## Chapter 8

# The functional organization of the ventral visual pathway and its relationship to object recognition

Kalanit Grill-Spector

**Abstract**

Humans recognize objects at an astonishing speed and with remarkable ease. However, the functional organization of the system that enables this remarkable human ability is not well understood. Here we examine whether the human ventral stream is organized more around stimulus content or recognition task. We scanned subjects while they performed one of two tasks: object detection (objects versus textures) or subordinate identification (e.g. pigeons versus birds). In order to limit success at recognition, pictures were presented briefly and then masked. For each subject we searched for cortical regions where activity was correlated with correct answers, separately for each category and task. Analysis by task revealed that, for each category, regions correlated with correct detection and correct identification were similar. However, analysis by stimulus revealed that different patterns of activation across occipitotemporal areas were correlated with successful identification of different categories. Analysis of regions whose activity was correlated with face recognition revealed a higher signal for faces (compared to birds and guitars) only in trials in which faces were perceived, but not in trials in which faces were not detected. Overall, these results indicate that the functional organization of higher-order areas in the human ventral stream is organized more around stimulus content than recognition task. These results provide new insights into the representations underlying our ability visually to recognize objects.

## 8.1 Introduction

Humans recognize objects and faces instantly and effortlessly. What are the underlying neural mechanisms in our brains that allow us to detect and discriminate among objects so efficiently?

Multiple ventral occipitotemporal regions anterior to retinotopic cortex (Grill-Spector *et al.* 1998) respond preferentially to various objects compared to textures (Malach *et al.* 1995). Indeed, several studies provide evidence that the activation of occipitotemporal object areas is correlated with subjects' perception of objects in a variety of experimental paradigms and tasks (Tong *et al.* 1998; Grill-Spector *et al.* 2000; James *et al.* 2000; Hasson *et al.* 2001; Kleinschmidt *et al.* 2002). Further evidence suggests that these regions play a critical role in object recognition, since lesions to the fusiform gyrus and occipitotemporal junction produce various recognition deficits (Damasio 1990; Damasio *et al.* 1990; Farah 1992), and electrical stimulation of these regions interferes with recognition (Puce *et al.* 1999).

Functional imaging studies have revealed that some of these regions respond maximally to specific object categories, such as faces (Puce *et al.* 1995; Kanwisher *et al.* 1997), places (Aguirre *et al.* 1998; Epstein and Kanwisher 1998), body parts (Downing *et al.* 2001), letter strings (Puce *et al.* 1996), tools (Martin *et al.* 1996), and animals (Martin *et al.* 1996; Chao *et al.* 1999). These results suggest that areas that elicit a maximal response for a particular category are dedicated to the recognition of that category. However, there are many difficulties underlying this idea. First, comparing activation between a handful of object categories is problematic because it depends on the choice of categories. Secondly, while there is maximal activation to one category the activation to other categories is not negligible (Ishai *et al.* 1999; Haxby *et al.* 2001). Thirdly, comparing the amplitude of activation to object categories does not exclude the possibility that the underlying representation might not be of whole objects (Fujita *et al.* 1992; Grill-Spector *et al.* 1998; Lerner *et al.* 2001; Tsunoda *et al.* 2001). Finally, objects from different categories differ in many dimensions and it is possible that the source of higher activation for a category is not restricted to visual differences.

How is the functional organization of these object-selective regions related to our ability to recognize objects? One view, proposed by Kanwisher (2000), is that the ventral temporal cortex contains a limited number of modules specialized for the recognition of special categories, such as faces, places, and body parts, and the remaining cortex, which exhibits little selectivity for particular object categories, is a general-purpose mechanism for the perception of any shape of any kind of visually presented object. A second model, proposed by Haxby *et al.* (2001), is an 'object form topography' in which occipitotemporal cortex has a topographically organized representation of form attributes. The representation of an object is reflected by a distinct pattern of response across ventral cortex, and this distributed activation produces the visual percept. In contrast to the Kanwisher model, Haxby and colleagues propose that submaximal activations across the ventral stream may be as important as the maximal activations (see also Avidan *et al.* 2002). A third view, posited by Tarr and Gauthier (2000), is that the organization is based according to the perceptual processes carried out and not by the content of information processed. Here the fundamental idea is that different cognitive processes require different computations, that are instantiated in different parts of the visual cortex.

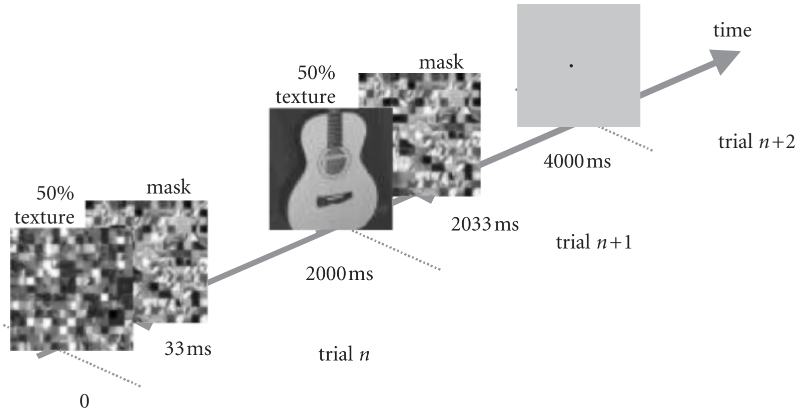
Here we examine whether the functional organization of object-selective regions in occipitotemporal cortex is based on the content of information processed or on computations dedicated to specific perceptual processes. To distinguish between these alternatives we scanned subjects while they performed different recognition tasks: object detection or subordinate-level identification (Rosch *et al.* 1976) on three object categories—faces, birds, and guitars. We asked two experimental questions: (1) within each task, are the same or different regions correlated with success at recognizing different categories; and (2) for each stimulus category, are the same or different regions correlated with success at different recognition tasks?

## 8.2 **Parsing object recognition into component stages**

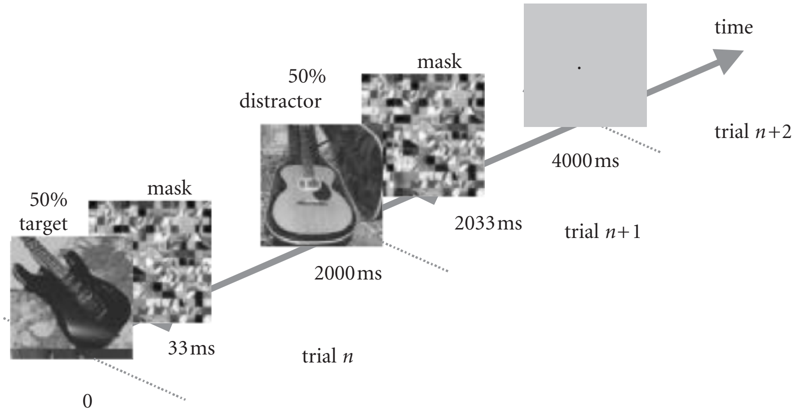
What are the stages of processing involved in visual object recognition? Hierarchical models of object recognition suggest that recognition involves several processing steps, proceeding from low-level stages that extract local visual information about features (Livingstone and Hubel 1988; Gallant *et al.* 1993), contours (von der Heydt *et al.* 1984), and boundaries (Lamme 1995; Zhou *et al.* 2000), to high-level stages that perform recognition by matching the incoming visual stimulus to stored representations of objects. Many models of recognition posit an intermediate stage at which the object is segmented from the rest of the image. Underlying this idea is the intuition that an efficient object recognition system should not operate indiscriminately on just any region of an image, because most such regions will not correspond to distinct objects. Instead, researchers have argued that stored object representations should be accessed only for image regions selected as candidate objects by a prior image segmentation process (Rubin 1958; Nakayama *et al.* 1995; Driver and Baylis 1996). However, other evidence (Peterson 1994) suggests that object recognition may influence segmentation, and may perhaps precede image segmentation.

To tease apart processing stages involved in visual recognition we varied stimulus exposure duration and measured behavioral performance on three different recognition tasks, each designed to tap into a different candidate stage of object recognition: detection, categorization, and identification. The *detection task* was designed to be a minimalist test of object segmentation that does not require recognition. Here subjects were asked to decide whether a gray-scale image contained an object or not. They were told that they did not have to recognize the object to report its presence. Half of the trials contained objects from 10 categories and half of the trials consisted of texture patterns created by randomly scrambling object images to 225 squares to equate mean luminance and local low-level features (see Fig. 8.1a). The second task was *object categorization*, in which subjects were required to categorize the object in the picture from a set of 10 possible categories: face, bird, dog, fish, flower, house, car, boat, guitar,

(a) Detection task: ‘object’ versus ‘not an object’



(b) Identification task: ‘electric guitar’ versus ‘not an electric guitar’



(c) Examples of stimuli

**Fig. 8.1** Experimental design; (a) Detection task: in each trial (of duration 2 s) an image was presented briefly and then masked. Subjects had to respond during the duration of the mask, whether the image contained an object or not. In each scan, subjects were presented with objects from one category (in this case guitars) but the subjects were not told in advance the content of the pictures. Trials of objects, textures, and blanks (no visual stimulation) were counterbalanced. (b) Identification task. Here all pictures contained objects from a single basic

**Table 8.1** Experiment 1: Behavioral data

| | **Detection** | **Identification** |
|---|---|---|
| Accuracy corrected for guessing | | |
| faces | 67 ± 6 | 32 ± 6 |
| birds | 70 ± 6 | 40 ± 6 |
| guitars | 67 ± 6 | 32 ± 5 |
| Reaction times (ms) | | |
| faces | 550 ± 24 | 652 ± 24 |
| birds | 594 ± 29 | 722 ± 40 |
| guitars | 586 ± 29 | 711 ± 48 |

Accuracy at identification was significantly lower than detection, as verified via an across-subject *t*-test (faces, $P < 10^{-8}$; birds, $P < 10^{-6}$; guitars, $P < 10^{-7}$) and reaction times were significantly longer ( faces, $P < 10^{-6}$; birds, $P < 10^{-4}$; guitars, $P < 10^{-5}$). There was no statistical significance between identification or detection performance across categories (verified via a *t*-test).

or trumpet. In the *identification task* subjects were instructed to discriminate a particular subordinate member of a category from other members of that category. Possible answers were: Harrison Ford, pigeon, German shepherd, shark, rose, barn, VW beetle, sailboat, electric guitar versus 'other', i.e. other male faces, other birds, other cars, etc. In all three tasks, the frequency of each category was 10% and for each basic-level category half of the images were from a single subordinate class. Images were presented in five durations (between 17 and 167 ms) in a counterbalanced order.

The behavioral data revealed that longer stimulus exposures were required for subjects to reach the same accuracy levels in the identification task compared to the other two tasks (see Table 8.1). Lower accuracy at identification compared to categorization occurred for each of the object categories tested. This indicates that identification occurs after detection and categorization. Surprisingly, the curves relating performance to stimulus duration were nearly identical for the categorization and detection tasks despite the greater complexity of the ten-alternative forced choice categorization task compared to the two-alternative forced choice object detection task (Fig. 8.2). Thus, as soon as subjects detected an object, they already knew its category (Grill-Spector and Kanwisher 2003).

These data suggest that there are at least two main processing stages involved in object recognition: *detection* and *identification*, with detection preceding identification. The question that we will address in the following sections is whether the extra processing needed for identification compared to detection reflects additional processing

---

level category and subjects were required to identify a particular subordinate category. In this example, subjects had to respond whether the object was an electric guitar or not; distractors were other kinds of guitar. (c) Examples of the stimuli used in these experiments.

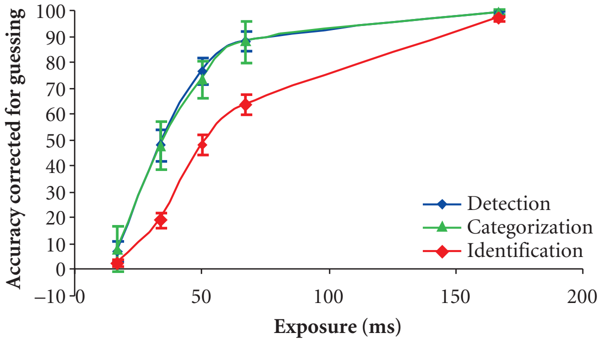


**Fig. 8.2** Behavioral data. Accuracy in both the detection and categorization tasks was significantly higher than identification for stimulus exposures of 33–68 ms ($t$-test, $P < 0.001$); vertical axis denotes accuracy (corrected for guessing) on object pictures. Blue, detection; green, categorization; red, identification. Error bars indicate SEM (standard error of the mean) across 13 subjects.

within the same cortical area, or whether it requires additional processing at subsequent stages.

## 8.3 Regions correlated to detection or identification of three object categories

Our behavioral data indicates that subjects' performance in detection and categorization is similar. Therefore, in the fMRI experiments, we chose to use only two tasks: *detection* and *identification*. Here we asked subjects to perform either a detection task (see Fig. 8.1a) that required subjects to detect the presence of an object without having to recognize it, or an identification task (Fig. 8.1b) that required subjects to discriminate between objects belonging to the same basic level category (Rosch *et al.* 1976). We manipulated subjects' ability to recognize objects by presenting pictures very briefly, for 33 or 50 ms[1] and then masking the images with a texture pattern. Due to the brief visual presentation, in some trials subjects could identify or detect objects, and in others they could not. When subjects viewed this display, we measured both behavioral performance and brain activation in a rapid event-related design experiment, using a 3T fMRI scanner.[2] In contrast to conventional fMRI experiments, in which areas are defined based on their amplitude of activation to different types of stimuli, here we localized areas that were correlated with detection or identification object, using 'individual subjects' behavioral data.

In the *detection task*, subjects were asked to decide whether or not a gray-scale image contained an object. Half the trials contained objects from one basic level category and half the trials contained texture patterns. The *identification task* required subjects to discriminate a particular subordinate member of a category (e.g. electric guitar) from other members of that category (e.g. other guitars) that share a common structure. Here all images belonged to one category (in this example guitars). Half of the images were different pictures of the target subordinate category and half of the images were other objects from the same basic level category. In separate scans subjects were asked to perform one of the two tasks (detection or identification) on one of three object categories: faces, guitars, or birds.

Importantly, in these experiments, in each trial subjects viewed an image they had never seen before, so performance could not be affected by prior knowledge of particular images. Further, objects from each category and subordinate class were depicted in various viewing conditions and with different backgrounds (see Fig. 8.1c) to reduce the probability that subjects would use a small set of low-level features to perform these tasks.

The behavioral performance of subjects is given in Table 8.1. Similar to the ten-category experiment, for all object categories accuracy was significantly lower and reaction times significantly longer for the identification task compared to the detection task. Reaction times for detection were on the average 125 ms shorter than identification, suggesting that it occurs prior to identification. This occured for all three categories, including faces (Tanaka 2001). The differences in performance in the detection and identification tasks indicate that subjects were indeed performing different perceptual tasks during these scans.

## 8.4 **Detection experiment**

For each subject we searched for regions that were correlated with successful detection separately for each category. Thus, we ran a statistical test searching for regions that showed a higher signal in trials in which objects were present and subjects successfully detected their presence (hits) compared to trials in which objects were present but subjects failed to detect their presence (misses). Texture stimuli were not included in the statistical analysis. This analysis was performed individually for each subject on a voxel by voxel basis. Importantly, in all trials a picture of an object from the same basic level category was shown for the same exposure duration. The only difference between trials was whether subjects succeeded or failed to detect the object.

For each of the categories, we found regions correlated with successful detection (examples of maps for faces and guitars are given in Fig. 8.3). These included the lateral occipital cortex (LO) and ventral occipitotemporal (VOT) areas including the occipitotemporal sulcus (OTS) and the fusiform gyrus. While we found for each object category regions that were correlated with successful detection, the pattern of activation across the human ventral stream was different for different categories. Hence, different subregions within higher-level areas were correlated with the detection of different categories.

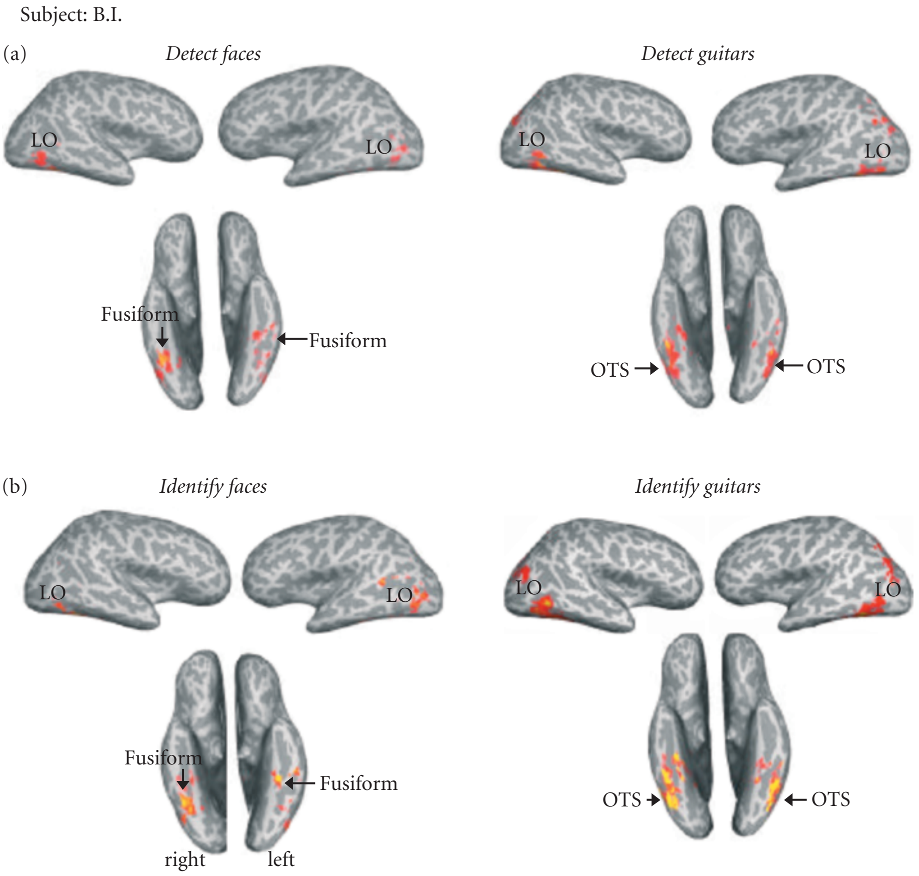


**Fig. 8.3** Areas correlated with correct detection and identification. (a) Areas correlated with hits > misses in the detection task shown on the inflated brain for one representative subject for two of the categories tested. Statistical analysis was performed using FS-fast software developed at MGH and brainalyzer software written by K.G.S. Brain reconstruction was performed using Freesurfer (Dale *et al*. 1999; Fischl *et al*. 1999, 2001). Color code indicates statistical significance (yellow, $P < 10^{-4}$; red: $P < 10^{-2}$). (b) Areas correlated with hits > misses in the identification task shown on the inflated brain of the same subject shown in (a). Color code indicates statistical significance (yellow, $P < 10^{-4}$; red, $P < 10^{-2}$).

## 8.5 **Identification experiment**

We performed a similar analysis for the identification experiment. Here we searched for regions that showed a higher signal in trials in which subjects were successful at identification. 'Hits' were defined as trials in which the target subordinate category was present and subjects answered correctly (e.g. electric guitar present and subjects responded 'electric guitar'). 'Misses' were trials in which the target subordinate category was present, but subjects answered incorrectly (e.g. electric guitar present but subject responded 'not an electric guitar'). Here catch trials contained other objects from the same basic level category (e.g. other guitars), but were not included in the statistical analysis.[3] This analysis was performed independently for each subject and category. Results are given in Fig. 8.3. Similar to the detection experiment, for each object category we found regions in the human ventral stream that were correlated with successful identification. Again, the pattern of activated areas across the human ventral stream that was correlated with successful identification was different for different categories.

## 8.6 Question 1: When we keep the category constant, are the same or different regions correlated with success at different tasks?

For each category we superimposed the maps of regions correlated with successful identification and successful detection and tested for conjunction effects. This superposition shows that when the category was kept constant there was a large degree of overlap between voxels that were correlated with detection and identification (yellow voxels in Fig. 8.4a). Across subjects, 60 ± 7% of the activated voxels were correlated with success at both object detection and object identification. Thus, most of the voxels were correlated with both identification and detection for each of the object categories. Thus, when the category was kept constant, similar regions across the human ventral stream were correlated with successful identification and detection.

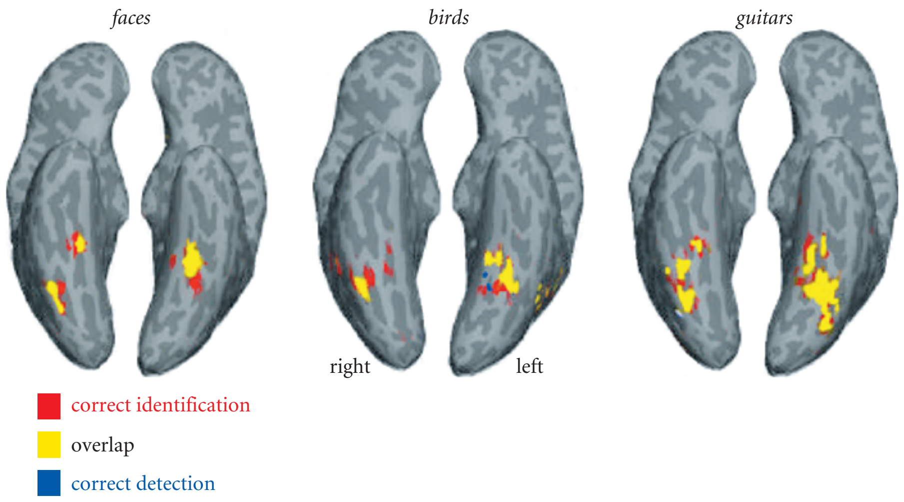
## 8.7 Question 2: When we keep the task constant are the same or different regions correlated with recognition of different object categories?

Next we superimposed maps of areas correlated with identification of faces, guitars, and birds, to test whether these areas are overlapping or distinct (see Fig. 8.4b). When we superimposed maps of areas correlated with the identification of two object categories, most of the voxels were not overlapping. Across subjects, only 33 ± 10% of the activated voxels were correlated with the identification of two different categories. Taliarach coordinates for the center of activated regions in the identification task are given in Table 8.2. Thus, when the task was kept constant, largely different regions across the ventral stream were correlated with success at identifying different object categories.

The results of the superposition analysis reveal that there was approximately twice as much overlap between areas that were correlated with detection and identification of a particular category, than the amount of overlap between areas that were correlated with identification of different object categories. This suggests that the organization in the ventral stream is more around visual content rather than perceptual task.

(a) Overlap by task

Subject: B.P.



(b) Overlap by category: identification task

Subject: B.P.

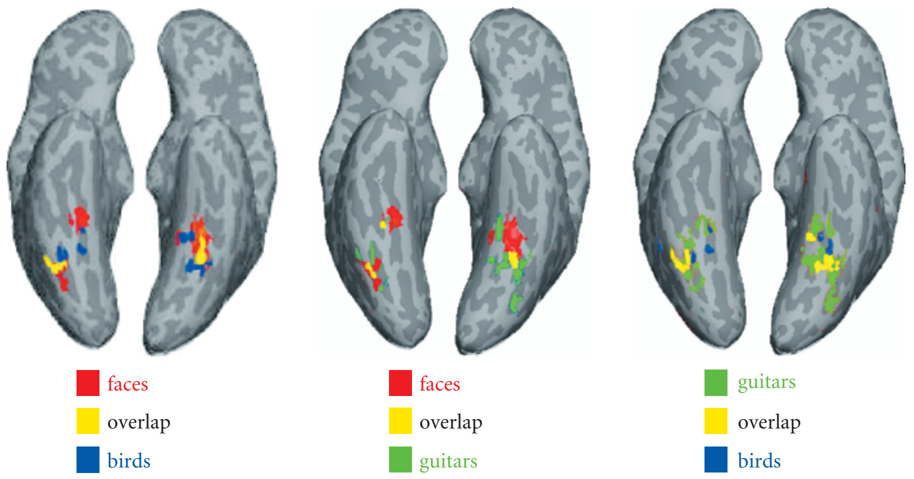


**Fig. 8.4** Overlap analysis. Ventral view of an inflated brain of the same subject in all tasks and categories. (a) Overlap by task: yellow, areas that were correlated with both correct identification and correct detection of a given object category; red, areas that were correlated only with correct identification; blue, areas that were linked only to correct detection. (b) Overlap by category: identification task: yellow, areas that were correlated to successful identification of two categories; red, areas that were correlated only with successful face identification; blue, areas that were correlated only with bird identification; green, areas that were correlated only with guitar identification.

**Table 8.2** Talairach coordinates

| | **Right** | | | **Left** | | |
|---|---|---|---|---|---|---|
| Ventral occipto-temporal (VOT) | | | | | | |
| Faces | 39 ± 3 | −49 ± 37 | −16 ± 5 | −37 ± 4 | −50 ± 7 | −14 ± 5 |
| Birds | 41 ± 5 | −52 ± 10 | −20 ± 3 | −38 ± 4 | −55 ± 10 | −17 ± 4 |
| Guitars | 46 ± 2 | −56 ± 8 | −15 ± 6 | −41 ± 6 | −59 ± 8 | −15 ± 5 |
| Lateral occipital (LO) | | | | | | |
| Faces | 45 ± 3 | −77 ± 6 | 2 ± 8 | −48 ± 3 | −76 ± 6 | 6 ± 3 |
| Birds | 48 ± 5 | −71 ± 10 | −9 ± 3 | −46 ± 2 | −74 ± 8 | 6 ± 8 |
| Guitars | 45 ± 6 | −74 ± 5 | −4 ± 8 | −41 ± 4 | −73 ± 5 | −5 ± 8 |

Nevertheless, there were regions that seemed to be correlated with only one recognition task. When we compared between tasks, most of the non-overlapping areas were regions that were correlated with successful identification but not detection (red voxels in Fig. 8.4a). This can be accounted for partially by the larger number of voxels that passed the statistical threshold in the identification task compared to the detection task. The critical question is whether these non-overlapping regions are dedicated to one recognition task?

Another concern in interpreting the overlap between detection and identification is the possibility that subjects were able to identify some of the objects in the detection experiment, even though they were not required to do so. We therefore conducted another experiment, but here we asked subjects to respond for each picture whether they could: (1) identify the object; (2) detect the object but not identify it; or (3) not detect it at all.

## 8.8 Comparing identification and detection directly within the same scan

In this set of experiments, five subjects saw in each scan different pictures from one object category. Half of the images were of the target subordinate category (e.g. different pictures of pigeons) and the rest of the pictures were other images from the same basic level (e.g. other birds). Subjects were asked to answer for each picture whether it was the target subordinate category (e.g. pigeon), or an object but not the target, or not an object. Subjects' behavioral performance is given in Table 8.3.

We first searched for regions that showed a higher signal for detected (but not identified objects) i.e. detection hits, versus not detected objects i.e. detection misses, independently for each object category. In contrast to the first detection experiment, here the detection hits consist of trials in which subjects could detect the presence of an object but could not identify it at the subordinate level. Consistent with the previous experiments, different patterns of activation across the human ventral stream were correlated with detection of different object categories. The time courses extracted from

**Table 8.3** Experiment 2: Behavioral data, percent responses out of all target trials

| | **Identified** | **Detected** | **Not detected** |
|---|---|---|---|
| Faces | 31 ± 6 | 41 ± 5 | 28 ± 6 |
| Birds | 28 ± 6 | 38 ± 3 | 34 ± 6 |
| Guitars | 26 ± 6 | 39 ± 3 | 35 ± 6 |

these areas showed a higher signal amplitude for trials in which objects were detected compared to trials in which objects were not detected (see Fig. 8.5a, left). Importantly, the signal from these voxels was highest for trials in which subjects were successful at identification, even though these trials were not included in the statistical analysis; however, the difference between identified and detected did not reach significance, except for the birds experiment ($P < 0.01$). This provides evidence that areas that were correlated with successful detection of objects were also correlated with successful identification of objects.

The design of this experiment also enabled us to search directly for regions that showed a higher signal for identification hits versus detection hits, or vice versa. We found regions that showed a higher signal for identification hits versus detection hits. They tended to be located in more anterior ventral regions along the fusiform gyrus and OTS, which we refer here as VOT regions. However, the time-course analysis (Fig. 8.5a, right) revealed that while the signal was maximal for identification hits, the signal was statistically significantly higher for detection hits than detection misses (significance verified via a *t*-test: faces, $P < 10^{-8}$; birds, $P < 10^{-3}$; guitars, $P < 10^{-3}$). We also searched for regions that showed a higher signal for 'identified' versus 'detected' for two or more categories. This revealed some activation in the fusiform gyrus. However, time-course analysis of these ROIs revealed that while the signal was highest for identification hits, it was also significantly higher for detection hits than detection misses ($P < 10^{-3}$). Thus, while we found regions that were also correlated with successful identification, the same regions were correlated with successful detection of these categories. We did not find any region that showed only a significant success effect for identification but not for detection. These results provide direct evidence that areas that were correlated with correct object identification were also correlated with correct object detection.

Previous studies (Grill Spector *et al.* 1999) indicate that there are functional differences between lateral object selective foci (LO) and anterior ventral occipitotemporal foci.[4] We therefore defined for each category regions of interest (ROIs) in LO and VOT[5] to test whether there are functional differences between these areas. Note that different foci in the vicinity of LO and VOT were correlated with successful recognition of these object categories. Both LO and VOT foci showed a main effect of success for both detection and identification performance. The difference between ventral and lateral foci was the magnitude of the effect across tasks. VOT foci exhibited a significant signal increase from detection misses to detection hits and also a significant signal increase from detection hits to identification misses. Both of these effects were statistically significant for all categories ($P < 0.01$). In LO the signal increased significantly from not detection misses to detection hits for all categories ($P < 0.001$), but the increase in the signal strength from detection hits to identification hits was statistically significant only for birds ($P < 0.01$). These results may suggest a hierarchy within the ventral stream, with LO regions contributing more to object detection/segmentation and VOT foci involved both in object detection and object identification.

Finally, we directly compared overlap by task and overlap by category in this experiment. Here, we performed the overlap analysis at two threshold levels[6] to ensure that our results do not depend on the choice of the threshold (see Fig. 8.5b). The results demonstrate that for both threshold levels the overlap by task was far greater than the overlap by category. At low thresholds the overlap by task was greater by twofold compared to overlap by category, and in higher thresholds the overlap by task was greater by threefold. These differences were statistical significant (LO, high threshold, $P < 10^{-5}$; low threshold, $P < 10^{-6}$; VOT, high threshold, $P < 10^{-6}$; low threshold, $P < 10^{-7}$). Importantly, at both threshold levels the majority of voxels that were correlated with correct recognition of one category were not the same voxels that were correlated with successful recognition of another category. Note that increasing the threshold level decreases the extent of activated areas. The finding that overlap by category is smaller at higher thresholds suggests that smaller ROI are more homogeneous and thus display a higher degree of category specificity. Thus, this analysis strengthens the conclusion that there is a higher degree of overlap by task compared to overlap by category and further show that this result does not depend on the choice of the threshold.

## 8.9 **False identification**

Our data so far suggest that activation in higher-order areas is predictive of success at object recognition. These regions were not activated to the same degree whenever an object stimulus was present; rather they were activated when the stimulus was there and the subject could report its presence consciously. The false-alarm data provide another demonstration that the activation was correlated with subjects' reports rather than with the presence of the stimulus.

In the first set of experiments we did not have a sufficient number of false-alarm trials to be able to measure a reliable signal. In the second set of experiments, 2 out of 5 subjects produced a sufficient number of false-alarm trials (17% and 22% of the non-target trials) to obtain a measurable signal. False alarms consisted of trials in which objects from the same basic level category as the target category (e.g. birds) were incorrectly identified as the target subordinate category (e.g. pigeons). Interestingly, the signal for falsely identified trials (green in Fig. 8.6) was almost as high as identification

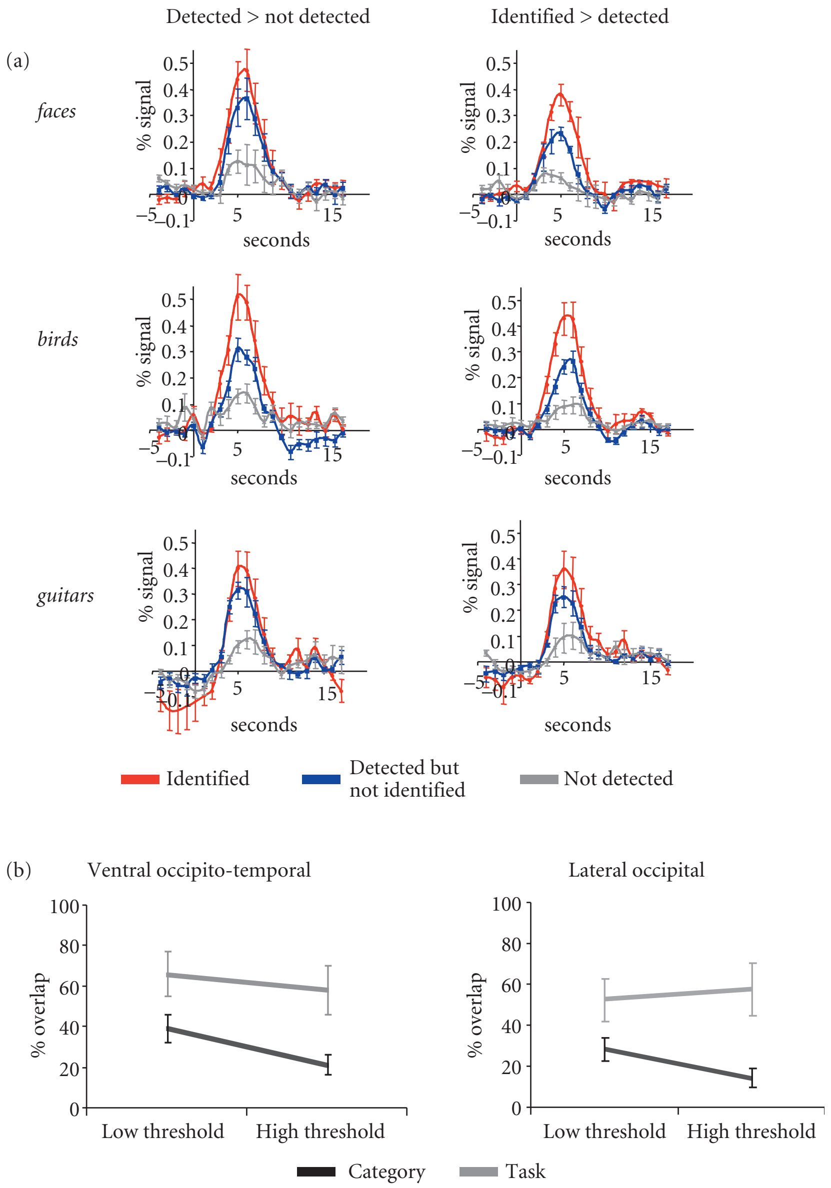


**Fig. 8.5** Comparing identification and detection performance within the same scan. (a) *Left:* time courses from areas that passed the statistical threshold for detected > not detected averaged across five subjects. Areas were defined independently for each subject and category, with a threshold of $P < 0.01$. Trials were not included in the statistical analysis, but the signal for these trials is highest. *Right*: time courses from areas that passed the statistical threshold for identified > detected. Areas were defined independently for each subject and category with a threshold of $P < 0.01$. Trials, although 'not detected' trials were not included in the statistical analysis (faces, $P < 10^{-8}$; birds, $P < 10^{-3}$; guitars, $P < 10^{-3}$). (b) We analyzed

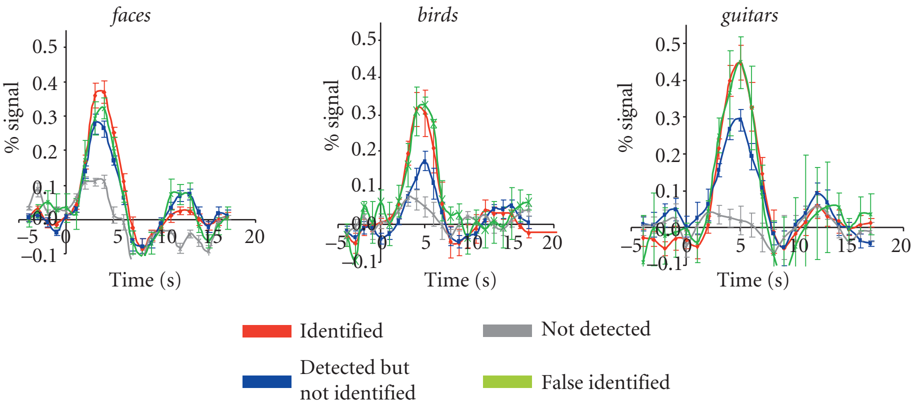


**Fig. 8.6** False identification. Time courses averaged across two subjects for areas that were correlated with identification and detection of the target subordinate category. Falsely identified trials were not included in the statistical analysis. The curve for falsely identified was similar to the curve for identified targets and was higher than the curve for detected but not identified targets. Differences between false identification and identification hits did not reach significance.

hits (red in Fig. 8.6) and significantly stronger than detection hits (blue in Fig. 8.6). Thus, the signal in higher-order visual areas was higher when subjects reported that they identified an object (whether or not their answer was correct) compared to when they detected the presence of an object but could not identify it. This result further supports the idea that the activity in higher-order visual cortex correlates with what subjects reported they perceived rather than what was physically present.

## 8.10 **Is the FFA a module for subordinate recognition?**

One of the major debates regarding specialization based on content or process focuses on the specialization of the fusiform face area (FFA). Two main hypotheses for the role of the FFA in recognition have been proposed. Kanwisher and colleagues have suggested that this is a region specialized for face recognition (Kanwisher *et al.* 1997; Kanwisher 2000). Others, in particular Tarr and Gauthier (2000), have suggested that this is a region for subordinate recognition (of multiple categories) that is

---

separately the amount of overlap by task and overlap by category from lateral foci and ventral occipitotemporal foci in two threshold levels. Numbers are given in percent overlap and are averaged across tasks or categories. Threshold values were: low threshold, $P < 0.01$; high threshold, $P < 0.001$.

automated by expertise. These authors argue that faces (and expert categories) are automatically accessed at the subordinate level and therefore face recognition recruits the FFA.

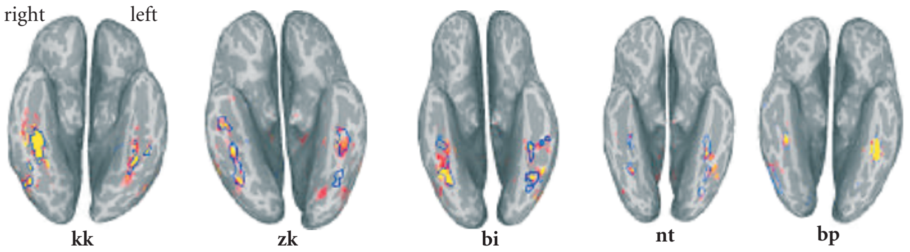
These two hypotheses posit different outcomes for FFA activation in our experiments. The first hypothesis that FFA is a module for face recognition predicts that the activation of the FFA should be correlated with both face identification and face detection but not with identification of other object categories. In contrast, the second hypothesis, that the FFA is involved in subordinate identification, predicts that the activation of the FFA should be linked to successful identification of all categories, but should not be correlated with correct detection of faces (or other objects).

To directly test these hypotheses for each subject, we used an independent localizer scan to define the FFA based on a face selectivity test. In the localizer scan we used a block design experiment in which subjects passively viewed pictures of faces, cars, novel objects (abstract sculptures) in outdoor scenes and textures. The FFA was defined for each subject as the regions in the fusiform gyrus that showed a higher signal for faces compared to cars and novel objects, with a significance level of $P < 10^{-4}$ (see blue in Fig. 8.7).

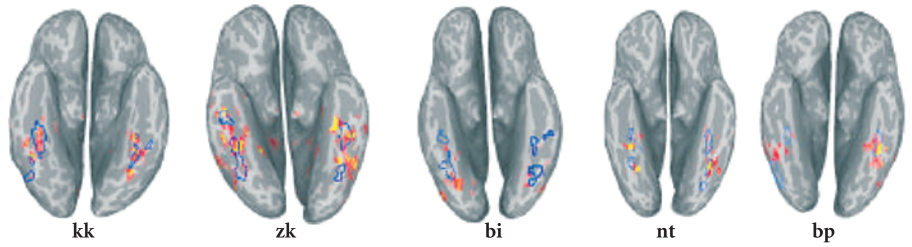
First, we compared the locus of regions activated by correct versus incorrect identification (from the previous experiment) with face selective areas (faces versus objects). This analysis was performed separately for each object category (faces, birds, and guitars). For all five subjects, the regions that were correlated with successful identification of faces (yellow and red voxels in Fig. 8.7a) were similar to regions that were defined as face-selective in the independent localizer scan (blue contours in Fig. 8.7a). This indicates that face-selective regions were involved in face identification. In the bird experiments, there was some degree of overlap between areas correlated with bird identification and the FFA. In contrast, there was very little correspondence between face-selective regions and areas correlated with guitar identification (see Fig. 8.7c). We found partial overlap between the left FFA and areas correlated with guitar identification only in two subjects. This analysis revealed that face-selective regions were always involved in face identification, but not in guitar identification.

We then extracted the fMRI time course from the FFA in all experiments (see Fig. 8.8a). The signal from the FFA was correlated with both face identification (identified > detected; $P < 10^{-2}$) and face detection (detected > not detected; $P < 10^{-3}$). While the overall amplitude of the FFA signal was lower in the bird experiments, it showed correlation with both bird identification (identified > detected; $P < 10^{-2}$) and detection (detected > not detected; $P < 10^{-2}$). In contrast, the signal in the FFA was not correlated with success at guitar identification. The difference in the signal amplitude between trials in which guitars were identified compared to trials in which guitars were not identified did not reach statistical significance. Thus, the fMRI signal from the FFA does not correlate with success at guitar identification. Critically, the lack of a differential signal between trial types in the guitar experiment does not seem to stem from a

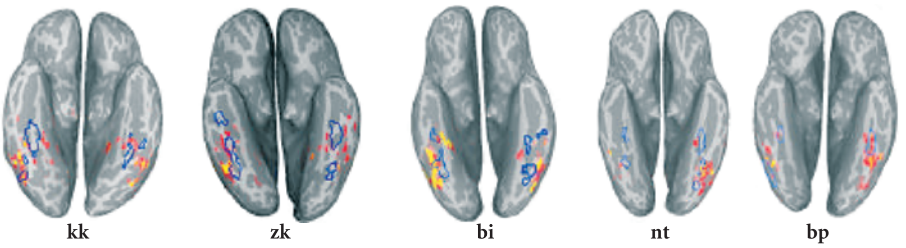
(a) Activation maps of areas correlated to face identification



(b) Activation maps of areas correlated to bird identification



(c) Activation maps of areas correlated to guitar identification



🟦 FFA - defined by localizer experiment:
faces > cars and novel objects $P < 10^{-4}$

**Fig. 8.7** Relation between the fusiform face area and areas correlated with identification of different object categories. The FFA was mapped independently for each subject in a separate localizer scan and was defined as areas in the fusiform gyrus that responded more strongly to faces compared to cars and novel objects, with a threshold of $P < 10^{-4}$. The blue contours indicate the boundaries of ventral face selective regions for each subject. We superimposed statistical maps of areas that were correlated with successful identification of: (a) faces, (b) birds, (c) guitars. Color of the statistical map indicates significance level (red, $P < 0.01$; yellow, $P < 0.0001$).

lack of a measurable signal, since the amplitude of the fMRI signal was not zero and the signal was not noisier than other experiments. The outcome of this analysis reveals that face selective regions were not correlated to success at identifying all subordinate-level object categories.

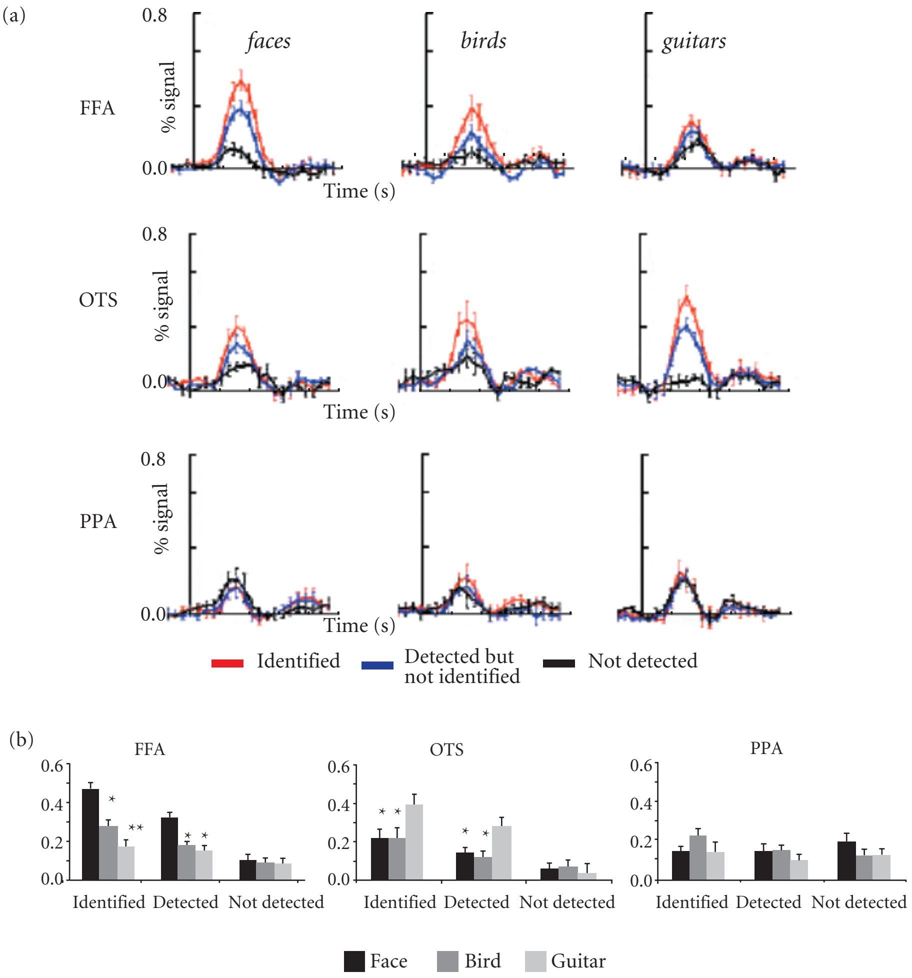


**Fig. 8.8** Activation in the fusiform face area (FFA), occipitotemporal sulcus (OTS), and parahippocampal place area (PPA) across experiments. (a) Raw time courses extracted from the FFA, OTS, and PPA averaged across five subjects for three experiments: *left,* faces; *middle*, birds; *right*, guitars. The FFA was defined by the localizer scan (see blue contours in Fig. 8.7). OTS voxels were defined as voxels in the OTS that were correlated with guitar recognition and did not overlap with the FFA. PPA was defined in the localizer scan as regions in the parahippocampal gyrus that showed higher activation for outdoor scenes containing sculptures versus faces and cars. (b) Mean activation amplitudes for five subjects averaged across three time points around the peak of activation (4-6 s after trial onset). Error bars indicate SEM. In both FFA and OTS (but not the PPA) there was a main effect of category and success. FFA, asterisks indicate significantly lower activation for guitars and birds than faces. OTS, asterisks indicate significantly less activation for faces and bird than guitars. $^{*}P < 10^{-2}$; $^{**}P < 10^{-4}$. In the PPA the differences did not reach statistical significance. We verified the main effects of success, category, and interaction between them via a two-way ANOVA analysis.

The activation maps in Fig. 8.7 indicate that for all subjects there was a region in the OTS lateral to the FFA that was correlated with guitar identification. For each subject we defined ROIs in the OTS that were correlated with guitar identification but did not overlap with the FFA. Time courses extracted from the OTS revealed an opposite profile of activation relative to FFA activation (see Fig. 8.8a). While OTS voxels reveal a main effect of success, the higher signal for identification hits versus detection hits in the OTS reached significance ($P < 10^{-3}$) for guitars but did not reach significance for faces or birds. Surprisingly, OTS voxels demonstrated a main effect of category preference for guitars (see Fig. 8.8b). This category preference was revealed despite the fact that the selection of these voxels was not based on category selectivity.

The category preference in the FFA (for faces) and OTS (for guitars) was not automatic. On trials in which objects were present, but not detected, the fMR signal from both the FFA and OTS was lowest. Importantly, in this condition there was no statistically significant difference between the activations to different categories, in either the FFA or OTS (see Fig. 8.8b). However, on trials in which objects were detected or identified, the signal was significantly higher in the FFA for faces compared to both birds and guitars (and the converse was true for OTS activation). This indicates that the higher activation for the preferred category does not occur automatically but only when a percept occurs.

Is the FFA a module dedicated to subordinate identification? The data presented here indicate that the FFA is not a module for subordinate categorization of all categories. First, the signal from the FFA is predictive of success at both detection and identification of faces (and birds). Secondly, the signal from the FFA was correlated with successful identification of faces but not guitars. Thus, it is not correlated with identification of all object categories. However, we found a different region in the OTS that was correlated with successful identification of guitars, showing that our failure to find this effect in the FFA was not due to some artifact. These results argue against the hypothesis suggested by Tarr and Gauthier (2000) that the role of the FFA is fine-grained subordinate discrimination between objects of any category.

## 8.11 **Discussion**

In sum, our data show that when the category was held constant but subjects performed different recognition tasks (fine-grained identification or object detection) similar regions in the human ventral stream were activated. However, when the task was kept constant and subjects were required to identify different object categories,

---

FFA, the interaction between faces and guitars was found for ‘detected’ compared with ‘not detected’ ($F > 6$; $P < 0.01$) and for ‘identified’ compared with ‘detected’ ($F > 4$; $P < 0.05$). OTS: ‘detected’ compared with ‘not detected’, $F > 6$, $P < 0.01$; ‘identified’ compared with ‘detected’, $F > 6.7$; $P < 0.001$.

different regions of the human ventral stream were activated. Even at the lowest thresholds there was twice as much overlap by task than content. This suggests that the human ventral stream is organized more around visual content than visual process (at least the processes treated here). Furthermore, we have shown that areas that were correlated with correct recognition of a category were also selective to that category, but only when the objects were detected or identified. Finally, our data show that the FFA is not a module for subordinate identification of any object category.

Our data reveal that similar regions within higher order areas were correlated with correct identification and detection when the object category was held constant. Surprisingly, even the most anterior regions along the ventral stream showed correlation with success at both detection and identification. This occurred despite the lack of necessity for explicit visual recognition in the detection task. In all areas that showed correlation with success at recognition, the signal was stronger for trials in which objects were successfully identified compared to trials in which objects were detected but not identified and lowest when objects were not detected at all. One possible explanation for the higher signal for identification compared to detection within the same regions is that identification requires longer processing times, and thus is a consequence of more neural processing. Since the bold signal sums up all the neural activations (Logothetis *et al.* 2001) this will be measured as a larger fMRI signal. Thus, one implication of these results is that the additional processing necessary for identification compared to detection occurs within the same cortical regions.

We found some differences between lateral-occipital and ventral occipitotemporal regions that were correlated with successful visual recognition. While the overlap by task analysis revealed that most of the voxels in LO and VOT were correlated with both identification and detection of an object category, some of the voxels in LO and VOT were correlated with detection but not with identification. Overall, LO regions contained a higher percentage of voxels that were correlated primarily with detection and not identification compared to VOT. Thus, if we consider a hierarchy of visual areas involved in object recognition, LO seems to be a candidate for a processing stage prior to VOT.

While there was substantially more segregation by category than by task, we did not find 100% overlap or segregation by either task or category. However, what clearly emerges is that the level of visual processing (at least those tested here) is not the major guideline for differentiating between regions in the ventral stream; object type seems to play a more critical role in differentiating among subdivisions of the human ventral stream. What still remains unknown is the precise nature of object representation in these regions, which could be whole objects, object fragments or even complex features and feature conjunctions.

Our data show that the recognition of a category is correlated with a distributed, distinct, and replicable pattern of activation across higher-order brain areas. This result is consistent with an ‘object form topography’ (Edelman *et al.* 1998; Haxby *et al.* 2001).

However, there are several important differences between the current results and those of Haxby *et al.* (2001). In their study Haxby *et al.* (2001) examined where there is information about object categories in ventral cortex. They showed that the response to a given category could be determined by the distributed pattern of activation across all ventral occipitotemporal cortex. Importantly, their analysis revealed that it was possible to predict the category of the object even when regions that showed maximal activation to a particular category were excluded. In contrast, the experiments described here examined which areas are used for visual recognition tasks. Our data suggest that the activation that is correlated with successful recognition of a category does not extend across all VOT. Rather, subregions within the VOT tend to be correlated with successful recognition of a category, and these were the same regions that showed higher activation for that category.

One of the predictions from Haxby *et al.* (2001) is that it is possible to predict from the pattern of activation across regions that showed a higher signal for house, the object category, (even faces, which are not the preferred category). To test directly whether our subjects used house-selective regions in the parahippocampal place area (PPA) to recognize objects, we extracted time courses from PPA ROIs.[7] Unlike the activation in the OTS and FFA, the time courses extracted from the PPA did not show either a main effect of success (see Fig. 8.8a) or category (see Fig. 8.8b). Thus, the signal from the PPA was not correlated with successful recognition of these three object categories. Thus, while Haxby *et al.* (2001) suggest that activation in the PPA conveys information about all object categories, our data indicate that the PPA does not seem to be utilized for the recognition of several categories. Moreover, the fact that there exists a region within the ventral stream that is not correlated with correct recognition further strengthens the conclusion that not all regions within the VOT contribute equally to visual recognition of specific categories.

Another mystery that remains unresolved is why the higher activation for specific categories is localized and replicable across subjects? Malach *et al.* (2002) have suggested that category preference emerges from resolution needs, which are tightly linked to eccentricity bias (Levy *et al.* 2001; Hasson *et al.* 2002; Malach, *et al.* 2002). However this explanation does not fully account for the experiments described here. Here, all identification experiments required high visual acuity, yet different regions of the ventral stream were correlated with the identification of different categories. One possibility (that remains to be examined) is that regions along the OTS have a foveal bias. Another possibility is that eccentricity bias is not the only parameter that governs object-form topography in the human ventral stream.

However, this explanation does not fully account for the experiments described here. Here, all identification experiments required high visual acuity—yet different regions of the ventral stream were correlated with the identification of different categories. One possibility (that remains to be examined) is that regions along the OTS have a foveal bias. Another possibility is that eccentricity bias is not the only parameter that governs object-form topography in the human ventral stream.

Finally, we address the issue of modularity. While our data argue against the hypothesis that the FFA is a module for subordinate recognition, they also pose some constraints on the hypothesis that the FFA is a module dedicated solely for face recognition. While the signal from the FFA was highest for faces, it was also correlated with success at bird identification and detection. This correlation between the FFA signal and success at bird detection and identification occurred despite the fact that our subjects were not bird experts. One possibility is that neurons within the FFA are selective to both human and animal faces or face features. Birds (and other animals) have faces and face parts such as eyes, mouths, etc. Thus one possible interpretation of these results is that the axis of differentiation between subregions within ventral cortex is animate/inanimate rather than faces/objects, and the activation of FFA may be necessary for recognition of animate categories that contain faces. However, this remains to be verified.

Another difficulty in interpreting the function of the FFA as a module for face recognition is that the signal in an adjacent region in the OTS showed a success effect for faces, even though the signal was smaller for faces compared to guitars. One possible explanation is that this was caused by partial voluming artifacts. Another alternative is that the OTS contains some general shape processors that are used for both object and face recognition. Hopefully, in the near future we should be able to distinguish between these two alternatives by imaging the brain at higher resolutions.

To conclude, our data shows that the functional organization of object and face selective regions in the ventral visual pathway is organized on stimulus content rather than object recognition task (detection vs. identification). The experiments described here provide important insights to the functional organization of higher-order visual areas in and their role in visual recognition.

## Notes

1. For each subject we determined the exposure of images for the fMRI experiments by running a behavioral experiment prior to the scan, in which we showed masked images for various durations. The exposure duration was defined as the minimal exposure duration in which each subject could detect at least 50% of the images. This duration was fixed for a given subject and varied between 33 and 50 ms across subjects who participated in the fMRI scans.
2. Scanning parameters: MGH, Siemens allegra, head-only scanner; 10 oblique slices, $3.125 \times 3.125 \times 4$ mm; covering the occipital, posterior parietal and temporal lobes; FOV = 20 cm; TR = 1 s; TE = 43 ms; flip angle = 60°.
3. Correct rejects are not plotted here since the behavioral data on the non-target object images is ambiguous. One possibility is that these objects were rejected because subjects correctly identified them as a different subordinate category. However, it is possible that subjects could not identify them at all, and therefore these pictures were rejected.
4. Here we refer to these regions as VOT, in previous publications we referred to them as LOa.
5. VOT foci were defined as foci within the OTS and fusiform gyrus that were correlated with successful recognition. LO foci were located around the lateral occipital sulcus (LOS) and inferior occipital gyrus (IOG), lateral to a lower meridian representation. Both VOT and LO foci lie beyond retinotopic cortex (see Grill-Spector *et al.* 2000, for details).

6. The thresholds used here were chosen to be the lowest and highest thresholds possible in the conditions of these experiments. The low threshold used was $P < 0.01$, similar to that of Experiment 1. Decreasing the threshold to lower values resulted in detecting voxels outside the brain that were typically noise. The higher threshold was a value of $P < 0.001$.
7. For most subjects further increasing the value of threshold resulted in detecting few active voxels. Here we used a somewhat a non-standard test to define the PPA, and we used it as a *post-hoc* analysis method to define the relevant ROI. We searched for regions that showed a higher signal for outdoor scenes containing abstract sculptures compared to faces, cars, and textures in the localizer scan. This contrast activates regions in LO and in a lower thresholds region in the parahippocampal gyrus. For this analysis we took only the regions within the parahippocampal gyues that corresponded anatomically with the PPA (Epstein and Kanwisher 1998) that were statistically significant in this test.

## Acknowledgements

Many thanks to Nancy Kanwisher for invigorating discussions and for her continuous support of this research. This research has been funded by HSFP LT0670 to K.G.S.

## References

**Aguirre, G. K., Zarahn, E., and D'Esposito, M.** (1998). An area within human ventral cortex sensitive to 'building' stimuli: evidence and implications. *Neuron,* **21**(2), 373–83.

**Allison, T., Puce, A., Spencer, D. D., and McCarthy, G.** (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb Cortex,* **9**(5), 415–30.

**Avidan, G., Hasson, U., Hendler, T., Zohary, E., and Malach, R.** (2002). Analysis of the neuronal selectivity underlying low fMRI signals. *Curr Biol,* **12**(12), 964–72.

**Chao, L. L., Haxby, J. V., and Martin, A.** (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat Neurosci,* **2**(10), 913–19.

**Dale, A. M., Fischl, B., and Sereno, M. I.** (1999). Cortical surface-based analysis I: Segmentation and surface reconstruction. *NeuroImage,* **9**, 179–94.

**Damasio, A. R.** (1990). Category-related recognition defects as a clue to the neural substrates of knowledge. *Trends Neurosci,* **13**(3), 95–8.

**Damasio, A. R., Tranel, D., and Damasio, H.** (1990). Face agnosia and the neural substrates of memory. *Annu Rev Neurosci,* **13**, 89–109.

**Downing, P. E., Jiang, Y., Shuman, M., and Kanwisher, N.** (2001). A cortical area selective for visual processing of the human body. *Science,* **293**(5539), 2470–3.

**Driver, J. and Baylis, G. C.** (1996). Edge-assignment and figure-ground segmentation in short-term visual matching. *Cognit Psychol,* **31**(3), 248–306.

**Epstein, R. and Kanwisher, N.** (1998). A cortical representation of the local visual environment. *Nature,* **392**(6676), 598–601.

**Farah, M. J.** (1992). Agnosia. *Curr Opin Neurobiol,* **2**(2), 162–4.

**Fischl, B., Sereno, M. I., and Dale, A. M.** (1999). Cortical surface-based analysis II: Inflation, Flattening, a surface-based coordinate system. *NeuroImage,* **9**, 195–207.

**Fischl, B., Liu, A., and Dale, A. M.** (2001). Automated manifold surgery: Constructing geometrically accurate and topologically correct models of the human cerebral cortex. *EEE Transactions on Medical Imaging,* **20**(1), 70–80.

**Fujita, I., Tanaka, K., Ito, M., and Cheng, K.** (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature,* **360**(6402), 343–6.

**Gallant, J. L., Braun, J., and Van Essen, D. C.** (1993). Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science,* **259**(5091), 100–3.

**Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., and Gore, J. C.** (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat Neurosci,* **2**(6), 568–73.

**Grill-Spector, K. and Kanwisher, N.** (2003, submitted). Visual recognition: as soon as you see it you know what it is.

**Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., and Malach, R.** (1998). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Hum Brain Mapp,* **6**(4), 316–28.

**Grill-Spector, K., Kushnir, T., Hendler, T., and Malach, R.** (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nat Neurosci,* **3**(8), 837–43.

**Hasson, U., Hendler, T., Ben Bashat, D., and Malach, R.** (2001). Vase or face? A neural correlate of shape-selective grouping processes in the human brain. *J Cogn Neurosci,* **13**(6), 744–53.

**Hasson, U., Levy, I., Behrmann, M., Hendler, T., and Malach, R.** (2002). Eccentricity bias as an organizing principle for human high-order object areas. *Neuron,* **34**(3), 479–90.

**Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., and Pietrini, P.** (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science,* **293**(5539), 2425–30.

**Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., and Haxby, J. V.** (1999). Distributed representation of objects in the human ventral visual pathway. *Proc Natl Acad Sci USA,* **96**(16), 9379–84.

**James, T. W., Humphrey, G. K., Gati, J. S., Menon, R. S., and Goodale, M. A.** (2000). The effects of visual object priming on brain activation before and after recognition. *Curr Biol,* **10**(17), 1017–24.

**Kanwisher, N.** (2000). Domain specificity in face perception. *Nat Neurosci,* **3**(8), 759–63.

**Kanwisher, N., McDermott, J., and Chun, M. M.** (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci,* **17**(11), 4302–11.

**Kleinschmidt, A., Buchel, C., Hutton, C., Friston, K. J., and Frackowiak, R. S.** (2002). The neural structures expressing perceptual hysteresis in visual letter recognition. *Neuron,* **34**(4), 659–66.

**Lamme, V. A.** (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *J Neurosci,* **15**(2), 1605–15.

**Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., and Malach, R.** (2001). A hierarchical axis of object processing stages in the human visual cortex. *Cereb Cortex,* **11**(4), 287–97.

**Levy, I., Hasson, U., Avidan, G., Hendler, T., and Malach, R.** (2001). Center-periphery organization of human object areas. *Nat Neurosci,* **4**(5), 533–9.

**Livingstone, M. and Hubel, D.** (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science,* **240**(4853), 740–9.

**Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A.** (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature,* **412**(6843), 150–7.

**Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., Ledden, P. J., Brady, T. J., Rosen, B. R., and Tootell, R. B.** (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci USA,* **92**(18), 8135–9.

**Malach, R., Levy, I., and Hasson, U.** (2002). The topography of high-order human object areas. *Trends Cogn Sci,* **6**(4), 176–84.

**Martin, A., Wiggs, C. L., Ungerleider, L. G., and Haxby, J. V.** (1996). Neural correlates of category-specific knowledge. *Nature,* **379**(6566), 649–52.

**Nakayama, K., He, Z. J., and Shimojo S,** (1995). Visual surface representation: a critical link between lower-level and higher-level vision. In Kosslyn, S. M. and Osherson, D. N. (Eds.), *An invitation to cognitive science: visual cognition.* MIT press, Cambridge, MA.

**Peterson, M. A.** (1994). Shape recognition can and does occur before figure-ground organization. *Curr Direct Psychol Sci,* **3**, 105–11.

**Puce, A., Allison, T., Gore, J. C., and McCarthy, G.** (1995). Face-sensitive regions in human extrastriate cortex studied by functional MRI. *J Neurophysiol,* **74**(3), 1192–9.

**Puce, A., Allison, T., Asgari, M., Gore, J. C., and McCarthy, G.** (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *J Neurosci,* **16**(16), 5205–15.

**Rosch, E. M., Mervis, C. B., Gray, M. D., Johnson, D. M., and Boyes-Braem, P.** (1976). Basic objects in natural categories. *Cognit Psychol,* **8**, 382–439.

**Rubin, E.** (1958). Figure and ground. In W. M. Beardslee (Ed.), *Readings in perception.* New York: Van Nostrand.

**Tanaka, J.** (2001). The entry point of face recognition: evidence for face expertise. *J Exp Psychol Gen,* **130**(3), 534–543.

**Tarr, M. J. and Gauthier, I.** (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nat Neurosci,* **3**(8), 764–9.

**Tong, F., Nakayama, K., Vaughan, J. T., and Kanwisher, N.** (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron,* **21**(4), 753–9.

**Tsunoda, K., Yamane, Y., Nishizaki, M., and Tanifuji, M.** (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nat Neurosci,* **4**(8), 832–8.

**von der Heydt, R., Peterhans, E., and Baumgartner, G.** (1984). Illusory contours and cortical neuron responses. *Science,* **224**(4654), 1260–2.

**Zhou, H., Friedman, H. S., and von der Heydt, R.** (2000). Coding of border ownership in monkey visual cortex. *J Neurosci,* **20**(17), 6594–611.