than basic-level category judgments (Loschky and Larson, 2010). A potential weakness in our study is that whereas behavior and brain activity reflect these biases, our visual decoders do not. In the future, it might be useful to develop a fully Bayesian visual feature decoder that incorporates such behavioral biases as priors. Experiments that manipulate priors combined with such decoders might also help us to read out neural correlates of perceptual decision-making that are distinct from visual representations.

Three other directions could inform alternative candidates for visual feature spaces that the brain might use to represent scenes. First, although SpEn is a highly successful representation of low-level natural image statistics for image classification tasks, our results must be interpreted in the light of its limitations. Other natural image statistics such as contrast energy and spatial coherence have also been successful in predicting behavioral ratings of naturalness as well as single-trial variability of the EEG response to individual scenes (Groen et al., 2012, 2013). Structural aspects such as curvature and interactions between edges may also be important (Walther and Shen, 2013; Choo et al., 2014). Second, important advances have been made in computational models of the ventral visual stream (e.g., Serre et al., 2007). These models are variants of neural networks built by units in each layer of the network pooling outputs of units from the previous layers. Such



models have been able to successfully predict representational similarities in various ventral areas in the inferior temporal cortex (IT) (Yamins et al., 2014). Third, the discriminability of scenes is a function of the specific task. For instance, Sofer et al. (2015) show that by modeling the discriminability of natural scenes in superordinate vs. basic-level scene categorization tasks, it is possible to explain the variance of behavioral performance across subjects. Thus, by characterizing the timing of neural decoders across multiple tasks, we can further seek to understand the neural basis of scene categorization. All these directions provide important candidates for future encoding models of MEG responses.

To conclude, the current study brought together behavioral, computational and neural correlates of rapid-scene gist categorization and novel data analysis techniques to provide a spatiotemporal neural portrait of it. Future theories of rapid-scene gist categorization could be constrained by testing a wide range of computational models of scene perception and visual feature spaces and using this analysis framework.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2016.03.027>.

### Author Contributions

P.R., B.C.H., S.P. and L.C.L. conceived of the study.  
B.C.H. and L.C.L. designed the stimuli.  
S.P. performed the experiments.  
P.R. and B.C.H. proposed the analysis methods and analyzed the data.  
P.R., B.C.H., S.P. and L.C.L. interpreted the results.  
P.R., B.C.H., S.P. and L.C.L. co-wrote the manuscript.

### Acknowledgments

We thank Adam Larson, Alexa Lee, and Stewart Lanphier at Colgate University for help with experiments and data analysis. We are grateful to Riitta Hari at Aalto University for the opportunity to perform MEG experiments. We thank Konrad Kording at Northwestern University for the opportunity to use high performance computing allocations to perform data analysis.

P.R. was supported by the Finnish Graduate School of Neuroscience, and European Research Council (Advanced Grant #232946 to R. Hari); B.C.H. was supported by Colgate Research Council and James S McDonnell Foundation (220020430); S.P. was supported by the FP7-PEOPLE-2009-IEF program [#254638]; L.C.L. was supported by the Office of Naval Research [#10846128].

### References

- Bar, M., 2004. Visual objects in context. *Nat. Rev. Neurosci.* 5 (8), 617–629.
- Bastin, J., Vidal, J.R., Bouvier, S., Perrone-Bertolotti, M., Bénis, D., Kahane, P., David, O., Lachaux, J.P., Epstein, R.A., 2013. Temporal components in the parahippocampal place area revealed by human intracerebral recordings. *J. Neurosci.* 33 (24), 10123–10131.
- Cant, J.S., Goodale, M.A., 2011. Scratching beneath the surface: new insights into the functional properties of the lateral occipital area and parahippocampal place area. *J. Neurosci.* 31 (22), 8248–8258.
- Cecotti, H., Marathe, A., Ries, A.J., 2015. Optimization of single-trial detection of event-related potentials through artificial trials. *IEEE Trans. Biomed. Eng.* 62, 2170–2176.
- Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2 (3).
- Choo, H., Shen, D., Walther, D., 2014. Local structure drives human scene categorization: converging evidence from computational analysis, behavior, and neural decoding. *J. Vis.* 14 (10), 1124.
- Cichy, R.M., Pantazis, D., Oliva, A., 2014. Resolving human object recognition in space and time. *Nat. Neurosci.* 17 (3), 455–462.
- Crouzet, S.M., Kirchner, H., Thorpe, S.J., 2010. Fast saccades toward faces: face detection in just 100 ms. *J. Vis.* 10 (4), 16.
- Dilks, D.D., Julian, J.B., Paunov, A.M., Kanwisher, N., 2013. The occipital place area is causally and selectively involved in scene perception. *J. Neurosci.* 33 (4), 1331–1336.
- Ehinger, K.A., Xiao, J., Torralba, A., Oliva, A., 2011. Estimating Scene Typicality from Human Ratings and Image Features.
- Epstein, R., Kanwisher, N., 1998. A cortical representation of the local visual environment. *Nature* 392 (6676), 598–601.
- Epstein, R.A., 2008. Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn. Sci.* 12 (10), 388–396.
- Epstein, R.A., Higgins, J.S., 2007. Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cereb. Cortex* 17 (7), 1680–1693.
- Epstein, R.A., Julian, J.B., 2013. Scene areas in humans and macaques. *Neuron* 79 (4), 615–617.
- Epstein, R., Graham, K.S., Downing, P.E., 2003. Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron* 37 (5), 865–876.
- Epstein, R., Harris, A., Stanley, D., Kanwisher, N., 1999. The parahippocampal place area: recognition, navigation, or encoding? *Neuron* 23 (1), 115–125.
- Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., Lin, C.-J., 2008. LIBLINEAR: a library for large linear classification. *J. Mach. Learn. Res.* 9, 1871–1874.
- Fei-Fei, L., Vanrullen, R., Koch, C., Perona, P., 2002. Rapid natural scene categorization in the near absence of attention. *Proc. Natl. Acad. Sci.* 99 (14), 9596–9601.
- Fei-Fei, L., Iyer, A., Koch, C., Perona, P., 2007. What do we perceive in a glance of a real-world scene? *J. Vis.* 7 (1), 10.
- Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P., Rossion, B., 2005. Diagnostic colours contribute to the early stages of scene categorization: behavioural and neurophysiological evidence. *Vis. Cogn.* 12 (6), 878–892.
- Greene, M.R., Oliva, A., 2009a. The briefest of glances the time course of natural scene understanding. *Psychol. Sci.* 20 (4), 464–472.
- Greene, M.R., Oliva, A., 2009b. Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cogn. Psychol.* 58 (2), 137–176.
- Groen, I.L., Ghebreab, S., Lamme, V.A., Scholte, H.S., 2012. Spatially pooled contrast responses predict neural and perceptual similarity of naturalistic image categories. *PLoS Comput. Biol.* 8 (10), e1002726.
- Groen, I.L., Ghebreab, S., Prins, H., Lamme, V.A., Scholte, H.S., 2013. From image statistics to scene gist: evoked neural activity reveals transition from low-level natural image structure to scene category. *J. Neurosci.* 33 (48), 18814–18824.
- Hämäläinen, M.S., Ilmoniemi, R.J., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. *Med. Biol. Eng. Comput.* 32 (1), 35–42.
- Hämäläinen, M.S., Hari, R., Ilmoniemi, R.J., Knuutila, J., Lounasmaa, O.V., 1993. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Mod. Phys.* 65 (2), 413.
- Hansen, B.C., Loschky, L.C., 2013. The contribution of amplitude and phase spectra defined scene statistics to the masking of rapid scene categorization. *J. Vis.* 13 (13), 1–21. <http://dx.doi.org/10.1167/13.13.21>.
- Hansen, B.C., Jacques, T., Johnson, A.P., Elleberg, D., 2011. From spatial frequency contrast to edge preponderance: the differential modulation of early visual evoked potentials by natural scene stimuli. *Vis. Neurosci.* 28 (03), 221–237.
- Hansen, B.C., Johnson, A.P., Elleberg, D., 2012. Different spatial frequency bands selectively signal for natural image statistics in the early visual system. *J. Neurophysiol.* 108 (8), 2160–2172.
- Hegdè, J., 2008. Time course of visual perception: coarse-to-fine processing and beyond. *Prog. Neurobiol.* 84 (4), 405–439.
- Joubert, O.R., Rousselet, G.A., Fize, D., Fabre-Thorpe, M., 2007. Processing scene context: fast categorization and object interference. *Vis. Res.* 47 (26), 3286–3297.
- Kanwisher, N., Dilks, D.D., 2013. The functional organization of the ventral visual pathway in humans. *New Vis. Neurosci.* 733–748.
- Kay, K.N., Naselaris, T., Prenger, R.J., Gallant, J.L., 2008. Identifying natural images from human brain activity. *Nature* 452 (7185), 352–355.
- Kersten, D., Mamassian, P., Yuille, A., 2004. Object perception as Bayesian inference. *Annu. Rev. Psychol.* 55, 271–304.
- Kirchner, H., Thorpe, S.J., 2006. Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vis. Res.* 46 (11), 1762–1776.
- Klein, A., Tourville, J., 2012. 101 Labeled brain images and a consistent human cortical labeling protocol. *Front. Neurosci.* 6.
- Kravitz, D.J., Peng, C.S., Baker, C.I., 2011. Real-world scene representations in high-level visual cortex: it's the spaces more than the places. *J. Neurosci.* 31 (20), 7322–7333.
- Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. *Proc. Natl. Acad. Sci. U. S. A.* 103 (10), 3863–3868.
- Lin, F.H., Witzel, T., Ahlfors, S.P., Stufflebeam, S.M., Belliveau, J.W., Hämäläinen, M.S., 2006. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *NeuroImage* 31 (1), 160–171.
- Linsley, D., MacEvoy, S.P., 2014. Evidence for participation by object-selective visual cortex in scene category judgments. *J. Vis.* 14 (9). <http://dx.doi.org/10.1167/14.9.19>.
- Loschky, L.C., Larson, A.M., 2010. The natural/man-made distinction is made before basic-level distinctions in scene gist processing. *Vis. Cogn.* 18 (4), 513–536.
- Loschky, L.C., Hansen, B.C., Sethi, A., Pydimari, T., 2010. The role of higher-order image statistics in masking scene gist recognition. *Atten. Percept. Psychophys.* 72 (2), 427–444.
- Loschky, L.C., Sethi, A., Simons, D.J., Pydimari, T., Ochs, D., Corbeille, J., 2007. The importance of information localization in scene gist recognition. *J. Exp. Psychol. Hum. Percept. Perform.* 33 (6), 1431–1450.
- MacEvoy, S.P., Epstein, R.A., 2009. Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Curr. Biol.* 19 (11), 943–947.
- MacEvoy, S.P., Epstein, R.A., 2011. Constructing scenes from objects in human occipitotemporal cortex. *Nat. Neurosci.* 14 (10), 1323–1329.
- Mosher, J., Baillet, S., Darvas, F., Pantazis, D., Yildirim, E., Leahy, R., 2005. Brainstorm electromagnetic imaging software. 5th International Symposium on Noninvasive Functional Source Imaging within the Human Brain and Heart (NFSI 2005) (May).
- Nasr, S., Liu, N., Devaney, K.J., Yue, X., Rajimehr, R., Ungerleider, L.G., Tootell, R.B., 2011. Scene-selective cortical regions in human and nonhuman primates. *J. Neurosci.* 31 (39), 13771–13785.
- Oliva, A., Schyns, P.G., 2000. Diagnostic colors mediate scene recognition. *Cogn. Psychol.* 41 (2), 176–210.
- Oliva, A., Torralba, A., 2001. Modeling the shape of the scene: a holistic representation of the spatial envelope. *Int. J. Comput. Vis.* 42 (3), 145–175.

- Oliva, A., Torralba, A., 2006. Building the gist of a scene: the role of global image features in recognition. *Prog. Brain Res. Spec. Issue Vis. Percept.* 155, 23–36.
- Park, S., Brady, T.F., Greene, M.R., Oliva, A., 2011. Disentangling scene content from spatial boundary: complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *J. Neurosci.* 31 (4), 1333–1340.
- Park, S., Konkle, T., Oliva, A., 2014. Parametric coding of the size and clutter of natural scenes in the human brain. *Cereb. Cortex* (bht418).
- Potter, M.C., 1976. Short-term conceptual memory for pictures. *J. Exp. Psychol. Hum. Learn. Mem.* 2 (5), 509.
- Potter, M.C., Wyble, B., Haggmann, C.E., McCourt, E.S., 2014. Detecting meaning in RSVP at 13 ms per picture. *Atten. Percept. Psychophys.* 76 (2), 270–279.
- Ramkumar, P., Jas, M., Pannasch, S., Hari, R., Parkkonen, L., 2013. Feature-specific information processing precedes concerted activation in human visual cortex. *J. Neurosci.* 33 (18), 7691–7699.
- Ranganath, C., Ritchey, M., 2012. Two cortical systems for memory-guided behaviour. *Nat. Rev. Neurosci.* 13 (10), 713–726.
- Ringach, D.L., 2002. Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *J. Neurophysiol.* 88 (1), 455–463.
- Serre, T., Oliva, A., Poggio, T., 2007. A feedforward architecture accounts for rapid categorization. *Proc. Natl. Acad. Sci.* 104 (15), 6424–6429.
- Sofer, I., Crouzet, S.M., Serre, T., 2015. Explaining the timing of natural scene understanding with a computational model of perceptual categorization. *PLoS Comput. Biol.* 11 (9), e1004456.
- Taulu, S., Simola, J., 2006. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Phys. Med. Biol.* 51 (7), 1759.
- Thorpe, S., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. *Nature* 381 (6582), 520–522.
- Thurgood, C., Whitfield, T.A., Patterson, J., 2011. Towards a visual recognition threshold: new instrument shows humans identify animals with only 1 ms of visual exposure. *Vis. Res.* 51 (17), 1966–1971.
- Vanrullen, R., Thorpe, S.J., 2001. The time course of visual processing: from early perception to decision-making. *J. Cogn. Neurosci.* 13 (4), 454–461.
- Walther, D., Shen, D., 2013. Structural, not spectral properties underlie human categorization of natural scenes. *J. Vis.* 13 (9), 1054.
- Walther, D.B., Caddigan, E., Fei-Fei, L., Beck, D.M., 2009. Natural scene categories revealed in distributed patterns of activity in the human brain. *J. Neurosci.* 29 (34), 10573–10581.
- Walther, D.B., Chai, B., Caddigan, E., Beck, D.M., Fei-Fei, L., 2011. Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proc. Natl. Acad. Sci.* 108 (23), 9661–9666.
- Yamins, D.L., Hong, H., Cadieu, C.F., Solomon, E.A., Seibert, D., DiCarlo, J.J., 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci.* 111 (23), 8619–8624.