Abstract

How does the brain recognize natural scenes? We approach this question by decoding natural scene categories from cortically-localized magnetoencephalography (MEG) responses to briefly presented scenes. We apply spatiotemporal searchlight decoding over the whole brain to study how basic- and superordinate-level category information is processed. We report spatiotemporal differences in the extent to which brain activity informs us about category-sensitive information at the two different levels. In particular, regions previously implicated in scene perception were involved at both levels, but the temporal sequence of information processing in these regions were very different as inferred from basic- and superordinate-level decoding accuracies. These fine-grained differences in temporal processing may shed light on the respective roles of individual cortical areas in scene perception.

1. Introduction

Humans are known to recognize natural scenes with high accuracies from extremely brief presentations (< 20 ms; [14, 12, 4, 22]). Our knowledge of how humans achieve this remarkable behavior comes from studies that examine agreement between behavior and computational representations of scene attributes (see e.g. [4]). To complement this knowledge, a comprehensive portrait of brain activity during scene recognition, coupled with an understanding of functional roles of different brain regions would help constrain computational theories of scene recognition. Functional magnetic resonance imaging (fMRI) studies have identified specialized scene-processing areas in the brain with distinct functional roles [2, 15] (see Figure 1). Yet, much less is known about the precise time course of how natural-scene information is processed in these regions.

Magnetoencephalography (MEG) is a non-invasive brain imaging technique that measures weak magnetic fields outside the head [6]. With knowledge of cortical structure, head geometry and assumptions about conductivity of brain tissue, skull and scalp, it is possible to invert these measured fields to obtain an estimate of the neural current source distribution that generated the measurements [6, 1]. Owing to its submillisecond-scale temporal resolution and millimeter-scale spatial resolution in sensory cortices, when combined with suitable inverse modeling and machine learning techniques, MEG is ideally suited to obtain a high-resolution portrait of neural information processing during a variety of tasks.

In this study, we recorded MEG responses from human subjects when they were asked to report the category of a scene after a very brief presentation. Scenes were drawn from one of six categories, three of which were natural, and the others, man-made, but subjects were asked to report the basic-level category label such as “forest” or “airport” in a six-alternative forced choice task. We then decoded scene category from spatiotemporal windows of inferred cortical current sources[18] to produce a comprehensive description of scene-category-specific information processing. In particular, we asked when and where in the brain, the distinction between basic-level (e.g. forests vs. airports) and superordinate-level (e.g. natural vs. man-made) scene categories are most apparent, by building separate decoders for each categorization.

2. Materials and Methods

2.1. Stimuli and Subjects

We presented 736 x 736 pixel grayscale images of natural scenes from one of six possible categories. Three categories were natural scenes (forests, coasts, or mountains) and the other three were man-made scenes (airports, cities, or suburbs). In each trial, scenes were flashed rapidly for for 33 ms. 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. Each category comprised 30 unique scenes (for details, see [11, 23]).

2.2. Measurements

We recorded MEG data using a Vectorview system (Elekta Oy, Helsinki, Finland) comprising 306 channels providing whole-head coverage. Sensor-level data were preprocessed using temporal signal space separation—a 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[29].

2.3. Source modeling

Measured magnetic fields can be “inverted” using Maxwell’s equations to obtain an inverse image of current sources in the brain that generated the observed measurements. We projected single-trial MEG sensor data to the cortical surface using a standard inverse-modeling technique called minimum norm estimation[7]. The cortical surface was reconstructed from single-subject MR images and a cortically-constrained source space was defined with an average spacing of 5 mm between source vertices. A boundary-element model was used to compute the forward model from current sources to measured fields. We then computed an inverse solution using a standard whitened, depth-weighted linear minimum-norm estimate. The entire inverse modeling was accomplished using the Brainstorm software package[18, 28].

2.4. Spatiotemporally-resolved decoding

For each spatiotemporal window, defined as a neighborhood of 25 vertices around each vertex and a temporal window of 20 ms trailing each time point, we built two linear support vector machine (SVM) decoders—one for the basic level, and the other for the superordinate level of categorization. Decoding accuracies were estimated using 5-fold cross-validation and their 95% confidence intervals were estimated by bootstrapping on the test set. Group-level decoding accuracies were obtained by averaging the single-subject median accuracies and were then thresholded at the respective chance levels (1/6 for the basic level and 1/2 for the superordinate level).

3. Results

We found differential basic-level and superordinate-level diagnostic information in the form of above chance-level decoding accuracies across various cortical regions implicated in scene processing including the perirhinal cortex (PRC), the orbital frontal cortex (OFC), the parahippocampal cortex (PHC), and the retrosplenial cortex (RSC) (see Figure 1 for an illustration of these regions on the inflated medial brain surface).

Given that early visual cortices represent natural image statistics [9, 19] and that low-level natural image statistics can discriminate between basic level scene categories well [20], we may expect that even though they are not functionally specialized for scene processing, they do encode scene category information. Indeed, we found that early visual cortices (striate and ventro-medial extrastriate regions) were sensitive to both basic and superordinate scene categories starting earlier than 100 ms (Figure 3). In the early visual cortices, we found that the burst of basic-level information processing peaked as early as 150 ms whereas superordinate-level information processing peaked only around 200 ms (Figure 2). By contrast, in the medial temporal cortex, the superordinate decoding accuracies rose faster and reach an initial peak in the left hemisphere by 100 ms, which is notably faster than the basic level in the left hemisphere. A somewhat similar, though less dramatic pattern was found in the right hemisphere (compare top and bottom panels in Figure 2).

For basic-level categorization, our decoders revealed a spatiotemporal information sequence that began with bilateral striate cortex around 90 ms (Figure 3). From there, maximum accuracies were observed bilaterally in the OFC around 100 ms. This was followed at roughly 150 ms by peak PRC and PHC accuracies in the left hemisphere and PRC, PHC, and RSC peak accuracies in the right hemisphere. Around 250 ms, bilateral peak accuracies were observed in the PRC, with OFC accompanying that peak in the left hemisphere. Later, around 350 ms, lateralized effects were observed—the PHC and RSC peaked in the left hemisphere, and OFC peaked in the right hemisphere (Figures 2 and 3).

For the superordinate categorization level, we observed different spatiotemporal decoding accuracies. The early (90 ms) bilateral classification accuracy in striate cortex, was followed (at 100 ms) by bilateral OFC accuracy peaks (as was observed for basic-level categorization). However, two subsequent peaks were observed in the left hemisphere. The first (10 ms after the OFC peak) was the PRC, which was then followed (10 ms after the PRC peak) by a PHC peak. This systematic three peak shift (OFC-to-PRC-to-PHC) was not observed in the right hemisphere, which instead consisted mainly of a PHC peak at 150 ms. Later, around 250 – 300 ms, the OFC accuracies peaked in the left hemisphere, followed by the PRC around 350 – 400 ms. However, in the right hemisphere, the RSC accuracies peaked during the 250 – 300 ms time window, followed by both RSC and PHC around 350 – 400 ms (Figures 2 and 3).

In summary, category-diagnostic signals were present in the OFC, PRC, PHC, and RSC for both levels of classification, with the OFC involved early on for both. However, for
Figure 1: An illustration of the key areas involved in scene perception in the medial temporal and frontal lobes: orbitofrontal cortex (OFC), perirhinal cortex (PRC), parahippocampal cortex (PHC), and retrosplenial cortex (RSC).

Figure 2: Time courses of decoding accuracies for the basic-level (above) and superordinate-level (below) decoding tasks from the left (left) and right (right) medial surfaces, respectively, at four regions of interest (OFC, PRC, PHC, RSC).
basic-level categorization, the 100 ms OFC was mostly followed by simultaneous PRC and PHC peaks, whereas for superordinate-level classification, the early OFC was followed by mostly systematic lagged PRC and PHC (respectively). Also, bilateral PRC peaked around 250 ms for basic level categorization, whereas superordinate accuracies suppressed PRC and PHC. Interestingly, around 350–400 ms, both basic and superordinate classification accuracies showed prominent PHC and RSC peaks, but lateralized to the left or right hemispheres respectively.

4. Discussion

We have shown earlier that scene categories can be decoded in a time-resolved manner from single-trial sensor-level MEG data [23]. Here, we present for the first time, a comprehensive spatiotemporal description of scene-category-specific information processing in the brain at the cortical level and at the millisecond timescale. By decoding natural scene categories from cortically-localized MEG responses, we reveal the category-specific neural code for basic- and superordinate-level representations in the context of a basic-level categorization task.

Behavioral studies have shown that we are able to identify the gist of a scene (basic-level scene category) from very short exposures although longer exposures are needed to be able to name objects and describe their relationships [3]. Here, we observed sequential category-specific information processing in the OFC, followed by the object-centric area PRC and subsequently, the PHC; but this was restricted to superordinate-level categorization. These regions have been implicated in different functions related to object and scene perception, and it is known that they have different anatomical connectivity with each other, with parts of the hippocampus, the medial prefrontal cortex and the dorsal and ventral visual areas [24]. Although a spate of recent studies has examined their functional properties using fMRI [33, 15, 10, 17, 30, 31, 27], their exact involvement in scene processing is still unresolved. Adding relative timing information as we have done here using MEG, might help narrow down their exact role.

One caveat in our experiment is that subjects were performing the basic-level categorization only. Therefore, it is unclear whether early basic-level category-diagnostic information in OFC reflects task demands, or a general functional property of OFC regardless of the goal of viewing natural scenes.

It has remained an important debate as to whether basic-level or superordinate-level distinctions are made earlier in the brain. Recent studies [13, 21, 8, 16] have challenged earlier findings [32, 25] that basic-level categorization takes place before superordinate-level categorization. At the computational level, the more abstract superordinate level coding between "natural" vs. "man-made" is easier to make than the basic level coding based on the amplitude spectra of natural scenes [20]. At the behavioral level, several studies have shown greater accuracy after less processing time (measured either by manual reaction times or masking stimulus-onset asynchrony) for the superordinate than for the basic level [4, 13, 26]. Here, we present neural evidence – in particular, differential timing for information processing across the two categorical levels – that may add a novel perspective to this debate.

In future work, our goal is to understand not only whether or not a certain spatiotemporal pattern of brain activity carries diagnostic information, but furthermore what that information might be. To this end, we intend to reverse-engineer the visual features and other scene attributes that are most diagnostic of scene categories (see e.g. [5, 34]) and then build generative models of brain activity using these attributes.

References

Figure 3: Representative temporal snapshots from a group average \((N = 8\) subjects) map of basic-level (left) and super-ordinate (right) median decoding accuracies. Median and 95\% CIs were estimated by 5-fold cross-validation followed by bootstrapping on all the held-out test sets.


[21] M. Poncelet, L. Reddy, and M. Fable-Thorpe. A need for more information uptake but not focused attention to access basic-


