Influences of Extra-Retinal Information on Visual Object Recognition: fMRI Study

Thesis Submitted for the degree
Doctor of Philosophy

by
Galvia Avidan

Submitted to the Senate of the Hebrew University of Jerusalem
July, 2002
This work was carried out under the supervision of

Dr. Rafael Malach

Dr. Ehud Zohary
Acknowledgements

During the course of my PhD studies I was very fortunate to work with a large group of clever and supportive people, who not only assisted me on scientific aspects, but were also my true friends. Thanks to all these wonderful people, I had the unique opportunity to work in a warm and caring working environment, which undoubtedly contributed much to this project.

I owe special thanks to my supervisors Dr. Ehud Zohary and Dr. Rafael Malach for introducing me into the world of vision research, and for educating me as a scientist by encouraging both independent and critical thinking. I would also like to express my deep appreciation to their wise, enthusiastic, and devoted guidance, and to thank them for the numerous hours they spent discussing scientific issues with me. Importantly they both also taught me that one could be immersed in science while still maintaining a life.

Having two supervisors, I encountered many students and colleagues from the vision labs in both the Hebrew University and the Weizmann Institute with whom I greatly enjoyed working. I would like to thank Prof. Shaul Hochstein for his original and wise advice and Michal Harel for her scientific assistance as well as for her concern and kindness. I would especially like to thank my dear friends Uri Hasson and Ifat Levy who were always there for me in any sense.

I would like to thank all my colleagues from the Imaging department in the Sourasky Medical Center who made it possible for me to work in a state of the art MRI facility and especially Dr. Talma Hendler who also provided valuable scientific advice.

Finally I would like to express my gratitude to my parents, Nili and Benny Avidan, whose unlimited love and support encouraged me not only to start this long and demanding journey but also made it possible for me to reach its end.
Abstract

Two main sources of information contribute to visual object perception. The first is extracted in a feedforward, ‘bottom-up’ fashion from the physical properties of the visual stimulus. The second type of information is more related to the observer’s state. This ‘top-down’ or extra-retinal information can dramatically modify object recognition and perception. In the first part of my thesis, I make a distinction between high order visual areas, and early retinotopic areas. The characteristic response of high order visual regions is more related to the actual perception of the visual stimulus, while that of the retinotopic regions is more tightly linked to its physical properties. I then focus my study on characterizing the principles of object representation within these high-order visual areas, and particularly I address the issue of whether these areas contain cortical modules specialized for specific object categories. Finally, by manipulating subjects’ attention I explore possible sources for the representation of different categories within these regions and also characterize the modulatory effect of attention on this representation. Detailed description of each study is given below.

An important characteristic of visual perception is the fact that object recognition is largely immune to changes in viewing conditions. I used a single, well defined visual property, that of image contrast, to follow the transformation in image representation along the entire constellation of ventral stream human visual areas. Contrast is a suitable parameter to study, since perceptually, object recognition is highly invariant to contrast changes beyond a minimal contrast level, whereas retinal responses are highly sensitive to all contrast levels. Our results reveal that the correlation between the contrast of the physical stimulus and fMRI response shows a gradual and consistent decline as one moves to high order visual areas along the ventral stream. Thus, object-related foci in the occipito-temporal cortex including the lateral occipital region (LO) and the more ventral and temporal region (posterior fusiform gyrus, pFs) tend to be contrast-invariant, specifically for face stimuli. These results reflect a hierarchical trend in the human
visual cortex, in which cortical responses gradually depart from the physical aspects of the visual stimulus and become more correlated with perceptual experience.

While in the first study I explored the build-up of object representation along the ventral visual cortex, in the second study, I focused on the properties of category-related object representation within the high order, occipito-temporal regions located at the top of the ventral stream hierarchy. It is not clear to what extent these regions are purely specialized for one stimulus category and to what extent other object representations coexist within them. Our approach employs fMRI adaptation, i.e. the reduction in fMRI signal following repeated presentation of visual stimuli, as an indicator for revealing highly active neuronal populations even when the measured fMRI signal is low thus circumventing for the limited spatial resolution of the fMRI method. My results show that while the overall fMRI signal in category-related regions (e.g. face-related pFs) was drastically reduced for the non-preferred category (e.g. non-face stimuli), the adaptation level obtained for these stimuli remained high. This effect is indicative of residual populations responding vigorously to the “non-optimal” stimuli that remain active despite an overall reduction in fMRI activity. These findings suggest that functionally distinct “hot spots” of neuronal activation reside within category-selective object areas that are strongly activated by the “non-optimal” stimuli. Thus, although a clear functional specialization does exist within ventral stream areas, it is not absolute, and representations of various object categories coexist within these regions.

In the third study I continued to explore the nature of category-related object representation, focusing on the role of attentional modulation in constructing these representations. Recent findings from our lab showed that the topographic organization of visual field eccentricity in human visual cortex extends into high order, ventral occipito-temporal (VOT) cortex. Within this region faces and buildings have specific eccentricity biases, suggesting that this category-eccentricity association may reflect differential resolution needs of recognition processes.
However, it was unclear to what extent this center/periphery differentiation depends on immediate, moment-to-moment, task demands and to what extent it is established through long-term effects.

In this study I examined the role of such moment-to-moment attentional shifts in constructing the category-eccentricity association within high-order occipito-temporal cortex. I also characterized the nature of the attentional mechanisms found in these regions, by comparing the modulatory impact of spatial attention and attention to object identity. I found that when activation by images of buildings and faces was contrasted, the differential activation to each object category in occipito-temporal object areas was maintained, even when attention was shifted to different spatial locations. However, under these conditions object-related activation was enhanced in a shape-specific manner. When spatial attention was modulated, by attending to large or small arrow-heads superimposed on the object images, weak correlation with the “spotlight size” was observed in higher order ventral occipito-temporal (VOT) regions. These results argue against the possibility that the source for the category differentiation of ventral occipito-temporal cortex is purely due to spatial attention. Rather, the results support the notion that the attentional modulation in VOT involves a combined effect of both shape selectivity and eccentricity bias.
# Table of Contents

## Chapter 1
**General Introduction**

1.1 *Organization principles of the visual system* ................................................................. 1  
1.2 *Object recognition* ........................................................................................................ 4  
  1.2.1 Object-related areas and their functional properties in the macaque brain............... 4  
  1.2.2 Object-related areas in the human brain ................................................................. 6  
    1.2.2.1 Organization of human object-related areas ......................................................... 6  
    1.2.2.2 Functional properties of human object-related areas ......................................... 9  
1.2.3 Top-down modulation of visual object recognition.................................................... 11  
  1.2.3.1 Correlation between neuronal activity and object recognition ......................... 11  
  1.2.3.2 Attentional modulation of object-related activation ........................................... 14  
1.3 *Methodological accounts* ............................................................................................ 16  

## Chapter 2
**Results**

2.1 *Contrast sensitivity in human visual areas and its relationship to object recognition* ................................................................. 17  
2.2 *Analysis of the neuronal selectivity underlying low fMRI signals* ......................... 33  
2.3 *Spatial vs. object specific attention in high-order visual areas* ................................ 43  

## Chapter 3
**General Discussion**

3.1 *Hierarchical organization of the visual cortex* ......................................................... 70  
3.2 *Models of cortical organization for object representation* .................................... 73  
3.3 *Attentional modulation of object recognition* ......................................................... 79  
3.4 *Correlations between fMRI signal and object perception* ...................................... 80  

**References** .................................................................................................................. 84  

**Appendix**  
Face-selective activation in a congenital prosopagnosic subject.
Chapter 1

General Introduction

1.1 Organization principles of the visual system

Much of our knowledge and understanding of the visual cortex was gained by anatomical, physiological, imaging and lesion studies performed in monkeys. The human visual cortex is studied by behavioral, neuropsychological and electrophysiological approaches and in recent years also by imaging techniques such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI). The following text will include relevant findings from both species that were obtained by the various techniques. It is important to note however that the extent of homology between these species is still largely undetermined and is being constantly updated (Van Essen and Drury, 1997).

Visual information in both the macaque and human brain is processed through a diverging tree of tens of cortical regions. The current dominating view is that each cortical area is specialized in processing specific aspects of the visual scene such as motion, color, form, and spatial information (Felleman and Van Essen, 1991; Van Essen and Drury, 1997). Anatomical studies showed that these areas are heavily interconnected in a reciprocal fashion, and are hierarchically organized to a large extent, composing distinct but intertwined processing streams (Felleman and Van Essen, 1991).

The concept of hierarchical organization of the visual system is also supported by physiological findings. Thus, while receptive field size in area V1 are typically very small (less than 1 deg), they increase progressively at successive stages, (Amir et al., 1993; Tootell et al., 1997; Van Essen, 1985) ultimately capturing a large portion of the entire visual field (tens of degrees) in some inferotemporal regions (Grill-Spector et al., 1998b; Gross et al., 1972; Tootell et al., 1998c). In addition, receptive field properties also change substantially at progressively later stages of the hierarchy, so that they have increased selectivity for more complex stimuli. Accordingly, while V1 cells respond to simple oriented bars and edges placed within their receptive fields (Hubel, 1988; Tootell et al., 1998b), cells in inferotemporal cortex are selective...
for complex patterns such as face stimuli (Baylis et al., 1987; Desimone et al., 1984; Tanaka, 1993). Such selective activations for complex stimuli were also found in human occipito-temporal cortex (Kanwisher et al., 1996; Kreiman et al., 2000; Malach et al., 1995). Another measure of the hierarchical level of a visual area can be derived from the latency of its responses evoked by the presentation of a visual stimulus (Lamme and Roelfsema, 2000), thus response latencies of inferotemporal (IT) neurons are generally greater than those of earlier areas (V1 to V4). These examples, as well as many others, which will be discussed below, support the notion that higher stages of the cortical hierarchy represent more advanced levels of processing.

The neural substrates mediating object and spatial processing of both macaque and human visual systems appear to be largely segregated into two anatomically distinct and functionally specialized streams: the ventral and dorsal pathways respectively (Ungerleider and Mishkin, 1982). The first pathway, the ventral pathway, interconnects the striate, prestriate, and inferior temporal areas (areas V1, V2, V4, TEO and TE). It is crucial for visual features relevant to object identification (i.e. ‘what’), such as color and shape. The second pathway, the dorsal pathway, interconnects the striate, prestriate and inferior parietal areas (V1, V2, V3, MT, MST and further stations in inferior parietal and superior temporal cortex). This pathway plays an essential role in perceiving the spatial aspects of visual stimuli (i.e. ‘where’), such as direction of motion and velocity as well as tracking eye movements.

Visual impairments produced by focal lesions in clinical cases (Farah, 1995) as well as several neuroimaging studies (Haxby et al., 1991; Haxby et al., 1994; Kohler et al., 1995; Ungerleider and Haxby, 1994) were able to demonstrate this segregation in the human visual cortex as well. Based on neurophysiological studies a somewhat different organization of the visual cortex to a “what” and “how” pathways, was also suggested (Goodale and Milner, 1992).

Another key aspect of the organization of the visual system is the spatially ordered- i.e. topographic, projection of information from the retina to the visual cortex. This involves a transformation from a
Euclidean coordinate system representation in the retina to a polar coordinate framework in the visual cortex. This principle was found initially in non-human primates (Schwartz et al., 1985; Tootell et al., 1988) and more recently in human early visual areas (DeYoe et al., 1996; Engel et al., 1997; Horton and Hoyt, 1991; Sereno et al., 1995). According to this transformation, which is applied in each of the early visual areas, the visual field is mapped along two orthogonal axes: polar angle and eccentricity. At progressively later stages of processing the retinotopic organization becomes increasingly coarse (but see Levy et al., 2001; Sereno et al., 2001).

Many of the physiological properties of the visual system described above can be attributed to the vast network of feedforward (ascending) connections and to the intrinsic circuitry within each visual area. However, the monkey visual cortex contains extensive reciprocating, feedback (descending) connections (Felleman and Van Essen, 1991; Rockland and Pandya, 1979). This observation is significant because it rules out the possibility that the visual cortex might operate by a strictly serial, feedforward processing scheme and implies that this pattern of connectivity could serve as the neural substrate for modulatory top-down influence. Moreover, feedback inputs from cortical areas outside the visual cortex, such as those in prefrontal cortex could also mediate such top-down influences. Thus, descending connections may contribute to the modulation of the visual response properties by processes such as attention, learning and memory, expectations, or contextual effects. Importantly, although feedback connections do exist, they can still be consistent with the notion of hierarchical processing in the broad sense. While the physiological properties of any given cortical neuron will reflect both ascending and descending influences, a given cell still represents a well-defined hierarchical position in terms of the information it represents explicitly and the way in which this information is used (Felleman and Van Essen, 1991). A somewhat different account for hierarchical processing and the role of top-down pathways in the visual cortex was suggested by Ullman (Ullman, 1996). Unlike the view presented above in which top-down processing is more cognitive than visual in nature and plays a role when visual input becomes ambiguous or after recognition already took place, according to Ullman’s model the use of top-down processing is far
more extensive and comprise an integral part of the recognition process. Thus object recognition is accomplished by a sequence-seeking model in which the computation is a bi-directional process performed by the combination of top-down and bottom-up, counter streams of processing. Another key aspect of the model is that in each direction, different alternatives for matching the incoming visual input with the stored representation are explored in parallel.

1.2 Object recognition

Visual object recognition is a key function of the primate brain. However, trying to understand this process is not a trivial task. Several reasons might account for this inherent difficulty. First, there is no clear answer to the seemingly simple question: ‘what is an object?’ The definition of an object depends heavily on the scaling level at which one examines a certain visual scene. For example, usually a face would be considered as an object, but at a different scale, the eyes or nose composing this face are also objects. Moreover, object recognition is underconstrained since the same 2D projection on the retina could result from numerous 3D objects in the world (i.e. the inverse problem). Finally, the physical appearance of objects changes constantly due to changes in viewing conditions such as different illumination, viewing angle, viewing distance, deformation in object shape, or occlusion.

Thus, in order to account for all these factors, the visual system must be highly adaptable and maintain the capability to continuously learn new objects and acquire expertise regarding familiar object categories. Inferotemporal (IT) cortex, which is a part of the ventral – ‘what’ pathway, participates in visual perception and recognition in both macaque and human brain. In the following sections I will view some of the properties of this cortical region, which enable object recognition.

1.2.1 Object-related areas and their functional properties in the macaque brain

The macaque inferotemporal (IT) cortex is generally considered a large region of cortex extending approximately from just anterior to the inferior occipital sulcus to a couple of millimeters posterior to the
temporal pole, and from the fundus of the superior temporal sulcus (STS) to the fundus of the occipito-temporal sulcus (for reviews, see Logothetis and Sheinberg, 1996; Miyashita, 1993; Tanaka, 1996). IT is considered to be the last exclusively visual area in the ventral pathway, with more than 85% of its neurons excited or inhibited by different visual patterns (Desimone et al., 1984). The observed properties of IT cells change significantly as one moves from its most posterior part termed TEO to the anterior part termed TE. Among the changing characteristics are topography, receptive field size, and stimulus selectivity. Area TEO has a coarse retinotopic organization; it has an almost complete representation of the contralateral visual field. In contrast, area TE is not retinotopically organized; its cells have large, ipsilateral, contralateral or bilateral receptive fields that almost always include the fovea (Gross et al., 1972). There is a systematic increase in the receptive field size along the posterior-anterior axis of IT, with receptive field diameters in TEO as small as 1.5 to 2.5 deg and those in TE reaching diameters of 30 to 50 deg. The critical features for the activation of cells in IT are moderately complex with the degree of complexity increasing along the posterior-anterior axis from area TEO to area TE (Tanaka, 1996; Tanaka, 1997).

Recent studies (Fujita et al., 1992; Tanaka, 1996) have revealed a systematic columnar organization in the TE region. Cells with relatively similar response properties are clustered in a columnar fashion perpendicular to the cortical surface, whereas cells in different columns respond to different features. The average width of a column is 400 µm. Data from optical imaging in TE have suggested that the borders between neighboring columns are not discrete (Tsunoda et al., 2001). This columnar organization may be used in various computations, such as representation of objects under varying conditions.

Many IT neurons indeed show various degrees of invariance to image transformation. The absolute response of the cells rarely exhibits size or position constancy, however, their selectivity to shape, i.e. their relative preference for the optimal stimulus over several sub-optimal stimuli, is preserved over large changes in stimulus size and position. In this sense more than half of the IT neurons can be thought of as
demonstrating size and position invariance (Gross et al., 1972; Ito et al., 1995; Lueschow et al., 1994; Tovee et al., 1994).

1.2.2 Object-related areas in the human brain

Several PET studies of the human visual cortex, conducted in the early 1990’s, have revealed selective activation in ventral temporal cortex which was related to face and object recognition (Allison et al., 1994a; Allison et al., 1994b; Haxby et al., 1991; Kosslyn et al., 1994; Sergent et al., 1992). Event-related potentials recorded from electrodes placed on the cortical surface of patients prior to surgery (Allison et al., 1999; McCarthy et al., 1999) also reveal object specific waveforms having stronger activation for a variety of objects as well as faces compared with scrambled control stimuli. Converging evidence from neuropsychological studies support the notion that occipito-temporal regions are not merely involved in object recognition but rather are necessary for this process to occur. Thus, damage to these regions results in a variety of recognition deficits (Damasio, 1990; Farah, 1995; Moscovitch et al., 1997).

One of the early attempts to map object related areas in the human visual cortex using fMRI was made by Malach et al. (Malach et al., 1995). They revealed a large cortical expanse, located in the lateral and ventral aspects of the occipito-temporal cortex, which exhibited preferential activation to a wide range of object images including 3D abstract shapes, compared with a variety of texture and noise patterns. This region was termed the lateral-occipital complex (LOC). Kanwisher et al. (Kanwisher et al., 1996) also reported similar results using intact and scrambled line drawings of object stimuli.

1.2.2.1 Organization of human object-related areas

Occipito-temporal, object-related regions are largely non-retinotopic (Grill-Spector et al., 1998b; Tootell et al., 1998c) and were suggested to be located at the end of a posterior-anterior hierarchical axis of object representation stretching along ventral stream visual areas (see Figure 1). Areas along this axis show a gradual trend of increasing contrast invariance stressing the transition from early retinotopic areas, which
largely depend on the physical properties of the stimulus to high order ones, in the build-up of abstract object representations (Avidan et al., 2002a). These latter results will be presented in details in Chapter 2.1. In addition areas along this axis become increasingly more sensitive to gradual scrambling of object images (i.e. gradual scrambling paradigm) thus implying a more holistic rather than part-based representation within high order object-related regions (Grill-Spector et al., 1998b; Lerner et al., 2001).

A hierarchical organization within the occipito-temporal object related regions themselves was also suggested. Thus, based on differential adaptation properties and on anatomical criteria, this large cortical expanse was subdivided into two tentative entities. A more dorsal region termed lateral occipital (LO) which is situated ventrally and posteriorly to area MT and extends into the posterior inferotemporal sulcus, and another focus situated in the vicinity of the posterior fusiform gyrus (pFs), which is anterior and lateral to area V4/V8 and extends into the inferotemporal sulcus (Figure 1). While the lateral occipital focus (LO) exhibited substantial sensitivity to stimulus’ size and position, pFs was largely invariant to these properties (Grill-Spector et al., 1999). In addition LO exhibits stronger motion sensitivity than pFs (Avidan et al., 2002c; Beauchamp et al., 2002; Kourtzi and Kanwisher, 1999) and see Chapter 2.3. A similar axis was found using the gradual scrambling paradigm (see above). These findings are also consistent with the hierarchical organization in the subdivisions of macaque IT cortex.

**Figure 1: Occipito-temporal object-related areas**

a. Occipito-temporal object-related areas of the human visual pathway are shown on two hemispheres of an inflated brain shown from a ventral view, and a left hemisphere shown in a lateral view. These areas include the lateral occipital region (LO), the posterior fusiform gyrus (pFs), and the collateral sulcus (CoS). LO and pFs were functionally defined by their preferential activation to face images compared with object and building images and separated based on anatomical criteria (see text). The CoS was defined by its preferential activation to building images compared with face and object images. Data are averaged across 14 subjects. b. Object-related areas are shown on a left, unfolded hemisphere of one representative subject. Retinotopic areas and area MT are marked by the white dotted lines and colored in light gray. (Adopted from Malach et al. 2002)
A number of functional mapping studies have revealed a complex network of specialized regions within occipito-temporal object-related cortex. The common theme to these studies is a clear selectivity in the response properties to various object shapes. A particularly remarkable example is the medio-lateral differentiation along ventral occipito-temporal cortex in response to face and building images. Thus faces activate a more lateral region located in the posterior fusiform gyrus (pFs) (Clark et al., 1996; Halgren et al., 1999; Kanwisher et al., 1997; Puce et al., 1995) while buildings, activate a more medial region located in the collateral sulcus (CoS) as well as the parahippocampal gyrus – PPA (Aguirre et al., 1998; Epstein and Kanwisher, 1998). Additional object categories such as tools (Martin et al., 1996), animals (Martin et al., 1996), human body (Downing et al., 2001b), words and letter strings (Hasson et al., 2002) and even chairs (Ishai et al., 1999) have been reported to manifest specific and differential activation patterns within these regions.

These consistent category-related activations are difficult to interpret since a variety of factors can account for them. Such factors include shape-related differences (Fujita et al., 1992), category-specific modules (Fodor, 1983; Kanwisher, 2000) task-related processing, and expertise effects (Gauthier et al., 1999). Moreover, the response to specific categories is by no means absolute: each specialized region shows substantial activation to other object categories (although to a lesser extent) as I show in Chapter 2.2. In the same chapter, I demonstrate that the source of this reduced activation are functionally distinct "hot spots" of neuronal activation, which reside within category selective cortex. Such results might imply a more distributed type of object representations (Haxby et al., 2001; Ishai et al., 1999).

Recent findings from our lab (Hasson et al., 2002; Levy et al., 2001) have led Malach et al. (Malach et al., 2002) to propose a new organizing principle of human high-order object areas. This scheme was based on the finding of topography in high order object areas, in which eccentricity bias, magnification factor and specific object shapes are linked in an orderly manner. Two such eccentricity maps were proposed, a posterior dorsal one, LO, located in lateral occipital cortex, and a ventral anterior one, VOT, in ventral occipito temporal cortex which includes the pFs and CoS regions (see Figure 1b). Malach et al.
hypothesized that the category-eccentricity association may reflect differential needs of recognition processes for high cortical magnification (associated with central vision) or large-scale integration (associated with peripheral vision). But this remains to be tested rigorously.

1.2.2.2 Functional properties of human object-related areas

Numerous fMRI studies conducted in the last several years have revealed that occipito-temporal object-related regions, possess various functional properties associated with high-level object representations. Among these properties are size and position invariance (Grill-Spector et al., 1999) as well as substantial contrast invariance (Avidan et al., 2002a), which were already mentioned. Additional properties are convergence of visual cues, completion and grouping effects as well as correlation of the neuronal activation in these regions with recognition performance, which will be discussed in section 1.2.3.1.

Convergence of visual cues

Several studies revealed that high order object related areas are activated by the presentation of an object shape, regardless of its physical properties. It has been shown that motion, texture, and luminance contrast (Grill-Spector et al., 1998a) as well as stereo cues (Gilaie-Dotan et al., 2002) activate object-related areas, provided that they are used to define an object form. In addition object related regions are activated to the same extent by 3D objects depicted either in grayscale photographs or in line drawings (Kourtzi and Kanwisher, 2000).

These studies demonstrate convergence of visual cues at the level of individual voxels. Convergence of cues occurs at the single neuron level, in the macaque IT (Sary et al., 1993). However, due to the limited spatial resolution of the fMRI method (mm range), the question if this is so in the human object related areas remains open. In order to circumvent the limited spatial resolution of the fMRI method and study the neuronal properties at the sub-voxel resolution, Grill-Spector et al. (Grill-Spector et al., 1999; Grill-Spector and Malach, 2001) developed the fMRI-adaptation approach. This approach utilizes the reduction in fMRI
signal following repeated presentation of identical images, to ‘tag’ neuronal activation according to the adaptation level obtained for stimuli presented in different formats (Avidan et al., 2002a; Grill-Spector et al., 1999; Grill-Spector and Malach, 2001). Kourtzi et al. (Kourtzi and Kanwisher, 2000) used the adaptation approach to test whether convergence of visual cues actually occurs at the neuronal level within human object-related regions. Using an event-related fMRI adaptation paradigm, they presented subjects pairs of object stimuli, that could be either identical, or exhibiting the same object shape but presented in a different format (i.e. grayscale or line drawings). A similar adaptation level was found in occipito-temporal regions when objects were repeated regardless of the format in which they were presented, thus suggesting convergence of visual cues in the neuronal level within these regions.

A recent study (Amedi et al., 2001) showed that a sub-region within the object-related occipito-temporal regions was activated preferentially not only by visual objects cues but also by tactile ones. Interestingly, auditory object cues did not activate this region (Amedi et al., 2002). These results suggest that this sub-region is a bimodal visou-somatosensory, object-related area, rather than a multimodal one.

**Grouping and completion effects**

The results described in the previous section suggest that occipito-temporal regions play a critical role in object recognition. However, they are still consistent with the possibility that the representation within these regions is based on local elements or features (Tanaka, 1996; Tanaka et al., 1991) and they do not imply global grouping processes. Kourtzi et al. (Kourtzi and Kanwisher, 2001) addressed this issue by applying the event-related adaptation paradigm discussed above. They found that object-related regions exhibited adaptation for pairs of identically perceived shapes, regardless of their contours. In contrast, when the perceived shapes within a stimulus pair were different but had identical contours, no adaptation was found. These results show that high level, "holistic" information about object shape is represented within occipito-temporal object related cortex, rather than simple contours.
Following the same line of evidence, using a modified Rubin face-vase illusion Hasson et al. (Hasson et al., 2001) showed that global grouping processes contribute to face-related activation within these regions. They found preferential activation within face related regions in both pFs (FFA) and LO during epochs in which subjects perceived the face and not the vase shape, despite the fact that their local features and contours were identical. Consistent with this trend substantial object completion effects (Lerner et al., 2002) and selective activation by subjective contours (Mendola et al., 1999) were also found in occipito-temporal regions. Both these studies indicate that the object-selective activation in these regions goes far beyond the information available only in local feature elements.

1.2.3 Top-down modulation of visual object recognition

As evident from the previous sections, activity in high order object related areas largely departs from the physical properties of the visual stimulus to become more correlated with the actual perception of objects. While some of these high order characteristics such as contrast invariance (Chapter 2.1) might be constructed in a bottom up, hierarchical fashion, other properties must rely on top-down or extra-retinal information. This information encompasses the internal state of the observer (e.g. attentional state) as well as all prior knowledge one has regarding a specific stimulus and the context of its presentation. Thus, object perception and recognition is likely to depend on a continuous interaction between these two processing streams (Ahissar and Hochstein, 1997; Ullman, 1996).

1.2.3.1 Correlation between neuronal activity and object recognition

Strong correlation between neuronal activity in a given area and perception, as measured behaviorally, has been taken as evidence for that area’s contribution to the perceptual function under investigation (e.g. area MT in motion perception (Newsome et al., 1989)). Many studies point to occipito-temporal visual areas as playing a crucial role in visual awareness, or more specifically, showing correlation with the subjects’ performance.
Dolan et al. (Dolan et al., 1997) found enhanced fMRI signal in category-related ventral visual areas during a second exposure to displays containing degraded stimuli once the stimuli had become recognizable. Using a similar paradigm Tovee et al. (Tovee et al., 1996) reported that the firing rate of some inferotemporal cells to a degraded set of pictures (containing a face) increased considerably after exposure to the gray-scale version of the same pictures. Similarly Grill-Spector et al (Grill-Spector et al., 2000) found a correlation between subjects' ability to recognize briefly presented object images and the level of the fMRI signal. They showed an enhancement of the fMRI signal in occipito-temporal regions for a set of briefly presented, masked stimuli on which subjects were extensively trained compared with a novel set of stimuli. Using an event-related paradigm and briefly presented masked objects, Bar et al. (Bar et al., 2001) found stronger fMRI activation in anterior fusiform regions during trials in which subjects were successfully able to recognize the stimuli compared with trials in which they were unable to do so. Using binocular rivalry it has also been shown that category selective activation in ventral occipito-temporal regions (pFs and CoS) is seen when the object is perceived as figure, compared with when the same object is suppressed (Tong et al., 1998). Similar correlations with perception were also reported from single unit recordings in macaque inferior temporal cortex when binocular rivalry was used (Leopold and Logothetis, 1996; Sheinberg and Logothetis, 1997).

While the above studies as well as others report enhancement of the fMRI signal, or single neuronal activity, following improvement in performance others show an opposite effect. Several priming studies (in which repeated presentation of a supra-threshold stimulus leads to improved behavioral performance) reported a decrease in fMRI signal in tandem with the behavioral improvement (Buckner et al., 1998; Chao et al., 2002; van Turennout et al., 2000). Decreases in fMRI signal are also evident when applying the fMRI-adaptation approach (Grill-Spector et al., 1999; Grill-Spector and Malach, 2001) in which response reduction is typically obtained following multiple exposures to a stimulus, under conditions which do not pose particular difficulty for object recognition. While adaptation and priming effects have some aspects in common, they differ in others. For example manipulation in the viewing conditions of
face stimuli (position, size, and rotation) elicits different adaptation levels in occipito-temporal cortex, despite similar performance level. Furthermore, the response reduction of the fMRI signal following priming has been shown to persist several days after the original event (Chao et al., 2002; van Turennout et al., 2000), a finding that cannot easily be attributed to adaptation processes, which typically are short-lived. It could be that the signal reduction associated with both priming and adaptation result from two different mechanisms: Priming may be the manifestation of a learning mechanism, while adaptation may serve as an on-line temporal filter for redundant information, which focuses neuronal resources on novel visual information. To summarize, two separate but related issues should be addressed when studying stimulus repetition effects. The first is the neuronal mechanism of adaptation vs. priming effects and the second is decrease vs. increase of neuronal responses following improvement in behavioral performance.

How can the discrepancy between signal increase and decrease following behavioral improvement can be resolved? Two recent studies provide some explanation: Henson et al. (Henson et al., 2000) found that repetition of familiar stimuli (both faces and symbols) led to a signal reduction in a right fusiform region, while an enhanced response was found to the repetition of novel stimuli in this same region. These findings suggest that familiarity may have a role in determining whether the fMRI signal would be increased or decreased following priming (but somewhat different results were obtained by that group lately (Henson et al., 2002)). Using a different experimental approach emphasizing the temporal aspects of the recognition process, James et al. (James et al., 2000) showed a more complex interaction in which, during a pre-recognition phase, primed objects produced stronger response compared with non-primed ones, while after recognition, the trend was reversed and stronger activity was found for the unfamiliar stimuli. This effect, however, seemed to be related to a leftward shift in the time-course obtained for the primed compared with the non-primed stimuli. Finally, using perceptual hysteresis effects produced by increasing and decreasing the contrast level of letter stimuli Keinschmidt et al. (Kleinschmidt et al., 2002) showed an increase in the fMRI signal in occipito-temporal object-related areas before subjects reported pop-out of the stimuli, and a decrease of the signal immediately following recognition onset.
Thus, while correlations between brain activation and recognition are both interesting and appealing, the studies described above provide a complex situation and the relation between brain activation and behavioral performance is currently poorly understood. The collective picture suggests that this relation may be dominated by several different mechanisms having different dynamic aspects, which are not yet entirely clear. Generally, it seems that whenever stimuli are shown in sub-threshold conditions and therefore a buildup of a new representation takes place (e.g. Kleinschmidt et al., 2002 before pop-out; James et al., 2000 during pre-recognition phase; Grill-Spector et al., 2000; Dolan et al., 1997) the signal increases. But under supra-threshold conditions, in which a prior representation of the stimuli already exist, as in the more classical priming experiments described above, or after a clear percept has already been formed (e.g. James et al., 2000 post-recognition phase; Kleinschmidt et al., 2002 after pop-out), a reduction in the BOLD signal is obtained. The exact relation between priming-related signal reduction, which were recently found to be rather long lasting, and the more short-term adaptation effect, are a subject for further research.

1.2.3.2 Attentional modulation of object-related activation

Another parameter that complicates the relation between brain activation and recognition performance even further is the effect of behavioral context in general, and of attention in particular. Selective attention plays an important role in various aspects of behavior, and affects a number of performance measures, such as reaction time, response accuracy and perceptual learning (Ahissar and Hochstein, 1993; Joseph et al., 1997). Two basic features capture the essence of visual attention (Desimone and Duncan, 1995). The first is the limited capacity for processing information; at any given time, only a small amount of information available on the retina can be fully processed. The second is selectivity - the ability to filter out unwanted information. It appears that brain activity, in many cortical areas, can be selectively amplified or suppressed as a function of the attentional set. Two of the leading theories regarding the role and mechanism of selective attention are space (location) based attention, metaphorically referred to as the
"spotlight" theory of attention (Posner, 1980; Treisman and Gelade, 1980) and object (feature) based attention (Desimone and Duncan, 1995; Duncan, 1984; Vecera and Farah, 1994). Neural correlates of selective attention were demonstrated in several brain areas of the macaque monkey including V4 and IT (Chelazzi, 1995; Connor et al., 1997; Moran and Desimone, 1985; Motter, 1993) and even V1 (Motter, 1993; Roelfsema et al., 1998).

In recent years, several neuroimaging studies investigated attentional effects in various visual areas under different experimental paradigms. Corbetta et al. (Corbetta et al., 1990) using PET studies, showed that attention to different features in an image such as its color, shape, or motion activated segregated visual areas, each specializing in processing a certain visual feature. Other studies (Haxby et al., 1991; Haxby et al., 1994; Kohler et al., 1995) found, using attention to spatial location and shape identity paradigms that there is a dissociation in human extrastriate cortex between the dorsal and ventral pathways. Evidence for spatial, spotlight-like attentional effects were obtained recently using fMRI in retinotopic visual areas (Brefczynski and DeYoe, 1999; Martinez et al., 1999; Somers et al., 1999; Tootell et al., 1998a) but also in high order object related ones (Downing et al., 2001a). O'Craven et al. (O'Craven et al., 1999) also found object-based attentional mechanism within these regions. In Chapter 2.3 (Avidan et al., 2002c) I describe my results which show attentional modulation of the activation in high-order object related regions. The results show that this modulation is the outcome of a complex interaction between space and object-based attentional mechanism. My study as well as several others (Beauchamp et al., 1997; O'Craven et al., 1997; Rees et al., 1997; Tootell et al., 1998a) found attentional modulation also in the motion sensitive area MT/V5. The elevated signal in this region was obtained when subjects attended movement in visual stimuli.

It has been suggested that attentional influences are mediated by inputs from cortical areas outside the visual cortex, such as those in parietal cortex (Corbetta et al., 1998; Wojciulik and Kanwisher, 1999) and prefrontal cortex (Courtney et al., 1997; Lumer et al., 1998). However, these studies are beyond the scope of this chapter.
1.3 Methodological accounts

Functional MRI BOLD (Blood Oxygenation Level Dependent) signal measures the hemodynamic responses to changes in neural activity in the brain. When neural activity increases, regional cerebral blood flow increases, causing the local concentration of deoxygenated hemoglobin to decrease. This decrease results in the fMRI signal in the brain. Therefore, fMRI signal is an indirect measure of localized changes in neural activity, typically averaged over 2-6 sec and over few millimeters of cortex (DeYoe et al., 1994; Kwong et al., 1992). Thus, the nature of the quantitative relationship between the fMRI signal and the underlying spiking activity of neurons is crucial for understanding the implication of fMRI data obtained from human subjects on the one hand, and for relating this data to single-neuron recordings in monkeys, on the other.

Several studies attempted to study this issue: Rees et al (Rees et al., 2000) measured fMRI population responses in the human motion sensitive area MT/V5, and showed that these responses were directly proportional to the averaged neuronal firing rate measured in the homologue region in the macaque. A similar analysis performed by Heeger and colleagues (Heeger et al., 2000), on data obtained from human and monkey V1 as a function of stimulus contrast revealed similar proportion between these two measures. Logothetis et al. (Logothetis et al., 2001) have taken a different approach and examined the relationship between the BOLD fMRI signal and the underlying neuronal activity in simultaneous intracortical electrophysiology and imaging experiments in anesthetized monkeys. They suggested that local field potentials (LFP), and not multi-unit responses showed the strongest correlation to the BOLD signal, implying that activation in an area might reflect its incoming input and local processing rather than its spiking activity. Thus, while some progress had been made in understanding the neuronal bases of the fMRI signal, this important issue is still far from being fully resolved.
Chapter 2

Results

2.1
Contrast sensitivity in human visual areas and its relationship to object recognition.

Galia Avidan, Michal Harel, Talma Hendler, Dafna Ben-Bashat, Ehud Zohary, and Rafael Malach

Contrast Sensitivity in Human Visual Areas and Its Relationship to Object Recognition

GALIA AVIDAN,1,2 MICHAL HAREL,3 TALMA HENDLER,4,5 DAFNA BEN-BASHAT,4 EHUD ZOHARY,1,2 AND RAFAEL MALACH3

1The Interdisciplinary Center for Neural Computation and 2Department of Neurobiology, Hebrew University of Jerusalem, Jerusalem 91904; 3Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100; 4Imaging Department, Whol Institute for Advanced Imaging, Sourasky Medical Center, Tel Aviv 64239; and 5Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

Received 8 August 2001; accepted in final form 8 February 2002

Avidan, Galia, Michal Harel, Talma Hendler, Dafna Ben-Bashat, Ehud Zohary, and Rafael Malach. Contrast sensitivity in human visual areas and its relationship to object recognition. J Neurophysiol 87: 3102–3116, 2002; 10.1152/jn.00669.2001. An important characteristic of visual perception is the fact that object recognition is largely immune to changes in viewing conditions. This invariance is obtained within a sequence of ventral stream visual areas beginning in area V1 and ending in high order occipito-temporal object areas (the lateral occipital complex, LOC). Here we studied whether this transformation could be observed in the contrast response of these areas. Subjects were presented with line drawings of common objects and faces in five different contrast levels (0, 4, 6, 10, and 100%). Our results show that indeed there was a gradual trend of increasing contrast invariance moving from area V1, which manifested high sensitivity to contrast changes, to the LOC, which showed a significantly higher degree of invariance at suprathreshold contrasts (from 10 to 100%). The trend toward increased invariance could be observed for both face and object images; however, it was more complete for the face images, while object images still manifested substantial sensitivity to contrast changes. Control experiments ruled out the involvement of attention effects or hemodynamic “ceiling” in producing the contrast invariance. The transition from V1 to LOC was gradual with areas along the ventral stream becoming increasingly contrast-invariant. These results further stress the hierarchical and gradual nature of the transition from early retinotopic areas to high order ones, in the build-up of abstract object representations.

INTRODUCTION

Recently, several neuroimaging studies have revealed a high order cortical region, located at the occipito-temporal junction (the lateral occipital complex, LOC), which possesses a number of functional properties associated with high-level object-related representations. Thus the LOC has been shown to manifest a high degree of size and position invariance (Grill-Spector et al. 1999) and to be activated by image completion (Lerner et al. 2001b) and grouping processes (Hasson 2001; Kourtzi and Kanwisher 2001). These processes are remarkably similar to those encountered in recognition performance. Finally, using a backward masking paradigm, it was shown that the activation pattern in the LOC is highly correlated with the subjects’ recognition performance rather than the physical duration of stimulus exposure (Grill-Spector et al. 2000).

One issue that remains unresolved is the characteristic of the process by which the functional transformation from the retinal image to high-level object representation is accomplished. It is well established that a sequence of ventral stream object areas is involved (Felleman and Van Essen 1991; Lerner et al. 2001a; Tootell et al. 1996), but the relative contribution of each stage in the process is unknown. For example, it is not clear whether the transformation is gradual, where each stage in the sequence is contributing a small increment, or whether it occurs in a few large steps.

Here we used a single well-defined visual property, that of image contrast, to follow the transformation in image representation along the entire constellation of ventral stream human visual areas. Using this approach, differences between areas in terms of their contrast response function could be explored and related to the putative hierarchical processing which exists between visual cortical areas along the ventral stream (Ungerleider and Mishkin 1982).

Contrast is a suitable parameter to study because perceptually, object recognition is highly invariant to contrast changes beyond a minimal contrast level. However, retinal responses are highly sensitive to all contrast levels. Consequently, the contrast response function can be used as a tool to explore to what extent activation in a given visual area is determined by the physical contrast of the stimulus and to what extent it is related to the subject’s perceptual performance. The answer to this question could shed light on the nature of the hierarchical processing in the visual system. More specifically, it will aid in determining to what extent the establishment of contrast invariance is a gradual process, whether contrast invariance in a given cortical area is specific to particular object shapes, and which areas are most closely related to recognition performance.

Using functional magnetic resonance imaging (fMRI), we studied the contrast response function along the entire set of ventral-stream human visual areas. Our results reveal that the correlation between physical stimulus contrast and fMRI re-
Results

CONTRAST SENSITIVITY IN HUMAN VISUAL CORTEX

spontaneous shows a gradual and consistent decline as one moves to high-order visual areas along the ventral stream. Concurrently the fMRI signal shows consistently increasing correlation to recognition performance. Thus the two subdivisions of the lateral occipital complex: the dorsal lateral occipital region (LO) and the more ventral and temporal region located in the posterior fusiform gyrus (pFs) showed the strongest tendency toward contrast invariance especially for face stimuli.

These results reflect a hierarchical trend in the human visual cortex in which cortical responses gradually depart from the physical aspects of the visual stimulus and become correlated with perceptual experience. Some of these results have been published previously in abstract form (Avidan-Carmel et al. 2000).

METHODS

Subjects

Twelve healthy subjects (6 women, ages 24–50), participated in one or more of the experiments. All subjects had normal or corrected to normal vision and provided written informed consent. The Tel-Aviv Sourasky Medical Center approved the experimental protocol.

MRI setup

Subjects were scanned in a 1.5 Signa Horizon LX 8.25 GE scanner equipped with a standard birdcage head coil. In the block-design experiments (experiments 1, 2, 4, and 5), blood-oxygenation-level-dependent (BOLD) contrast was obtained with gradient-echo echo-planar imaging (EPI) sequence (TR = 3,000, TE = 55, flip angle = 90°, field of view 24 × 24 cm², matrix size 80 × 80). The scanned volume included 17 nearly axial slices of 4-mm thickness and 1-mm gap. In the event-related experiment (experiment 3), the scanning parameters were changed (TR = 1,500, TE = 55, flip angle = 70°) and the scanned volume included eight oblique slices. T1-weighted high-resolution (1 × 1 × 1 mm) anatomical images and three-dimensional spoiled gradient echo sequence were acquired in order to allow accurate cortical segmentation, reconstruction and volume-based statistical analysis.

Visual stimulation

Stimuli were generated on a PC and projected via an LCD projector (Epson MP 7200) onto a tangent screen positioned over the subject’s forehead and viewed through a tilted mirror located above subjects’ eyes.

Experiments

EXPERIMENT 1: FACES AND OBJECTS. Ten subjects participated in the experiment. The experiment (Fig. 1), which lasted 450 s, consisted of 12 different stimulus conditions and had 57 epochs which were presented in a block design paradigm. Stimuli were 19 × 17° black on white line drawings of faces and objects, and control stimuli were texture patterns (for an example, see Fig. 1). The face stimuli were either a woman, man or a child and the object stimuli included: man-made objects, images of vehicles and images of buildings. Illumination level of the white background was 97 cd/m² and of the black line drawings at 95.6% contrast was 2 cd/m² as measured directly from the tangent screen. Stimulus contrast was defined as follows where L is luminance

\[
\text{Stimulus Contrast} = \frac{L_{\text{max}} - L_{\text{min}}}{L_{\text{max}} + L_{\text{min}}}
\]

All contrast levels were verified by direct measurement from the tangent screen. Each face and object was presented in five different contrast levels (0, 4.4, 6.1, 9.8, and 95.6%), and each pattern stimulus was presented in two different contrast levels (9.8 and 95.6%). In the following text, these contrast levels will be rounded for convenience to integer level (e.g., contrast of 4.4% will be referred to as 4% contrast etc.). Each of the face and object conditions was repeated twice. Epochs were pseudo-randomized so that minimum interaction will be possible between occurrences of same stimuli presented in different contrast level. In three epochs (of 8) high-contrast stimuli appeared before the low-contrast epochs of the same stimuli. In all other cases, low-contrast epochs appeared before high-contrast epochs. Minimum of two intervening stimulus epochs were inserted between epochs containing the same visual stimuli but presented in different contrast levels. When using only two intervening stimulus epochs, the low-contrast level condition always preceded the high-contrast epoch of the same visual images to avoid possible adaptation effects. In addition, stimulus presentation order within each epoch was randomized, thus further minimizing such interactions. The pattern epochs were repeated four times each and, in addition, there were 29 interleaving blank epochs. Each stimulus epoch lasted 9 s, and each blank epoch lasted 6 s with the exception of the first and last blanks which lasted 21 and 15 s, respectively.

Within an epoch, each of the 18 images was presented for 200 ms followed by 300 ms of fixation point on a blank screen to minimize eye-movement effects. Contrast of stimuli was varied by changing the gray level of the line-drawing image (black at 100% contrast) while keeping the background (white) constant.

Subjects were instructed to fixate on a fixation point located in the middle of each stimulus and to covertly name each stimulus in the “object” conditions and to indicate whether the face was that of a child, man or woman. Although the latter task appears somewhat easier, in fact the performance of the two tasks was similar (see Fig. 6). To enable subjects to differentiate between low- or zero-contrast epochs and blank epochs, stimulus epochs had a gray fixation dot while blank epochs had a red fixation dot. One of the subjects who participated in the experiment was tested on an early version of the experiment, which did not include the 10% condition of the pattern stimuli, and, for two subjects, the contrast was changed slightly to 0, 5, 7, 10, and 100%.

Psychophysics. For each subject, recognition performance was measured while they were still in the magnet following the fMRI scan. As in the original experiment, subjects were required to name the objects as specifically as possible and to name subordinate categories within the faces (man, woman, child), only when they were asked to perform overt naming. The sequence of stimuli was identical to the MRI scan except that each stimulus was presented for 200 ms followed by 1800 ms to allow overt naming of each stimulus. In addition the pattern epochs were omitted.

EXPERIMENT 2: FACES, CARS, AND HOUSES. Eight subjects participated in experiment 2, which consisted of 12 different stimulus conditions. Stimuli were pictures of faces, houses, and cars presented at contrasts of 4, 6, 10, and 100%, and each condition was repeated twice. Stimuli were generated in the same way as in experiment 1. Presentation of stimuli and task (i.e., subordinate category naming) were identical to experiment 1.

EXPERIMENT 3: EVENT-RELATED CONTROL. Five subjects participated in experiment 3 in which we used the 100 and 10% contrast conditions. Stimuli were pictures of faces and objects that were used in experiment 1. Sixty-four presentations of 16 different faces and 16 different objects (2 contrast levels each) were presented in a counter-balanced event-related paradigm. Each stimulus was presented for 500 ms followed by 5,700 ms, and the experiment lasted 444 s. The experiment began with 24-s blank and ended with 18-s blank. In addition there were two more long blank epochs along the experiment, each lasted 9 s. The subjects’ task was to covertly name each of the stimuli that were presented.
EXPERIMENTS 4 AND 5: ATTENTION CONTROLS. Six subjects participated in both experiments 4 and 5. In these experiments, we used 54 different pictures of faces presented in 100 and 10% contrast. Each experiment lasted 228 s and consisted of 13 visual epochs of 9 s followed by a short blank of 6 s. The first visual epoch consisted of images of texture patterns and was not included in the statistical test. In addition there were two long blanks at the beginning and end of the experiment (21 and 12 s, respectively). Each picture was presented for 200 ms followed by 800 ms of blank. During the visual epochs, there was a small light-gray fixation point centered on each image while during the blank epochs the color of the fixation point was red.

In experiment 4, the color of the fixation point was changed once or twice per epoch to a darker gray, and subjects had to perform a one-back memory task on the color of the fixation point. In four of the six subjects that were scanned in each experiment, we measured performance during the fMRI scan. Subjects were instructed to covertly name each stimulus. B: a segment from the time axis of the experiment. An interleaved short-block design was used, with each block (epoch) consisting of 18 different stimuli from the depicted type (see METHODS for more details). The experiment lasted 450 s and included visual epochs of 9 s, and blank epochs of 6 s.

Mapping borders of visual areas

The representation of vertical and horizontal visual field meridians were mapped in all subjects to delineate borders of retinotopic areas (DeYoe et al. 1996; Grill-Spector et al. 1998a; Sereno et al. 1995). Visual stimulation was presented at a rate of 4 Hz in 18-s blocks and consisted of triangular wedges that compensated for the expanded foveal representation. The wedges were presented either vertically (upper or lower vertical meridians) or horizontally (left or right horizontal meridians). The wedges consisted of either gray-level natural images or black and white objects—from texture pictures (Grill-Spector et al. 1998a). Subjects were asked to fixate on a small central cross. Visual epochs alternated with 6-s blanks. Four cycles of the stimuli were shown.

Because the exact parceling of the ventral areas V4 and V8 is still debatable in the literature (Hadjikhani et al. 1998; Zeki and Marini 1998), we defined a combined focus V4/V8 for which the posterior border is the upper visual meridian representation. The anterior border was defined as the border passing through half field representation (a lower visual field representation, a horizontal one and an upper visual field representation). The collateral sulcus activation was defined as activation that was located anterior to and outside from the upper visual field representation and was therefore not retinotopic (see Fig. 2A).
FIG. 2. Functional activation maps, detailed meridian map, and contrast response functions of visually active areas from experiment 1. a: functional activation maps of all visually active areas from 1 subject are shown superimposed on the 2 hemispheres of an inflated brain seen from a ventral view (left), and on a flattened map of the right hemisphere (middle). Visual areas were delineated by superimposing meridian maps that were obtained on a separate scan. The meridian borders are indicated on the central flattened map by white dotted lines and denote the retinotopic areas V1, V2, Vp, and V4/V8 in the ventral visual pathway and V3, V3A, V7, and intra-parietal sulcus (IPS) in the dorsal visual pathway. Additional visual activation was found in the lateral occipital (LO) and posterior fusiform (pFs) and in the collateral sulcus (Coll.). Color scale indicates the degree of correlation to the statistical paradigm. R, right; L, left; A, Anterior; and P, posterior. In addition, we also show (rightmost panel) a detailed meridian map of the same hemisphere, the retinotopic borders are now indicated by the red dotted lines. Yellow: horizontal visual meridian; green: lower visual meridian; blue: upper visual meridian. b: contrast response functions of visually active areas of the ventral pathway (averaged across all subjects). y axis denotes functional magnetic resonance imaging (fMRI) activation level (% signal change) and x axis denotes log contrast level. Red curves denote activation profiles for faces and green for object images. Early visual areas manifested strong contrast dependence at suprathreshold contrast levels (>10%), for both faces and objects while higher order, object-related areas (LO, pFs) were more invariant to contrast changes at this range. This invariance was stronger for faces compared with common objects. Error bars indicate ±SE calculated across subjects.
Results

AVIDAN, HAREL, HENDLER, BEN-BASHAT, ZOHARY, AND MALACH

Fig. 3. Raw time-course data from experiment 1. Averaged time-course data across 10 subjects. The percent signal change from a blank baseline is shown (y axis) against time for areas V1, LO, and pFs. Repetitions of the same experimental condition were averaged. Gray level indicates different contrast level for the 3 stimulus types used in the experiment (i.e., faces, objects, and patterns). Black bar in the first condition of area V1 represents the epoch length (i.e., 9 s) and the onset and offset of the visual stimulation. Error bars indicate ± SE across subjects.

Data analysis

fMRI data were analyzed with the “BrainVoyager” software package (Brain Innovation, Maastricht, Netherlands) and with complementary in-house software. The data of each subject from each scan were analyzed separately. The first three images of each functional scan were discarded and a hemodynamic lag of 3 s was assumed. The functional images were superimposed on two-dimensional (2D) anatomical images and incorporated into the three-dimensional (3D) data sets through trilinear interpolation. The complete data set was transformed into Talairach space (Talairach and Tournoux 1988). Preprocessing of functional scans included 3D-motion correction and filtering out of low frequencies up to five cycles per experiment (slow drift).

Statistical analysis was based on the General Linear Model (Friston et al. 1995). In this analysis a linear combination of several predictor variables is used to predict the variation of an observed variable y

\[ y(t) = \hat{y}(t) + e(t) = b_0 + b_1 x_1(t) + b_2 x_2(t) + \ldots + b_n x_n(t) + e(t) \]

where y is the observed signal time course, the \( x_i \) are explanatory variables, the \( b_i \) are the regressor values and \( e(t) \) is an error term for the unexplained deviation of the estimated \( \hat{y}(t) \) from the measured signal value \( y(t) \) at each time point. The GLM analysis is performed independently for the time course of each individual voxel. The results of a GLM analysis of a voxel time course are estimates for the regression weights \( b_i \) such that the predicted values \( \hat{y}(t) \) are as close as possible to the measured values \( y(t) \) at each time point. The least-squares method is used for estimating the regression weights such that the error values \( e(t) \) are minimized

\[ \sum e(t)^2 = \sum [y(t) - \hat{y}(t)]^2 = \min \]

Specifically in all experiments described in this paper, each experimental condition (except for blank) was defined as a separate predictor. A box-car shape was used for each predictor according to the stimulus epochs and a hemodynamic lag of 3 s was assumed. Activation maps were obtained by running this analysis (see Fig. 2A, left and middle) and visualizing the multiple regression value (r value) at each voxel (similar to a correlation r value after standard correlation analysis). The r values represent goodness of fit of the specified full model to the signal time course at a given voxel.

When mapping the relative contribution of two functional responses (Fig. 5A, all face epochs vs. all object epochs in that example), the color coding represents the relative contribution of either set. If both predictor sets contribute roughly equally to the activation at a voxel, this voxel will be colored in blue. Green and red colors show strong contribution of one predictor set over the other. The exact color used depends on the level of differential contributions by each predictor set.

Percent signal change for each subject in each experiment was calculated as the percent activation from a blank baseline

\[ \text{Percent Signal} = \frac{\text{signal} - \text{mean[signal (blanks)]}}{\text{mean[signal (blanks)]}} \times 100 \]

Due to the incorporation of the 3-s hemodynamic lag, all time points of each epoch were included in the calculation of the mean signal for each epoch. The data shown in Fig. 3 are taken from all three time points of each epoch and include also one time point preceding each epoch and two time points following each epoch.

All epochs belonging to the same condition were averaged together to provide an average condition epoch time course (Figs. 2B, 3, and 5B) error bars indicate the standard error of the mean in each condition across all subjects.

The cortical surface was reconstructed from the 3D-spoiled gradient echo scan. The procedure included segmentation of the white matter using a grow-region function, the smooth covering of...
Results

CONTRAST SENSITIVITY IN HUMAN VISUAL CORTEX

23

a sphere around the segmented region, and the expansion of the reconstructed white matter into the gray matter. The sulci were smoothed using a cortical “inflation” procedure. The surface was cut along the Calcarine sulcus and unfolded into the flattened format. The obtained activation maps were superimposed on the unfolded cortex and the Talairach coordinates were determined for the center of each region of interest (ROI).

Contrast invariance ratio: to evaluate quantitatively the differences in the fMRI signal for the 100 and 10% contrast epochs for the faces and objects separately, we calculated contrast invariance ratio (Fig. 4, Table 1)

Contrast Invariance Ratio

\[ = 1 - \frac{\% \text{ signal (100% epochs)} - \% \text{ signal (10% epochs)}}{\% \text{ signal (100% epochs)} + \% \text{ signal (10% epochs)}} \]

The contrast invariance ratio was calculated using the signal from the suprathreshold contrast levels of 100 and 10% because this was the range in which subjects’ performance was most invariant as evident in Fig. 6. Recognition performance was reduced substantially below 10% contrast.

Normalization of fMRI signal and correlation to psychophysical data. To compare fMRI signal and recognition performance of the subjects, the fMRI signal was normalized for each subject for each visual area so that it will range between 0 and 100%. The signal was normalized by referring to the activation level for the maximal contrast level to faces and object images, respectively, as the maximum activation level (100%) and calculating the activation for the rest of the contrast levels (4, 6, and 10%) relative to that signal level. Recognition performance (percent correct) of each subject is presented in Fig. 6. To evaluate the degree of correlation between the normalized fMRI signal and the recognition performance of the subjects, we first calculated the difference between the norm. fMRI signal in the two suprathreshold contrast values of 100 and 10% as well as the difference of the recognition performance between these levels for each subject separately for the face and object stimuli. We then defined an activation-to-performance distance measure between these differences in each area. The distance measure was defined as [(100 – norm. fMRI pcs at 10%) – (%correct at 100% – %correct at 10%)]. A one-way paired t-test was conducted to find whether the distance measure in area V1 was significantly greater then in the high-order visual areas: LO, pFs, and the Coll.

R E S U L T S

Experiment 1

The aim of experiment 1 was to characterize the fMRI response of various visual areas to different contrast levels of visual stimuli. The stimuli we used were black-and-white line drawings of either objects or faces that were presented at five different contrast levels (0, 4, 6, 10, and 100%) in a short-block design paradigm (see METHODS and Fig. 1). Control stimuli were texture patterns that were presented at two contrast levels (10 and 100%). Epochs containing visual stimuli (including the 0% contrast epochs) were indicated by a gray fixation point at the center of each image, while interleaved blank epochs were identified by a red fixation point, located at the center of the screen. This was done mainly to differentiate between low-contrast epochs that engaged subjects’ attention when attempting to recognize the objects from the blank epochs, which required only fixation.

Contrast response in visually active areas

Data from 10 subjects were analyzed. The basic statistical test applied to the data searched for voxels that were activated by all visual images (objects, faces, and patterns) irrespective of their contrast level, compared with blank (visual > blank). This test revealed activation in the entire set of visual areas as demonstrated in one representative subject in Fig. 2A. Activation maps are presented in two different formats. On the left, the data are shown on both hemispheres of an inflated brain seen from a ventral view. In the middle, data are shown on a flattened map of the right hemisphere. The visual areas of each

---

**TABLE 1. Talairach coordinates**

<table>
<thead>
<tr>
<th></th>
<th>Left</th>
<th>Right</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>LO</td>
<td>-46 ± 3</td>
<td>-69 ± 6</td>
</tr>
<tr>
<td>pFs</td>
<td>-40 ± 5</td>
<td>-54 ± 7</td>
</tr>
</tbody>
</table>

Talairach coordinates (Talairach and Tournoux 1988) for lateral occipital (LO) and posterior fusiform (pFs) foci, (obtained by the statistical test: faces > objects) derived across the 10 subjects who participated in experiment 1. LO and pFs were separated based on anatomical criteria: LO focus was located in the lateral aspect of the occipital lobe while the pFs focus was located within the vicinity of the fusiform gyrus. Values represent the mean ± SD in millimeters.

---

FIG. 4. Contrast-invariance ratio. Contrast invariance ratio for suprathreshold contrast levels (see METHODS) was calculated for each of the visual areas for each of the 3 types of visual stimuli used in exp. 1. Left: the data obtained for the face stimuli; middle: data for objects; right: data for the pattern stimuli. Note the gradual increase in the contrast invariance ratio going from early retinotopic visual areas, which manifested high sensitivity to changes at suprathreshold contrast levels, to higher object areas, which manifested a high degree of contrast invariance for faces and objects. This gradual increase in the contrast invariance ratio was not significant for the pattern images. Error bars indicate SE calculated across subjects. *; significance level as calculated by a t-test comparing the invariance ratio obtained in each area to the ratio obtained for area V1 (*P < 0.05, **P < 0.005, ***P < 0.0005).
subject were delineated by superimposing meridian maps that were obtained in a separate scan on the flattened activation maps obtained in the current experiment (see METHODS). The meridian borders are indicated on the central flattened map by white dotted lines. A detailed meridian map of the same subject is shown on the right, the same retinotopic borders as in the middle image are indicated by the red dotted lines.

This statistical test highlighted activation in the retinotopic areas, stretching from V1 to V4/V8 ventrally (see METHODS) and to V3A dorsally. High-order activation was found in two main foci: LO focus, which was situated ventrally and posteriorly to area MT and extending into the posterior inferotemporal sulcus, and a focus in the vicinity of the pFs, which is anterior and lateral to area V4/V8 and extending into the inferior temporal sulcus (see Table 1 for Talairach coordinates). The latter focus (pFs) may overlap the fusiform face area (FFA) described previously (Kanwisher et al. 1997) (see Fig. 5A). Another focus was situated within the anterior portion of the collateral sulcus (Coll., see METHODS).

In the dorsal pathway, there were two additional foci, one located adjacent to the upper visual field representation of area V3A, probably corresponding to area V7 (Hadjikhani et al. 1998; Mendola et al. 1999), and another region, located within the IPS.

After establishing an anatomical definition for each activation focus, we derived the average activation profiles (i.e.,

![Image](image_url)

**FIG. 5.** Heterogeneity of object-selective activation. a: activation maps of face and object images: data are shown on the 2 hemispheres of 1 subject. Voxels were color coded according to the relative contribution of 2 predictors: face epochs (red) and object epochs (green) regardless of contrast level. Blue indicates balanced activation for both predictors. Retinotopic borders are indicated by the white dotted lines. b: contrast response functions of specific shape-selective regions: contrast response functions were obtained from object-related voxels in the collateral sulcus (Object Coll.) Face-related voxels were sampled from LO and pFs. Similar to the results shown in Fig. 2b for higher-order areas, clear contrast invariance existed for faces and to a lesser degree for objects.
contrast response functions) for each cortical area for each subject using the flattened brain format. Figure 2B shows the contrast response function (averaged across all subjects) for the various areas in the ventral pathway. Red and green graphs denote activation profiles for face and object images, respectively. A conspicuous difference was observed in the contrast response function of early and intermediate versus higher visual areas. Early and intermediate visual areas (V1-Vp, V4/V8, respectively) manifested strong contrast dependence at suprathreshold contrast levels, for both faces and objects, so that when the contrast was lowered from 100 to 10%, signal intensity was reduced to about half. On the other hand, higher-order, object-related areas (LO, pFs) showed a significantly lower contrast dependence at this range. This effect was clearly evident for face images and was somewhat weaker for objects.

Figure 3 shows time-course data averaged across 10 subjects from areas V1, LO, and the pFs for the face, object, and pattern stimuli. As shown in Fig. 2B, the marked difference in terms of contrast response function between activation in primary visual cortex (V1) and higher-order areas (LO, pFs) is evident.

To make sure that activation in lower visual areas was not underestimated due to the statistical test used (visual > blank), activation in areas V1 and V2 was also sampled from 4 subjects by a statistical test comparing activation in 100% contrast epochs versus blank and ignoring the rest of the epochs (all 100% epochs > blank). Comparing the number of activated V1/V2 voxels in this test versus the former one (using the same statistical threshold for each subject) revealed no significant difference (paired t-test P < 0.19) but a trend of reduction in the number of voxels in the latter test (all 100% epochs > blank) probably due to its weaker statistical power.

To obtain a quantitative comparison of the level of suprathreshold contrast invariance across the different areas, we calculated a “contrast invariance ratio” (see Methods). High levels of this ratio indicate greater invariance to contrast changes. Figure 4 exhibits the contrast invariance ratio for each of the visual areas presented in Fig. 2B for each type of visual stimuli used in the experiment. The leftmost bar graph represents the data obtained for the face stimuli, the middle one for the objects, and the rightmost histogram for the pattern stimuli. Note the gradual increase in the contrast invariance ratio going from early retinotopic visual areas (V1-Vp), which manifested high sensitivity to changes in contrast, through intermediate areas (V4/V8) which showed less sensitivity to contrast changes, to higher, nonretinotopic areas (LO, pFs, Coll.), which manifested a high degree of contrast invariance, particularly for faces but also for objects compared with area V1. [V1, contrast invariance ratio: 0.59 ± 0.10 (mean ± SE), 0.53 ± 0.07 for faces and objects, respectively; LO, contrast invariance ratio: 1.06 ± 0.04, 0.78 ± 0.06; pFs, 0.94 ± 0.05 0.75 ± 0.07 for faces and objects, respectively, and see Table 2 for ratios of all areas.]

It should be noted that in this experiment there were epochs that contained 0% contrast stimuli in which subjects were instructed to attempt to recognize objects, so that imagery and/or expectation effects could be detected. Across all areas the activation for the 0% contrast epochs was not significantly different from zero (t-test, P < 0.10). Thus it seems that under the specific conditions of the present experiment there was no clear evidence for a component of imagery or expectation-related activation in any of the studied areas. Such effects were reported by other studies, (Ishai et al. 2000; Kastner et al. 1999). This difference may be due to the fact that in the present experiment subjects were not required to actively try to imagine stimuli in low-contrast epochs.

An interesting question is whether the trends toward increased invariance continued beyond the LOC. To test this possibility, we looked at activation found for the same stimuli in different contrast levels. This was done by repeated presentation of the same stimuli in different contrast levels, which might cause contrast adaptation due to repeated presentation of the same stimuli in different contrast levels. This was done by pseudo-randomization of the experimental design (see Methods for details). However, in three epochs (of 8), high-contrast stimuli appeared before the low-contrast epochs of the same stimuli. In all other cases, low-contrast epochs appeared before high-contrast epochs.

<table>
<thead>
<tr>
<th>Experiment 1</th>
<th>Faces</th>
<th>Objects/Houses†</th>
<th>Patterns/Cars†</th>
</tr>
</thead>
<tbody>
<tr>
<td>V1</td>
<td>0.59 ± 0.10</td>
<td>0.53 ± 0.07</td>
<td>0.54 ± 0.10</td>
</tr>
<tr>
<td>V2</td>
<td>0.74 ± 0.08</td>
<td>0.59 ± 0.04</td>
<td>0.63 ± 0.05</td>
</tr>
<tr>
<td>Vp</td>
<td>0.82 ± 0.09</td>
<td>0.66 ± 0.03</td>
<td>0.62 ± 0.04</td>
</tr>
<tr>
<td>V4/V8</td>
<td>0.88 ± 0.06*</td>
<td>0.74 ± 0.06*</td>
<td>0.65 ± 0.07</td>
</tr>
<tr>
<td>Coll.</td>
<td>0.89 ± 0.12</td>
<td>0.81 ± 0.05**</td>
<td>0.68 ± 0.04</td>
</tr>
<tr>
<td>LO</td>
<td>1.06 ± 0.04***</td>
<td>0.78 ± 0.06*</td>
<td>0.71 ± 0.12</td>
</tr>
<tr>
<td>pFs</td>
<td>0.94 ± 0.05**</td>
<td>0.75 ± 0.07**</td>
<td>0.71 ± 0.13</td>
</tr>
<tr>
<td>Obj. Coll.</td>
<td>0.90 ± 0.08**</td>
<td>0.79 ± 0.05*</td>
<td>0.68 ± 0.04</td>
</tr>
<tr>
<td>Face LO</td>
<td>1.00 ± 0.04***</td>
<td>0.58 ± 0.12</td>
<td>0.94 ± 0.28</td>
</tr>
<tr>
<td>Face pFs</td>
<td>0.96 ± 0.04*</td>
<td>0.59 ± 0.13</td>
<td>0.68 ± 0.44</td>
</tr>
<tr>
<td>Dorsal pathway</td>
<td>V3</td>
<td>0.92 ± 0.07*</td>
<td>0.51 ± 0.08</td>
</tr>
<tr>
<td></td>
<td>V3A</td>
<td>1.03 ± 0.07**</td>
<td>0.77 ± 0.05</td>
</tr>
<tr>
<td></td>
<td>V7</td>
<td>1.05 ± 0.09**</td>
<td>0.81 ± 0.07*</td>
</tr>
<tr>
<td></td>
<td>IPS</td>
<td>1.03 ± 0.12*</td>
<td>0.81 ± 0.07*</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>LOC</td>
<td>0.91 ± 0.04</td>
<td>1.24 ± 0.14*</td>
</tr>
<tr>
<td></td>
<td>Coll.</td>
<td>0.65 ± 0.11</td>
<td>0.92 ± 0.06</td>
</tr>
<tr>
<td></td>
<td>House Coll.</td>
<td>0.42 ± 0.15</td>
<td>0.89 ± 0.07</td>
</tr>
<tr>
<td></td>
<td>Face LOC</td>
<td>0.97 ± 0.06</td>
<td>1.41 ± 0.20</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>V1</td>
<td>0.43 ± 1.10</td>
<td>0.36 ± 0.37</td>
</tr>
<tr>
<td></td>
<td>LOC</td>
<td>0.88 ± 0.18*</td>
<td>0.84 ± 0.13</td>
</tr>
</tbody>
</table>
Contrast response in specific object-category regions

An interesting question is whether the contrast response function in high-order areas is related to the shape selectivity of the different foci of activation. To explore this issue, we conducted additional statistical tests that looked for the specific functional signature of object-selective brain regions (Epstein and Kanwisher 1998; Ishai et al. 1999; Kanwisher et al. 1997). Figure 5A shows activation maps of the left and right hemispheres of one subject. In this map, voxels were color coded according to the relative contribution of two predictors (Friston et al. 1995; Goebel et al. 1998) (and see METHODS for details of analysis). Specifically, voxels were color coded according to their activation by face epochs (red), object epochs (green), and both (blue), regardless of their contrast level—note that this test is somewhat different from conventional object-selectivity tests, which typically use exclusively high-contrast images. Retinotopic borders are indicated by the white dotted lines.

Face-related voxels (red) appeared mainly within LO and the pFs (arrows). Voxels that showed the highest selectivity for faces within the pFs are marked as the fusiform face area (FFA) (Kanwisher et al. 1997). Except for this clear preference for face-related activation, both LO and pFs tended to exhibit a rather balanced activation for both faces and objects as indicated by the blue color in the vicinity of these areas. Voxels that showed preferential activation for objects versus faces (green) appeared mainly in the collateral sulcus. In addition, such balanced activation was also found in several brain regions typically stretching from area V3A, V7 and the IPS dorsally to area Vp, V4/V8 ventrally. Figure 5B depicts results that were obtained by using the objects > faces test and the faces > objects test. Activation profiles from the objects > faces test, were sampled from the collateral sulcus (object coll.) and activation profiles from the faces > objects test were sampled from LO and the pFs (face LO, face pFs). Similar to the results shown in Fig. 2B for higher-order areas, contrast invariance existed for faces and to a lesser degree for objects (see Table 2 for contrast invariance ratios).

Psychophysical experiment

An issue of interest is the correspondence between brain activation and human performance. To explore this relationship to the contrast response, all 10 subjects from experiment 1 also participated in a psychophysical experiment that was conducted in the magnet at the end of the scanning session under the same viewing conditions. In this experiment, subjects were shown the same set of images as in experiment 1 (except for the pattern stimuli), only this time they were asked to overtly name each stimulus. The recognition performance of the subjects is presented in Fig. 6. The averaged recognition performance (percent correct) for each contrast level across the 10 subjects was: faces: 100%: 100 ± 0%; 10%: 96.6 ± 2.2%; 6%: 73.6 ± 6.2%; 4%: 23.9 ± 7.9%; objects: 100%: 99.7 ± 0.3%; 10%: 97.5 ± 0.7%; 6%: 61.9 ± 6.73%; 4%: 26.6 ± 7.9%; means ± SE.

To compare the fMRI signal of the subjects to their recognition performance, we normalized the fMRI signal for each subject (see METHODS). This was done separately for the signal for faces and objects in areas V1, LO, pFs, and the Coll. A distance measure between the norm. fMRI signal and recognition performance was calculated for each subject for the face and object stimuli (METHODS). A t-test revealed that fMRI signal in the high-order object related areas LO, pFs, and the Coll. was significantly (paired t-test, P < 0.05) more correlated to recognition performance compared with area V1.

Control experiments

In addition to the main experiment (experiment 1), we conducted several control experiments. Because the contrast response function of LO and the pFs was not significantly different (ANOVA: 2-factor analysis, P < 0.75), for simplicity of presenting the control data in this section we averaged them together to a combined focus termed the lateral occipital complex (LOC).

Experiment 2

While in LOC the activation caused by face stimuli was invariant to changes from 100 to 10% contrast, activation
for the object images did not reach the same degree of invariance. A main difference between these two types of stimuli is that the object images, unlike faces, included a diverse set of shapes. To examine the impact of shape diversity, we conducted another experiment (experiment 2) in which we used two well-defined object categories, houses and cars, which have a narrower shape diversity compared with common objects. In addition, we included the face images used in the original experiment (experiment 1). Each image was presented in 4 contrast levels (4, 6, 10, and 100%). The results of this experiment are summarized in Table 2. Three different statistical tests were used: visual > blank, faces > houses, and houses > faces. In agreement with the results obtained in experiment 1, the response to faces was highly invariant to contrast changes within the LOC. In the collateral sulcus and LOC, the contrast invariance ratio for house stimuli was indeed higher than the one obtained in experiment 1 when using objects from various categories (see Table 2). On the other hand, the contrast invariance ratio in LOC for the second category of images (cars) was not substantially different from the results obtained for the mixed object stimuli in experiment 1 (see Table 2). Thus it seems that shape diversity was not the only factor contributing to the lower contrast invariance for objects compared with faces.

**Experiment 3**

It could be argued that the contrast invariance measured in high-order visual areas is a result of a saturation (“ceiling”) of the fMRI hemodynamic signal and thus does not reflect a neuronal contrast invariance. To rule out this possible confound, we conducted another experiment in which we used the 100 and 10% face and object images from experiment 1. However, this time we used an event-related presentation paradigm, which reduces the signal by approximately an order of magnitude, thus ensuring that it would not saturate. The results of that experiment are depicted in Fig. 7 which shows the activation profiles for faces (red) and for objects (green) of V1 and the LOC (see Table 2 for exact ratios). As in the block-design experiment, area V1 was highly sensitive to contrast changes, while the LOC showed a high degree of contrast invariance for object stimuli and complete invariance to contrast for face images. These results match the results of the block-design experiment from the anatomical point of view as well.

**Experiments 4 and 5**

Attention and task demands were previously shown to modulate the activation in high-order visual areas. It could be argued that the contrast invariance measured in LOC is a result of such effects and thus does not reflect contrast invariance of the neurons in that area. To rule out this possible confound, we conducted two additional experiments in which we used 100 and 10% face stimuli that were presented in a block-design fashion (see METHODS for details). The aim of the first experiment (experiment 4) was to explore whether attention could be the source for the contrast invariance found in LOC. This was done by instructing the subjects to perform an attention demanding foveal task and thus focusing their attention away from the face stimuli presented in the experiment. Specifically, the fixation point, centered on each image, changed its color once or twice in each visual epoch from light to darker gray.

**FIG. 7.** Results of the event-related experiment. **A:** activation profiles for faces (red bars) and objects (green bars) of visually active voxels located in area V1 and in LOC obtained from the event-related experiment using 10 and 100% contrast levels. These voxels were selected by applying the statistical test visual > blank. As in the block-design experiment, V1 was highly sensitive to contrast changes, whereas LOC showed a high degree of contrast invariance at suprathreshold contrast levels. Note that the invariance to contrast changes was maintained in LOC despite a approximately 10-fold reduction in fMRI signal. Error bars indicate SE calculated across subjects. **B:** activation map of the flattened right hemisphere of 1 subject, conventions are as in Fig. 24.
Subjects had to perform a one-back memory task and to report via pressing on a response box whether the fixation point changed its color or not. Note that the task was identical during the 10 and 100% contrast epoch. The aim of the second experiment was to explore whether changing task demands could affect the contrast response found in LOC. In that experiment (experiment 5), subjects had to perform a one-back memory task on the identity of the face stimuli. Note that this task is markedly different from the covert-naming task used in experiment 1.

As in the analysis of experiment 1 also in the analysis of experiments 4 and 5, LOC voxels were sampled from a statistical test searching for all visually active voxels (visual > blank). The results of these two experiments are shown in Fig. 8, the bar graph shows the contrast invariance ratio (left y-axis) calculated for the LOC in experiment 1 (original, for the face stimuli), experiment 4 (attention to fixation), and experiment 5 (attention to faces). In addition, the right y-axis represents subjects performance in all 3 tasks during the 100% contrast epochs and during the 10% contrast epochs (right y-axis). Note the similarity of the contrast invariance ratio obtained in the 3 different experiments. This implies that the contrast invariance found in the LOC is not a result of specific task demands.

Overall activation level (averaged percent signal change across subjects) was slightly reduced in the attention-to-fixation task (experiment 4) compared with the attention-to-faces task (experiment 5; attention to fixation: 100% contrast: 1.14 ± 0.08%, 10% contrast: 1.11 ± 0.08, attention to faces: 100% contrast: 1.24 ± 0.14, 10% contrast: 1.28 ± 0.13). Regarding subjects performance: in both experiments (experiments 4 and 5), task performance was not significantly different (P < 0.15) during the 100% versus 10% contrast epochs [experiment 4: 100% contrast: 84 ± 10% (mean ± SD), 10% contrast: 80 ± 12 experiment 5: 100% contrast: 94 ± 5 10% contrast: 91 ± 5]. It is important to note that performance was high for all three tasks tested although in the attention-to-fixation task (experiment 4) performance was somewhat lower, which implies that this task was the most demanding.

**Discussion**

Hierarchical processing reflected in the contrast sensitivity of visual areas

Our results show that the contrast response profile of visual areas changes along the cortical hierarchy, moving from strong contrast dependence in early visual areas, to a contrast invariance of varying degree in high order object areas. Is this transformation along the ventral visual pathway a gradual process or involves abrupt transition along particular visual areas? In the present experiment, we took advantage of the large coverage offered by the fMRI method and obtained a detailed analysis of the contrast sensitivity across the entire constellation of human visual areas for an identical set of stimuli. In our previous backward-masking experiment (Grill-Spector et al. 2000), the visual mask employed to limit image exposure activated by itself early visual areas and thus precluded the analysis of their object-related signal. The present study provides a comprehensive comparison of a specific functional response across the various visual areas. A comparison across different visual areas was also performed in other studies but to different factors than the current one (e.g., Polonsky et al. 2000; Tootell et al. 1998). The main question that such analysis allowed us to answer is whether the transition from early contrast-sensitive areas to high-order invariant regions was a gradual, monotonic process, or whether it happened in a single large step. Our results (Figs. 2B and 4) clearly point to a gradual process, which follows nicely the putative cortical hierarchy (i.e., V1, V2, Vp, V4/V8, and finally LOC).

Another related question is whether the transformation in the sensitivity to contrast changes achieves its highest level at the LOC, or whether it continues at more frontal cortical regions. This is particularly relevant in the case of the object images, which showed lower contrast invariance effects compared with faces. Interestingly, our analysis of frontal cortical regions did not show a significantly enhanced invariance—so it appears that the contrast invariance effect reaches its highest degree already at the LOC level.

Object-selective heterogeneity and the contrast invariance levels

Although our results clearly show a gradual increase in contrast invariance as one moves toward occipito-temporal cortex, we did find substantial changes in the level of this invariance for different image categories within LOC. More specifically, activation to face images, as well as activation in face-related regions was much more invariant to contrast changes compared with activation elicited by common objects. The source of such heterogeneity is not clear at this stage. One possibility is that the higher stages of the cortical hierarchy are better activated by face images compared with other objects. In this sense, the movement from object activation to face-specific activation is an extension of the general hierarchical trend to increased contrast invariance discussed earlier.

An alternative possibility is that the face images were more similar to each other within a block compared with the mixed objects epochs and this similarity affected the level of contrast...
invariance. To test this possibility, we ran experiment 2 in which the invariance to three specific object categories (faces, cars, and houses) was compared. However, the results of that experiment were mixed: we found an elevation of the contrast invariance in the collateral sulcus for the house stimuli compared with the case when a diverse set of objects was used (mixed-objects condition in experiment 1, see Table 2). However, the activation for the car images was very similar to that obtained for the mixed-objects condition. Thus it seems that the level of the contrast invariance is determined by a complex interaction of various factors, and shape diversity was certainly not the only factor contributing to the lower contrast invariance for objects compared with faces.

Finally, it should be noted that several lines of evidence have suggested that face recognition may be a special process, stressing the importance of the holistic representation of faces comparing to other object categories (Farah et al. 1998; Kanwisher 2000; Moscovitch and Moscovitch 2000). Hence, it may be that the unique properties of face processing are the source for the greater contrast invariance obtained for faces compared with other object categories. This, however, requires further investigation.

Correlation to object-recognition performance

Our results show a clear transition in the activation of cortical visual areas from strong contrast dependence in primary visual areas toward substantial contrast invariance in higher order occipito-temporal visual areas. A similar trend was found in the recognition performance of the subjects measured on the same stimuli (Fig. 6). Such correlation to recognition performance was found previously using other manipulations that degrade object recognition (Grill-Spector et al. 2000; James et al. 2000).

The correlation between fMRI activation and recognition performance may seem surprising given the indirect relationship between neuronal activity and the MRI signal (Logothetis et al. 2001). However, both in the case of the backward masking experiments as well as in the present contrast experiment, the manipulation involved crossing the recognition threshold. Thus it is plausible that neuronal populations were increasingly recruited as the contrast level was manipulated across recognition threshold concomitantly with recognition performance, leading to the positive correlation between the two. It should be emphasized that under different experimental situations, such as fMR-adaptation this correlation does not hold (Grill-Spector and Malach 2001). Furthermore, the correlation between psychophysical performance and fMRI signal in LOC was found when the subjects performed a specific task, i.e., object recognition. Different tasks’ requirements and different stimulus types may show tighter correlation to activity in other brain regions. Indeed, it has been shown that when the task and stimuli were tailored for optimally activating other areas such as primary visual cortex, V1 activity was more correlated with performance than in our case (Boynton et al. 1999; Huk and Heeger 2000; Ress et al. 2000). Following this rational, the present results further emphasize the involvement of the LOC in human object recognition.

From a broader perspective of the object recognition processes, the transformation toward contrast invariance that was found along the human visual ventral stream (Fig. 9) is yet another example of a visual process enabling object constancy (Grill-Spector et al. 1999; Gross 1972; Ito et al. 1995; Sary et al. 1993). In this respect, the present results extend our previous findings of position and size invariance in the LOC (Grill-Spector et al. 1999). A common theme to all these processes is that the cortical representation departs from the variable retinal activity patterns caused by changes in the viewing conditions (such as retinal size, retinal position, etc.) and becomes more attuned to the invariant, intrinsic properties of objects in the visual world.
real environment. Such transformation of object representation is an essential characteristic of visual perception.

Could hemodynamic nonlinearities account for the contrast invariance?

The hemodynamic signal is assumed to be an indirect measure of the neuronal response. Thus it is important to establish that the hemodynamic activation profile obtained by fMRI mirrors the average activity of the neurons in the same brain area. This has been recently suggested by Heeger et al. (2000), who showed that the contrast-response function obtained using fMRI in human V1 is closely correlated with the average single-unit activity measured in V1 of the macaque monkey. In two other recent papers, the close correlation between fMRI and neuronal activity was shown for human and monkeys MT (Heeger et al. 1999; Rees et al. 2000).

In our experiment, we found that the fMRI signal reached an asymptotic level at contrast levels more than 10% in high-order visual areas (LOC). A major concern is that in the LOC, the hemodynamic signal may reach saturation while the neuronal response would continue to increase with elevated contrast. Thus it could be that the contrast invariance measured in high-order visual areas is a result of hemodynamic signal saturation and not a characteristic feature of these areas.

The fact that the contrast invariance was found in specific cortical regions and not in others argues against a generalized hemodynamic effect, which presumably should not show such highly localized heterogeneity. However, to address this issue directly, we compared the results obtained by blocks of stimuli attended more to the stimuli that were difficult to recognize and objects, whereas LOC showed strong contrast invariance for the object stimuli and full invariance for the face stimuli. These results demonstrate that the contrast invariance in high-order visual areas is not a result of hemodynamic signal saturation, rather, it reflects a true characteristic feature of neuronal activity in high-order object areas.

Could attention effects account for the contrast invariance?

It could be argued that attention effects might contribute to the contrast invariance found in the LOC. Thus if subjects attended more to the stimuli that were difficult to recognize and if attention produces enhanced activation in the LOC (Wojciklik et al. 1998), this might lead to “flattening” of the contrast response because the lowered activation due to reduced contrast will be compensated by the increase in activation due to attention. Our control experiments, in which the subject’s task required attending the faces at different contrasts, or alternatively, attending the fixation point that had an unrelated contrast level, clearly rule out this possibility. Thus despite the fact that subjects did not attend the face stimuli, their contrast invariance level remained the same as in the case where they were required to recognize the face images or to remember their shape (see Fig. 8).

Comparison of the contrast response function in other animal and human studies

Our findings of contrast invariance in higher visual areas are compatible with previous single-unit studies in primates. Thus Rolls and Baylis (1986) reported that responses of neurons in the superior temporal sulcus (STS) were relatively invariant to contrast changes of face stimuli. The contrast response function was also characterized physiologically for extrastriate visual areas such as area MT (Cheng et al. 1994; Sclar et al. 1990) and V4 (Cheng et al. 1994). Using sinusoidal luminance gratings, Reynolds et al. (2000) showed that the neuronal response in area V4 increased with log contrast. The contrast response function obtained in area V4 in our experiment (see Fig. 2B) is comparable with these physiological findings.

Form perception is considered to be a faculty mediated by the ventral stream, which was thought to receive its major input from the parvocellular pathway (Livingstone and Hubel 1988). The magno- and parvocellular pathways have markedly different contrast response functions with the magnocellular system showing higher sensitivity and early contrast saturation (Merigan and Maunsell 1990; Merigan et al. 1991). To test this view, Ferrera et al. (1992, 1994) studied the responses of neurons in area V4 after inactivating the magno- or parvocellular layers within the LGN. They found no evidence for a clear dominance of one of the two pathways in this area. Neither was there a clear spatial segregation of the two inputs within V4. Thus it is plausible that visual areas within the inferotemporal cortex, which receive major ascending inputs from area V4, (Nakamura et al. 1993) would also have mixed magnocellular and parvocellular contributions. However, one should not conclude from the contrast invariance observed in LOC in our study that it is due to magnocellular input, it may well be that this effect is produced intrinsically at the level of the LOC itself.

Several neuroimaging studies characterized the contrast response function of human visual areas. Studying attentional effects, Kastner et al. (2000) found monotonic increase in the contrast response function in areas V1, V2/Vp, V4, V3A, and MT. These findings are consistent with our findings for ventral retinotopic visual areas (i.e., V1, V2, Vp, and V4; see Fig. 2B). Tootell et al. (1995) studied area MT and V1 and showed that similar to the physiological findings obtained in monkeys, human area MT exhibits high sensitivity to contrast and its activity saturates at low contrast levels. The fMRI activation in area V1, on the other hand, increased as a function of log contrast without obvious saturation.

Neuronal mechanisms responsible for contrast invariance

While the sensitivity to contrast changes observed in area V1 and even in the retina are fairly well understood physiologically (Kaplan and Shapley 1986; Ohzawa et al. 1982; Sclar et al. 1990). The mechanisms responsible for contrast invariance observed in higher order areas, such as the present results and those reported by Rolls and Baylis (1986) for cells in the monkey’s STS, are still not clear.

A simple mechanism that could produce such invariance is a high sensitivity to low contrast combined with saturation nonlinearity in the neuronal response (i.e., a ceiling effect). Enhanced contrast sensitivity in higher-order visual areas may be a consequence of the large receptive field size, characteristic...
of neurons in these areas (Sciar et al. 1990). This simply follows from the assumption that spatial summation of inputs will increase sensitivity in successive visual areas. The gradual increase in receptive-field size in the ventral stream (Amir et al. 1993; Tootell et al. 1997; Van Essen 1985), reaching its highest level in LOC with a bilateral-visual field activation pattern (Grill-Spector et al. 1998b), may therefore be the reason for the contrast invariance observed in LOC.

An alternative mechanism that could produce such invariance is a nonspecific contrast gain control operating on a fast time scale. Such mechanism will tend to shift the dynamic range of the contrast response function so that it will optimally register small changes from the adapting contrast (Muller et al. 1999; Ohzawa et al. 1982). Contrast gain control effects predict a higher degree of invariance for blocks of images compared with single presentations, and this was not found in our single-event experiment. However, a more direct comparison of these conditions (block design vs. single event) should be conducted to properly explore this possibility.

A simple sensitivity of LOC to low contrast like that observed in MT/MST is unlikely because the level of the contrast invariance we observed was not identical for all stimulus types as expected from such sensitivity. While full invariance was observed for the face stimuli, it was weaker for the object stimuli and it was not significantly different from V1 for the pattern stimuli. (see Fig. 4).

We thank M. Behrmann, U. Hasson, and I. Levy for fruitful discussions and comments. We thank E. Oncon for technical assistance.

This study was funded by Israel Academy Grant 8009/00-1 and German-Israeli Foundation Grant I-0576-040.01/98.

REFERENCES


MENDOLA JD, DALE AM, FISCHL B, LIU AK, AND TOOTELL RB. The representation of illusory and real contours in human cortical visual areas revealed.


**ROLLS ET AND BAYLIS GC.** Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. *Exp Brain Res* 65: 38–48, 1986.


2.2
Analysis of the neuronal selectivity underlying low fMRI signals

Galia Avidan, Uri Hasson, Talma Hendler, Ehud Zohary, and Rafael Malach

Current Biology, June 25 2002, 12:964-972
Results

Analysis of the Neuronal Selectivity Underlying Low fMRI Signals

Galia Avidan,1,2,6 Uri Hasson,3 Talma Hendler,4,5
Ehud Zohary,1,2 and Rafael Malach1,3,6
1The Interdisciplinary Center for Neural Computation
2Department of Neurobiology
Hebrew University of Jerusalem
Jerusalem 91904
3Department of Neurobiology
Weizmann Institute of Science
Rehovot 76100
4Imaging Department
Whol Institute for Advanced Imaging
Sourasky Medical Center
Tel Aviv 64239
5Faculty of Medicine
Tel Aviv University
Tel Aviv 69978
Israel

Summary

Background: A prevailing assumption in neuroimaging studies is that relatively low fMRI signals are due to weak neuronal activation, and, therefore, they are commonly ignored. However, lower fMRI signals may also result from intense activation by highly selective, albeit small, subsets of neurons in the imaged voxel. We report on an approach that could form a basis for resolving this ambiguity imposed by the low (mm range) spatial resolution of fMRI. Our approach employs fMR-adaptation as an indicator for highly active neuronal populations even when the measured fMRI signal is low.

Results: In this study, we first showed that fMRI-adaptation is diminished when overall neuronal activity is lowered substantially by reducing image contrast. We then applied the same adaptation paradigm, but this time we lowered the fMRI signal by changing object shape. While the overall fMRI signal in category-related regions such as the face-related pFs was drastically reduced for non-face stimuli, the adaptation level obtained for these stimuli remained high. We hypothesize that the relatively greater adaptation level following exposure to “nonoptimal” object shapes is indicative of small subsets of neurons responding vigorously to these “nonoptimal” objects even when the overall fMRI activity is low.

Conclusions: Our results show that fMR-adaptation can be used to differentiate between neuronal activation patterns that appear similar in the overall fMRI signal. The results suggest that it may be possible to employ fMR-adaptation to reveal functionally heterogeneous islands of activity, which are too small to image using conventional imaging methods.

Introduction

An implicit and widely held assumption in functional magnetic resonance imaging (fMRI) research is that the strong magnetic resonance (MR) signals within a cortical area reflect the functional specialization of this region, so that weak activation is treated as “suboptimal” and is considered to be functionally irrelevant. However, a weak fMRI signal may not be a simple reflection of global, suboptimal, neuronal activation, but it may also result from a variety of heterogeneous neuronal activity patterns. An extreme example of such a case is a small set of intensely active neurons whose signal is averaged out by the vast majority of nonresponsive neurons within which they are embedded. An illustration of such a situation is depicted in Figure 1. As illustrated on the left, weak activation to a certain set of images (houses and words in this example) can reflect the activity of a uniform population of face-selective neurons, which are suboptimally activated by these object categories. However, as illustrated on the right, the same weak fMRI activation could reflect the presence of a few intensely active groups of neurons, which are not sensitive to faces at all but rather are preferentially activated by other object categories (e.g., houses or words). These neurons are embedded among a large number of face-selective neurons, and due to their relative low density, their activity is not sufficient to produce a strong fMRI signal in the measured voxel. Note that conventional MR imaging cannot differentiate between these two extreme alternatives. Here, we propose to use the fMR-adaptation phenomenon, in which repeated presentation of a stimulus leads to significant reduction in the fMRI signal, to differentiate between activity patterns that appear to be similar in conventional fMRI imaging (for a review, see [1]).

What are the factors that control the adaptation process? Previous studies have demonstrated that the level of adaptation was dependent on the similarity between the repeated stimuli and in fact could be used to probe the level of invariance and generalization of neuronal populations in the visual cortex [2, 3]. However, another parameter that was not considered so far in human fMRI studies is the activation level of the neuronal population; that is, it could be that the level of adaptation correlates to the level of activation produced by the adapting stimulus prior to the adaptation process. This issue is of crucial importance because if indeed the fMR-adaptation level is linked to neuronal activation, then the level of adaptation could be used as a measure for the level of activity of cortical neurons, thus bypassing the spatial resolution limits of the fMRI method.

We chose to study the relationship between the neuronal activation and adaptation level by creating suboptimal neuronal activation through drastic reduction of image contrast. Single-unit studies clearly indicate that lowering image contrast reduces neuronal activation and that this effect is not specific to a particular group
Results

Neuronal Selectivity and Low fMRI Signals

Figure 1. Alternative Neuronal Organizations within a Face-Selective Voxel
Such a voxel may contain a homogenous population of neurons (left) optimally activated by faces and suboptimally activated by non-face stimuli. Alternatively (right), it may contain a heterogeneous mixture of neurons with small subsets of neurons ("hot spots") optimally tuned to non-face object categories. Both alternatives would produce an identical fMRI signal in a conventional experimental design.

of neurons [4, 5]. We have shown previously [6] that drastic contrast reduction leads to weak fMRI signals in both early and high-order visual areas. Here, we show that, in high-order, object-related areas, adaptation was significantly reduced when fMRI activation was reduced by lowering image contrast, thus demonstrating that adaptation was indeed correlated to the level of neuronal activation.

We then applied this method as an analysis tool to reveal that object shape changes produce markedly different neuronal activation patterns in category-specific cortex (e.g., face- and house-selective regions) compared to contrast changes, although their conventional fMRI signals are comparable. Similar to contrast reduction, shape changes in category-specific cortex lead to a drastically reduced signal for the nonselective stimuli (e.g., houses and words in the fusiform face area [FFA]). Crucially, unlike contrast changes, repetition of the same "nonoptimal" stimulus manifested a robust adaptation effect. Thus, adaptation effects can expose differential effects of contrast and shape changes. Such differences were pointed previously in single unit studies of mid-level visual areas [7]. We hypothesize that these high adaptation levels are due to the presence of small neuronal populations, which are strongly activated by the nonoptimal object stimuli.

Results

Dependency of Adaptation on Neuronal Activation
To explore the dependency of adaptation on neuronal activation, we conducted the experiment shown in Figure 2A. In this experiment, fMRI activation was measured for face images (see the Experimental Procedures, Experiment 1) and for non-face images (cars, Experiment 2; not shown) at different contrast levels. In both experiments, we measured the adaptation level by comparing the fMRI activation to "identical" versus "different" image presentations.

As reported previously using similar images [2], adaptation for high-contrast images was particularly evident at high-order object areas: a lateral-occipital focus (LO), which is situated ventrally and posteriorly to area MT and extends into the posterior inferotemporal sulcus, and a focus in the vicinity of the posterior fusiform gyrus (pFs), which is anterior and lateral to area V4/V8 and extends into the inferior temporal sulcus (see Table 1 for Talairach coordinates). The latter focus (pFs) may overlap the fusiform face area described previously [8]. Another focus that exhibited substantial adaptation was situated within the anterior portion of the collateral sulcus (CoS), and this focus may overlap the parahippocampal place area (PPA) reported previously [9] (see
Table 1 for Talairach coordinates). This is demonstrated in Figure 2B, which shows a functional map of the adaptation effect for high-contrast car images averaged across nine subjects.

Figure 3 shows the activation profiles in these anatomical regions. Note that, at low contrast, the averaged fMRI signal was reduced by a factor of 3 and the adaptation was concurrently reduced (Figure 3A: different faces, LO: high contrast: 1.28 ± 0.17% signal change, low contrast: 0.43 ± 0.12; pFs: 1.46 ± 0.24 versus 0.54 ± 0.14). The reduction of adaptation at low contrast was highly significant and did not depend on the object category used (faces or cars). (Three-way ANOVA: significant main effect of stimulus repetition: LO: F = 37.8, p < 10^{-4}; pFs: F = 20.5, p < 10^{-4} and of stimulus contrast: LO: F = 25.9, p < 10^{-4}; pFs: F = 20.2, p < 10^{-4}, and significant interaction of these effects: LO: F = 13.1, p < 10^{-4}; pFs: F = 6.8, p < 0.002. No significant effect of stimulus type, i.e., faces/cars: LO: F = 0.6, p < 0.45; pFs: F = 0.9, p < 0.34).

Similar results were obtained in the collateral sulcus for the car images used in Experiment 2 (Figure 2B: car images, two-way ANOVA: significant main effect of stimulus repetition: F = 45.5, p < 10^{-4} and of stimulus contrast: F = 45.5, p < 10^{-4}, and significant interaction of these effects: F = 19.4, p < 10^{-4}). This link between adaptation and fMRI activation suggests that the level of adaptation can now be used as an indirect measure of neuronal activity levels within the measured voxels.

Table 1. Talairach Coordinates

<table>
<thead>
<tr>
<th>Left</th>
<th>Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>X Y Z</td>
<td>X Y Z</td>
</tr>
<tr>
<td>Face Contrast Adaptation Experiment (Experiment 1)</td>
<td></td>
</tr>
<tr>
<td>LO: F = 25.9, p &lt; 10^{-4}; pFs: F = 20.2, p &lt; 10^{-4}, and significant interaction of these effects: LO: F = 13.1, p &lt; 10^{-4}; pFs: F = 6.8, p &lt; 0.002. No significant effect of stimulus type, i.e., faces/cars: LO: F = 0.6, p &lt; 0.45; pFs: F = 0.9, p &lt; 0.34.</td>
<td></td>
</tr>
<tr>
<td>LO</td>
<td>pFs</td>
</tr>
<tr>
<td>-42 ± 5</td>
<td>-69 ± 8</td>
</tr>
<tr>
<td>pFs</td>
<td></td>
</tr>
<tr>
<td>-36 ± 4</td>
<td>-48 ± 10</td>
</tr>
<tr>
<td>Car Contrast Adaptation Experiment (Experiment 2)</td>
<td></td>
</tr>
<tr>
<td>LO: F = 45.5, p &lt; 10^{-4} and of stimulus contrast: F = 45.5, p &lt; 10^{-4}, and significant interaction of these effects: F = 19.4, p &lt; 10^{-4}).</td>
<td></td>
</tr>
<tr>
<td>LO</td>
<td>pFs</td>
</tr>
<tr>
<td>-44 ± 2</td>
<td>-66 ± 4</td>
</tr>
<tr>
<td>pFs</td>
<td></td>
</tr>
<tr>
<td>-34 ± 4</td>
<td>-53 ± 11</td>
</tr>
<tr>
<td>CoS</td>
<td></td>
</tr>
<tr>
<td>-27 ± 3</td>
<td>-44 ± 6</td>
</tr>
</tbody>
</table>

Talairach coordinates [27] for LO and pFs, derived from all subjects in Experiments 1 and 2, and for the collateral sulcus (CoS), obtained from all subjects in Experiment 2. Values represent the mean ± SD in mm.
Category-Adaptation Experiment: The Issue of Cortical Modularity

Having found that adaptation level could provide an indirect indication of the level of neuronal activation, we could now apply this method to study the activity patterns in ventral-stream, high-order areas. A fundamental issue, which appears particularly amenable for analysis using this approach, is the modular nature of high-order object representations, i.e., to what extent cortical regions are purely specialized for a single category of object images. Perhaps the best-studied example of such specialization is the fusiform face area [8, 10, 11] and the parahippocampal place area [9, 12]. The first was described as a cortical module specialized in processing faces, and the latter was associated with processing houses and scene images. The main feature of the FFA, which led to its characterization as an independent module, was the low level of its activation by non-face objects. The same characteristic was found for the PPA but for non-house or place stimuli. Is this weak activation a result of suboptimal neuronal activation throughout the FFA and PPA, similar to that produced by low-contrast images? Or is it the result of strong neuronal activation produced by a small set of highly selective neurons within those regions? To answer this question, we performed the category-adaptation experiment (Experiment 3), following the same experimental design as in the contrast-adaptation experiments but using three different object categories (faces, houses, and words). Figure 4A shows the activation level in face-selective voxels both in LO and in the pFs (corresponding to the FFA) for faces and the non-face images of houses and words. Note that, while the overall activation was drastically reduced for the non-face stimuli, the adaptation level remained robust and did not differ significantly for all stimulus types in both LO and pFs (two-way ANOVA: significant main effect of stimulus repetition: LO: F = 4.5, p < 10^{-2}; pFs: F = 8.2, p < 10^{-3} and of stimulus type: LO: F = 33.1, p < 10^{-6}; pFs: F = 21.8, p < 10^{-5}, but no significant interaction between these factors: LO: F = 0.7, p < 0.5; pFs: F = 0.4, p < 0.7 and between hemispheres and stimulus repetition as calculated by a three-way ANOVA: pFs: F = 1.5, p < 0.2). Similar results were obtained in the collateral sulcus (corresponding to the PPA, Figure 4B), which was associated previously with specific house-related or place-related activity [9, 12]. This region was strongly activated by the house images, and its activity was drastically reduced for the face and word images. (Two-way ANOVA: significant main effect of stimulus repetition: F = 8.0, p < 10^{-4} and of stimulus type: F = 20.0, p < 10^{-4}, but no significant interaction between these factors: F = 0.6, p < 0.6).

The similar level of adaptation for "optimal" and non-optimal stimuli in these regions exhibits a marked discrepancy from contrast-related effects. Furthermore, it
Results

Current Biology
38

Figure 4. Category Adaptation
(A) Activation profiles in face-related areas during the category-adaptation experiment: LO (top), pFs (bottom). Internal localizer data (see the Experimental Procedures) is shown in the left graphs, and the measured activation is shown in the right graphs. Note the significant adaptation effect evident for all stimulus categories, although the overall signal was clearly face-selective in these regions. Dark and light colors indicate different and identical conditions, respectively. Green, blue, and red indicate house, word, and face stimuli, respectively. Asterisks denote the significance level as in Figure 3, but it is calculated for each stimulus category. Error bars indicate \( \pm \) SEM. Abbreviations are as in Figure 2; H., houses; W., words; F., faces.

(B) Activation profiles in the house-related area (CoS); again, note the significant adaptation for all stimulus categories despite the drastic reduction in activation for the face and word images. Conventions are as in Figure 4A.

strongly suggests that, despite the weak overall signals, there was robust functional activation by the nonoptimal stimuli in all these regions.

**Direct Comparison between Contrast and Category Adaptation**

In the two sets of experiments described above (Experiments 1 and 2 and Experiment 3), the fMRI-activation level was manipulated in two different ways: image contrast changes in one case, and shape changes in the other. While adaptation was reduced when lowering the image contrast, it remained robust for both optimal and nonoptimal object categories despite the drastic reduction in the overall activation level for the nonoptimal categories. However, it could be argued that the lack of adaptation in the low-contrast conditions in the contrast-adaptation experiments is a result of a floor effect, i.e., a nonlinear effect that precluded the fMRI signal from being reduced further below the low-contrast nonadapted state. Furthermore, the results were obtained in two separate experiments, thus enabling only indirect comparisons.

In order to address these issues, we conducted another experiment, which combined the critical conditions from both the contrast-adaptation and the category-adaptation experiments. In this experiment, we measured the adaptation level in face-related regions for two stimulus types: face images presented in medium-contrast level (nonoptimal category presented in an optimal fashion), Face-related regions were independently localized by including localizer conditions (high-contrast faces and houses).

Behavioral data were collected during the experiment. Since adaptation levels may depend on the perceived similarity between presented images, we quantified the ability of subjects to discriminate between the different images during the different conditions. To that end, we calculated a response alteration measure for each subject (see the Experimental Procedures). Statistical analysis revealed that the response alteration measure differed significantly between the different and identical image presentations (stimulus repetition effect). Importantly, this measure was not significantly different for the two contrast levels that were used in the experiment (i.e., high-contrast faces compared to medium-contrast faces). Furthermore, no significant interaction was found between these factors. Two-way ANOVA: a significant main effect of stimulus repetition: \( F = 116, p < 10^{-3} \), but not of stimulus contrast: \( F = 0.3, p < 0.6 \), and no significant interaction between these factors: \( F = 10^{-4}, p < 1.0 \). Thus, any differences between the high-contrast houses and the medium-contrast faces in terms of the adaptation level could not be ascribed to the subjects’ inability to discriminate between the different stimuli in these two conditions.

Figure 5A shows the activation profile obtained in face-related voxels defined by the independent localizer (faces > houses) in LO and the pFs for the medium-
Results

Neuronal Selectivity and Low fMRI Signals

Figure 5. Direct Comparison of Contrast and Category Adaptation

(A) Activation profiles of face-related regions in LO (top) and pFs (bottom) obtained for the medium-contrast face stimuli and for the high-contrast house stimuli used in Experiment 4. Gray and white bars indicate activation to faces and houses, respectively. Filled and dotted bars indicate activation to different and identical conditions, respectively. Note that, while significant adaptation was obtained for both stimulus categories, the adaptation level was stronger for the house stimuli than for the face stimuli. This occurred despite the higher signal for the face compared to the house stimuli. Asterisks denote the significance level as calculated between “diff.” and “ident.” presentation of each condition (paired t test, an asterisk indicates $p < 0.01$). The error bars indicate $\pm$ SEM.

(B) The graphs show the adaptation index for the medium-contrast faces and the high-contrast houses obtained in LO and pFs. An index of zero indicates no adaptation, while larger values indicate stronger adaptation. Note that the adaptation index for the house stimuli was significantly larger (stronger adaptation) than the adaptation index obtained for the medium-contrast face stimuli, further stressing that the overall level of fMRI activation was not the limiting factor in determining the level of adaptation. Asterisks denote the significance level as calculated between the adaptation index for houses and faces (paired $t$ test, a double asterisk indicates $p < 0.01$). The error bars indicate $\pm$ SEM.

Contrast faces and the high-contrast houses. The adaptation index (Figure 5B) for the medium-contrast faces was significantly weaker (paired $t$ test, $p < 0.01$) than the adaptation index for the high-contrast houses despite the fact that the fMRI-signal level for the face stimuli was actually higher. This trend is in agreement with the results of the previous experiments.

These results demonstrate that strong adaptation can be obtained in face-related regions (LO, pFs) for nonoptimal images yielding a low fMRI signal (high-contrast houses), while much weaker adaptation is obtained for stimuli eliciting a stronger fMRI signal due to contrast reduction. This indicates that the level of adaptation is not a consequence of the level of the overall fMRI signal, but rather is determined by the underlying neuronal activity.

Discussion

Problems Inherent in the Spatial Resolution of fMRI

While substantial attention is given to the relatively low temporal resolution of the BOLD signal [13–15], major difficulties are also inherent in the limited spatial resolution of fMRI and indeed in any functional neuroimaging method that records the averaged activity of neuronal populations. We have previously discussed this difficulty with respect to the assessment of the invariant properties of cortical neurons (for a review, see [1]). Here, we address the issue of putative small populations of functionally distinct neurons, which are at a subvoxel resolution. Our results show that, although the existence of such populations can not be revealed by direct measurement of the BOLD signal, it may be possible to use adaptation effects to highlight their existence.

Relationship between Adaptation and Cortical Activity

Our suggestion that the existence of fMRI-adaptation can be taken as a measure of high neuronal activity level is based on the results of the contrast-adaptation experiments (Experiments 1 and 2), which showed that the adaptation level was significantly correlated to the activation level. Thus, substantial reduction in fMRI activation practically abolished the adaptation (low-contrast condition, Figure 2). We attribute this reduced adaptation to an overall reduction in the neuronal activation.

Of the single neuron studies of adaptation in the inferior-temporal (IT) cortex, the only report that is somewhat relevant is that of Li et al. [16], which indeed found an association between optimal activation and familiarity effects in a small subset of IT neurons. However, the adaptation procedure was substantially different than the one used here, so it is difficult to make direct comparisons.

The Issue of Cortical Modularity

An influential concept of cortical organization, delineated most forcefully by Fodor [17], has been the exis-
tence of highly specialized and relatively isolated cortical modules for specific cognitive tasks. As mentioned above, the most documented example of such a putative module has been the fusiform face area [8, 10, 11, 18], but other areas of cortical specialization have been suggested as well [9, 12]. However, it is not clear to what extent these regions are purely specialized for one stimulus category and to what extent other object representations may coexist within them. The fact that we observed similar adaptation levels to optimal and non-optimal stimuli illustrates that the strength of fMRI signals might not reveal the full functional profile of these regions. The most straightforward interpretation of our results is that, within the category-selective object areas, there are functionally distinct “hot spots” of neuronal activation that are strongly activated by the nonoptimal stimuli.

Thus, our results point to the possibility that, although a clear functional specialization within ventral stream areas does exist [8], it is not absolute, and representations of various object categories coexist within these regions. Such mixing might occur on a columnar level, so that shape-selective columns [19] of various object categories interdigitate within the same cortical region. The source of such a heterogeneous mosaic is not clear, but it might be related to complex acuity demands [20, 21], differential expertise effects [22], or distributed object representation [23, 24].

However, lacking direct evidence for such hot spots, it should be emphasized that alternative, albeit more complex, schemes might account for the present results. For example, it could be the case that the low fMRI activation to house images found in the face-related FFA is due to a uniformly distributed, weak input from the neighboring, house-related PPA. If this is the case, then the adaptation found for house images in the FFA is a reflection of the neuronal adaptation that occurs in the PPA and is transmitted to the FFA through lateral cortico-cortical connections. Since we found adaptation effects throughout the object-selective cortex, such a model implies extremely exuberant intracortical connections. However, until more direct studies using techniques such as single-unit recordings or optical imaging of adaptation effects in primate IT are performed, this issue can not be fully resolved.

Conclusions

In conclusion, we propose a method based on fMRI-adaptation that allows better differentiation of neuronal activation patterns, which appear ambiguous using conventional fMRI. Using this approach, we demonstrated a qualitative difference in the impact of contrast and shape changes in high-order human object areas. Finally, we hypothesize that our results point to the presence of highly active neuronal islands that are selective for nonoptimal stimuli in category-selective human object areas.

Experimental Procedures

Subjects

Fourteen healthy subjects (eight women, ages 24–50) participated in one or more of the experiments. All subjects had normal or corrected-to-normal vision and provided written informed consent. The Tel Aviv Sourasky Medical Center approved the experimental protocol.

MRI Setup

Subjects were scanned in a 1.5-T Signa Horizon LX 8.25 GE scanner equipped with a standard birdcage head coil. Blood oxygenation level-dependent (BOLD) contrast was obtained with gradient echo-planar imaging (EPI) sequence (TR = 3000, TE = 55, flip angle = 90°, field of view 24 × 24 cm², matrix size 80 × 80). The scanned volume included 17 nearly axial slices of 4-mm thickness and 1-mm gap. Three subjects who participated in Experiment 1, one subject who participated in Experiment 3, and the seven subjects who participated in Experiment 4 were run with a quadrature surface coil (Nova Medical), which covered the posterior brain regions. T1-weighted, high-resolution (1 × 1 × 1 mm) anatomical images and 3D spoiled gradient echo sequences were acquired on each subject to allow accurate cortical segmentation, reconstruction, and volume-based statistical analysis.

Visual Stimulation

Stimuli were generated on a PC, were projected via an LCD projector (Epson MP 7200) onto a tangent screen positioned over the subject’s forehead, and were viewed through a tilted mirror located above the subjects’ eyes.

Experiments

Experiment 1: Face Contrast Adaptation

Experiment 1 (Figure 2A, 8 subjects), lasted 480 s, included 6 different stimulus conditions, and had 51 epochs that were presented in a counterbalanced block design paradigm: stimulus and blank epochs were interleaved. A stimulus epoch consisted of either 12 different stimuli (different condition) or 12 repetitions of the same stimulus (identical condition) all presented in the same contrast level. Stimuli were presented in 4.4%, 8.1%, or 95.6% contrast level (low, medium, and high contrast, respectively, in the following text). Stimuli were 12° × 12° black on white line drawings of faces. The illumination level of the white background was 97 cd/m², and that of the black line drawings at 95.6% contrast was 2 cd/m², as measured directly from the tangent screen. Stimulus contrast was defined as follows, where L is luminance:

\[
\text{Stimulus Contrast} = \frac{L_{\text{max}} - L_{\text{min}}}{L_{\text{max}} + L_{\text{min}}}
\]

Contrast levels were verified by direct measurement from the tangent screen. The contrast of stimuli was varied by changing the gray level of the line drawing while keeping the background (white) constant. Each experimental condition was repeated 4 times; in addition, there were 26 interleaving blank epochs, and the experiment started with an epoch of pattern stimuli, which was excluded from all the statistical analysis. Each stimulus epoch lasted 12 s, and each blank epoch lasted 6 s, with the exception of the first and last blanks, which lasted 21 s and 15 s, respectively.

Within an epoch, each of the 12 images was presented for 800 ms followed by 200 ms of fixation point on a blank screen. Subjects were instructed to fixate on the fixation point located in the middle of each stimulus and to covertly categorize each stimulus as a man, woman, or child. To enable subjects to differentiate between low-contrast epochs and blank epochs, stimulus epochs had a gray fixation point, while blank epochs had a red fixation point.

Experiment 2: Car Contrast Adaptation

This experiment was identical in its design to Experiment 1, but the stimuli that were used were car images. Nine subjects were run on the experiment. Subjects had to covertly categorize each stimulus as a private car, a bus, or a truck.

Experiment 3: Category Adaptation

This experiment was identical in its design to Experiment 1, but instead of using three different contrast levels, we used three different stimulus categories: faces, houses, and words all presented in a high-contrast level. Eight subjects were run on this experiment, and their task was to covertly name each stimulus.
Experiment 4: Category and Contrast Adaptation

Seven subjects were run in this experiment, which was similar in its design to Experiment 1. The experiment included 4 conditions, which consisted of either 12 different stimuli (different condition) or 12 repetitions of the same stimulus (identical condition). Two conditions included medium-contrast faces, and two conditions included high-contrast houses. In order to independently localize face-related regions, we included two additional localizer conditions of high-contrast faces and houses. Each stimulus was repeated 4 times. Within an epoch, each of the 12 images was presented for 800 ms, followed by 200 ms of a blank screen. Subjects were instructed to fixate on the fixation point located in the middle of each stimulus and to perform a two-alternative forced choice recognition task. Each face had to be categorized as a man or a woman, and each house image had to be categorized as a public or a private facility. Subjects provided their responses via a “Neuroscan Co.” response box, and data were collected by in-house software.

Due to a technical failure, behavioral data were collected for five out of the seven subjects who participated in the experiment.

Data Analysis

Details of the analysis are provided in [6]. Briefly, fMRI data were analyzed with the BrainVoyager software package (Brain Innovation). The cortical surface was reconstructed and unfolded into the flattened format. Statistical analysis was based on the General Linear Model [25]. Statistical maps (Figure 2) were colored according to the relative contribution of selected predictors [6]. The percent signal change (PSC) for each subject in each experiment was calculated as the percent activation from a blank baseline:

\[
\frac{\text{Percent Signal} = \frac{\text{signal} - \text{mean signal (blanks)}}{\text{mean signal (blanks)}} \times 100}{1}
\]

“Internal Localizer” Test

Details of this analysis are given in [26]. Briefly, we used one set of epochs to define anatomical regions of interest, while another set was used to estimate the PSC, providing an independent measure of activation.

Regions of interest in Experiment 4 were defined using the localizer conditions, applying a statistical test that searched for all face-related voxels (faces > houses) while ignoring all the experimental conditions. The adaptation index was defined as: 1 - (activation [PSC] in the identical condition/activation in the different condition) for each stimulus type.

Analysis of Behavioral Data from Experiment 4

We calculated a response alteration measure as follows: a binary value ("0" or "1") was assigned to each of the two possible key presses in each epoch; the standard deviation across these values was calculated separately for each epoch and was then averaged across all the repetitions of the same experimental condition for each subject. A two-way ANOVA was then calculated across all subjects in order to evaluate whether the response alteration measure differed significantly across the four experimental conditions.

Acknowledgments

We thank Marlene Behrmann and Ifat Levy for fruitful discussions and comments. We thank Michal Harel for 3D brain reconstruction and Eli Okon for technical assistance. This work was funded by Israel Science Foundation 8009/00-1 and German-Israeli Foundation I-0576-040.01/98 grants.

Received: September 27, 2001
Revised: March 14, 2002
Accepted: April 17, 2002
Published: June 25, 2002

References


2.3
Spatial vs. object specific attention in high-order visual areas.

Galia Avidan\textsuperscript{1,2,3}, Ifat Levy\textsuperscript{1,3}, Talma Hendler\textsuperscript{4,5}, Ehud Zohary\textsuperscript{1,2}, and Rafael Malach\textsuperscript{3}

Affiliation
\textsuperscript{1}The Interdisciplinary Center for Neural Computation, Hebrew University of Jerusalem, Jerusalem 91904, Israel.
\textsuperscript{2}Department of Neurobiology, Hebrew University of Jerusalem, Jerusalem 91904, Israel.
\textsuperscript{3}Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel.
\textsuperscript{4}Imaging Department, Whol Institute for Advanced Imaging, Sourasky Medical Center, Tel Aviv 64239, Israel.
\textsuperscript{5}Faculty of Medicine, Tel Aviv University, Tel Aviv 69978 Israel.

Submitted
Results

**Abstract**

Recently we reported that the topographic organization of visual field eccentricity in human visual cortex extends into high order, ventral occipito-temporal (VOT) cortex. Within this region faces and buildings have specific eccentricity biases, suggesting that this category-eccentricity association may reflect differential resolution needs of recognition processes. However, it is still not clear to what extent this center/periphery differentiation within high-order occipito-temporal cortex depends on immediate, moment-to-moment, task demands and to what extent it is established through long-term effects. Here, we examined the role of moment-to-moment attentional demands in establishing this association and also characterized the nature of the attentional mechanisms found in these regions by comparing the relative modulatory impact of spatial attention and attention to object shape.

We found that when fMRI activation by images of buildings and faces was contrasted, the differential activation to each object category in occipito-temporal object areas was maintained, even when attention was shifted to different shapes and spatial locations. However, under these conditions object-related activation was enhanced in a shape-specific manner. When spatial attention was modulated, by attending to large or small arrow-heads, regardless of object shape, an attention-related eccentricity map was revealed in the more dorsal lateral occipital (LO) object area. However this map was only slightly extended into the more ventral occipito-temporal cortex (VOT). These results argue against the possibility that the source for the differentiation of ventral occipito-temporal cortex is a purely spatial attention mechanism and support the notion that the attentional modulation involves a combined effect of both shape selectivity and eccentricity bias.

**Introduction**

Our understanding of the organizing principles of the human occipito-temporal visual cortex has been growing rapidly in recent years. A number of functional mapping studies have revealed a complex network of specialized regions showing clear selectivity in their response properties for various object
shapes. A particularly striking example is the medio-lateral differentiation along ventral occipito-temporal cortex in activation to face (Puce et al., 1995; Clark et al., 1996; Kanwisher et al., 1997; Halgren et al., 1999) and building images (Aguirre et al., 1998; Epstein and Kanwisher, 1998).

In parallel with this differentiation, we have recently found evidence for a topographic organization in which, for identical object shapes, medial occipito-temporal cortex, along the collateral sulcus, is activated preferentially to more peripheral stimuli compared to lateral regions along the fusiform gyrus (Levy et al., 2001; Malach et al., 2002). Furthermore, we have shown that face-related activation is associated with central visual field bias, while building-related activation is associated with a peripheral bias. Similar association with central field bias was found for letter strings and words (Hasson et al., 2002). This eccentricity-bias map was termed the ventral occipito-temporal cortex (VOT) (Malach et al., 2002). We hypothesized that the category-eccentricity association may reflect differential needs of recognition processes for high cortical magnification (associated with central vision) or large-scale integration (associated with peripheral vision).

One aspect of this finding, which remains unresolved, is the extent to which the center/periphery differentiation depends on immediate, moment-to-moment task demands and the extent to which it is established through long-term developmental or learning effects. A prime instance of moment-to-moment task demands are attentional shifts. According to space-based theory of attention (Posner, 1980; Treisman and Gelade, 1980), attention can be directed to certain locations in space and hence can be metaphorically referred to as a ‘spotlight’ having a variable size. Space-based attentional effects were indeed found in retinotopic visual areas (Tootell et al., 1998; Brefozynski and DeYoe, 1999; Martinez et al., 1999; Somers et al., 1999) but also in high-order object related ones (Downing et al., 2001).

Following these findings, the center-periphery organization in the ventral occipito-temporal cortex (VOT) might be explained as a by-product of varying spatial attentional demands. For example, whenever one encounters building images, which may require large-scale feature integration, a large spotlight of attention may be activated leading to the peripherally-biased collateral sulcus activation. While in
Results

encountering face images or letters, which require high acuity demands, a foveal spotlight of attention is recruited, leading to the centrally biased activation.

Alternatively, it could be that object identity, rather than spatial attention is the dominant factor in determining the activation within those regions. Such a finding would be more compatible with an object-based attentional mechanism (Duncan, 1984; Vecera and Farah, 1994; Desimone and Duncan, 1995; O’Craven et al., 1999), according to which attention is directed toward specific objects rather than specific visual field locations.

In the present work we tried to examine the nature of the attentional mechanisms found in high-order occipito-temporal cortex by comparing the modulatory impact of spatial attention and attention to object identity. We found that when images of buildings and faces were contrasted while attention was shifted to different shapes and spatial locations, the medial-lateral differentiation of buildings and faces remained, although the level of activation was modulated by these attentional shifts. When spatial attention was modulated by attending to a central arrow or to a larger arrow in the periphery, an attention-related eccentricity map was found in the dorsally located lateral occipital region, that was only slightly extended into VOT cortex.

These results argue against shape-independent spatial attention as the underlying cause for the center-periphery differentiation of object areas in ventral occipito-temporal cortex. However, they are compatible with the notion that attentional effects in high order object areas depend on both spatial and object components.
Results

Materials and Methods

Subjects

Eight healthy subjects (4 women, ages 24-50), participated in the experiments. All subjects had normal or corrected to normal vision and provided written informed consent. The Tel-Aviv Sourasky Medical Center approved the experimental protocol.

MRI Setup

Subjects were scanned in a 1.5 Signa Horizon LX 8.25 GE scanner equipped with a quadrature surface coil (Nova Medical Inc., Wakefield, Massachusetts), that covered the posterior brain regions. Blood oxygenation level dependent (BOLD) contrast was obtained with gradient-echo echo-planar imaging (EPI) sequence (TR = 3000, TE = 55, flip angle = 90°, field of view 24 × 24 cm², matrix size 80 × 80). The scanned volume included 17 nearly axial slices of 4mm thickness and 1mm gap. T1-weighted high resolution (1×1×1mm) anatomical images and 3D spoiled gradient echo sequence were acquired on each subject to allow accurate cortical segmentation, reconstruction and volume-based statistical analysis.

Visual Stimulation

Stimuli were generated on a PC, projected via LCD projector (Epson MP 7200) onto a tangent screen positioned over the subject's forehead, and viewed through a tilted mirror located above subjects' eyes.

Experiments

Attention experiment

The experiment lasted 450 sec, it included 10 different stimulus conditions, and had 57 epochs, which were presented in a counterbalanced block design paradigm, stimulus and blank epochs were interleaved. A stimulus epoch consisted of 9 different stimuli; to minimize eye-movements each stimulus
Results

was presented for 150 msec followed by 850 msec blank. Stimuli were 12X12 deg. grayscale photos of faces and buildings. Superimposed on each image were three red dots that formed together a subjective arrow-head shape. The arrow-head was either small (0.96X1.64 deg.) located at the center of the image or large (6.2X12 deg.) and located in the periphery of the image (Figure 1). The size of the dots composing the large arrow-head were enlarged to compensate for the magnification factor. The three dots composing the arrow-head were used instead of a full arrow in order to ensure that judging the arrows’ direction could not be resolved by local information but would be based on global integration. This was especially crucial during the ‘large arrow’ condition.

Different epochs contained either small or large arrows. In each image the arrow was pointing randomly to one of the four cardinal directions with a small jitter. Additionally, the arrow-heads (either small or large) appeared alone in two conditions, which were repeated twice each. A small (0.3X0.3 deg.) fixation-cross appeared in the middle of all images throughout the experiment.

Each stimulus epoch lasted 9 sec. and each blank (fixation cross only) epoch lasted 6 sec. with the exception of the first and last blanks, which lasted 18 sec. each, blank epochs were repeated 29 times.

Subjects were instructed to fixate on the fixation cross, located in the middle of each stimulus, and to attend either the object stimuli or the arrows. Instructions were conveyed by an image of a letter, which preceded each stimulus epoch and appeared for 2 sec, the letters were ‘F’, ‘H’, or ‘A’. ‘F’ indicated that the epoch would contain faces and that the subjects should attend the faces and covertly categorize each of them as female or male. ‘H’ indicated that the epoch would contain buildings (houses) and that subjects should categorize each of them as a private or a public facility, in the following text we will refer to either of these conditions as “attend pictures”. In contrast, the letter ‘A’ indicated that subjects had to attend the superimposed arrow and to indicate the direction to which it pointed. Each of these conditions was repeated three times. Task performance was verified for each subject during a training session performed outside the scanner on a different set of images. For five subjects performance of the original experiment was measured under similar experimental conditions, but outside the scanner, at least two months after the original scan, so that priming effects would be avoided. Subjects provided their responses via a
keyboard and responses were collected by in-house software. For each subject both mean reaction time and the proportion of correct responses were calculated for each experimental condition (Table 1).

MT experiment
The experiment lasted 420 sec, visual stimuli were low contrast white rings surrounding the fixation point, 0.5 cycle/degree. Two different 18 sec experimental conditions were used: a stationary condition, in which visual stimuli were presented every 3 seconds and a moving condition in which the rings expanded for 2 sec and contracted for 2 sec at a rate of 7°/sec. Subjects’ were instructed to fixate on the fixation point throughout the experiment. For each subject area MT/V5 was defined by a correlation test contrasting the moving and stationary conditions.

Data Analysis
fMRI data were analyzed with the “BrainVoyager” software package (Brain Innovation, Masstricht, Netherlands) and with complementary in-house software. The cortical surface of each subject was reconstructed from the 3D-spoiled gradient echo scan. The procedure included segmentation of the white matter using a grow-region function, the smooth covering of a sphere around the segmented region, and the expansion of the reconstructed white matter into the gray matter. The sulci were smoothed using a cortical “inflation” procedure. The surface was cut along the Calcarine sulcus and unfolded into the flattened format. The obtained activation maps were superimposed on the unfolded cortex and the Talairach coordinates (Talairach and Tournoux, 1988) were determined for the center of each ROI (region of interest). Preprocessing of functional scans included 3D-motion correction and filtering out of low frequencies up to 5 cycles per experiment (slow drift).

Statistical analysis was based on the General Linear Model (GLM), (Friston et al., 1995). The GLM analysis was performed independently for the time course of each individual voxel for each subject. Each experimental condition (except for blank) was defined as a separate predictor. The first three images of
Results

Each functional scan were discarded, a box-car shape was used for each predictor and a hemodynamic lag of 3 sec was assumed.

Percent signal change (PSC) for each subject was calculated as the percent activation from a blank baseline:

\[
\text{Percent Signal} = \frac{\text{signal} - \text{mean}[\text{signal (blanks)}]}{\text{mean}[\text{signal (blanks)}]} \times 100
\]

Multi-subject analysis: In addition to subject-by-subject analysis (Figure 4) we also analyzed the data in a multi-subject approach (Figures 2, 3 and 5). To obtain the multi-subject maps, time series of images of brain volumes were converted into Talairach space, z-normalized and concatenated. The statistical tests were performed on the concatenated time course. In the relevant figures, the multi-subject functional maps are projected on the flattened Talairach normalized brain of one subject.

For creating the two-color maps presented in Figures 2, 3 and 5, the GLM model was recalculated using a subset of the predictors. The highlighted voxels in each figure are those for which this reduced model explained a substantial portion of the variance. When mapping the relative contribution of two functional responses the color-coding represents the relative contribution of either set. In all figures brightest green represents exclusive contribution of one predictor set over the other while brightest red represents exclusive contribution of the second predictor set. The exact color used between these two extremes depends on the level of differential contributions by each predictor set.
Results

The aim of the experiment was to characterize attentional modulation effects within face and building related regions. During the experiment, subjects viewed an identical set of visual stimuli that consisted of complex natural objects (either faces or buildings) with superimposed 3-dot sets, forming either a small or large subjective arrow-head (Figure 1). Subjects were instructed to attend to different aspects of the stimuli during different epochs. In one case, the subjects attended the pictures of either the faces or buildings (in the text we will refer to this condition as "attend pictures") and in the other they attended the arrows (small or large) (see Materials and Methods for more details).

Behavioral data for this task are given in Table 1. Performance as measured by both reaction time and proportion of correct responses was high during all experimental conditions. Note specifically the high performance during all the ‘attend arrow’ conditions, which implies that subjects’ attention was indeed engaged by that task.

We first looked for an overall attention effect by contrasting activation to the same stimuli under different selective attention tasks (attend pictures vs. attend arrows). The results of this analysis are shown in Figure 2. As can be seen, the different tasks resulted in a clear differentiation of dorsal from ventral stream areas. Thus, attending the pictures generally activated occipito-temporal regions (blue-green), while attending the arrows activated occipito-parietal regions (yellow-red). The strongest signal enhancement by attending the pictures (green) was evident in the ventral occipito-temporal cortex (VOT), which includes the posterior fusiform region (pFs) and the collateral sulcus (CoS). Interestingly, the lateral occipital region (LO) did not exhibit a substantial modulation by either attentional state as evident in the more yellow and blue colors assigned to it.

The strongest signal enhancement when attending the arrows (red) was found in the vicinity of the intraparietal sulcus (IPS), post-central sulcus (PCS) and in the motion sensitive area MT/V5. The preferential activation in the parietal regions obtained by attending to the rotating arrows can be attributed to the task
Results

performed with these stimuli which involved spatial information (naming the direction to which the arrows pointed) and hence could contribute to the activation in these areas, associated with the dorsal processing stream (Ungerleider and Mishkin, 1982). The preferential activation for arrows, obtained in MT/V5, may be related to the flickering rate created by the arrows (Tootell et al., 1995), in addition, subjects also reported perceiving an apparent motion effect with these stimuli.

Since the current experiment did not include experimental conditions optimally designed for activating area MT/V5, for 6 out of 8 subjects, the localization of this region was verified by a comparison to the localization of MT/V5 obtained in a separate scan using low contrast contracting and expanding rings vs. stationary rings, which are known to be very effective stimuli for this area (Tootell et al., 1995) (see Materials and Methods). The location of MT/V5 as marked in Figure 2 is an estimated averaged location based on this experiment.

To what extent was the differential activation to buildings and faces dependent on specific attention? To examine this issue we directly compared the activation maps obtained from epochs in which attention was directed to the building or face images, with maps obtained from epochs during which subjects attended the arrows. Figure 3 shows the results of this comparison. Importantly, these two sets of epochs were physically identical, and differed only in terms of the task performed by the subjects. Figure 3a was obtained by comparing the activation to face images (blue-green) with the activation to buildings (yellow-red) when subjects attended the pictures of objects. Figure 3b was obtained by applying the same test to the epochs in which subjects attended the arrows.

Generally, the differential activation to buildings and faces in occipito-temporal regions was maintained during both attentional states. However, some differences were evident, as can be seen also from the enlarged insets of these regions. Note that the activations in VOT, when subjects attended the pictures (Figure 3a) was enhanced and extended toward more anterior regions, as indicated by the white arrows placed on the right hemisphere in Figure 3a which show the anterior edge of activation in the ‘attend
arrows’ conditions. Thus, our results show that while shape-related selectivity in LO and VOT was not dependent on attention to specific shapes, it could be substantially intensified by it.

Although the focus of this study was attentional modulation in ventral stream areas, interesting differences between the two attentional states were also found in parietal regions. While regions in the vicinity of the intra-parietal sulcus exhibited clear preference for building stimuli over face stimuli in the ‘attend pictures’ condition (red, Figure 3a), this selectivity was substantially reduced during the ‘attend arrows’ condition as evident in the more yellow and blue colors assigned to this region in Figure 3b. This could result from the fact that during the latter condition subjects attended the arrows (large or small) regardless of the underlying stimuli (faces or buildings).

Another difference between the two attentional states was that when subjects attended the arrows (Figure 3b) activity was extended also to the vicinity of the post-central sulcus. This is indicated by the white arrow placed on the right hemisphere in Figure 3b which shows the anterior edge of activation in the ‘attend pictures’ conditions. These findings are consistent with the strong selectivity found in those regions for the arrow stimuli (Figure 2).

To obtain a quantitative measure of the activity level during the different experimental conditions we performed time course analysis for each subject in four different regions of interest (ROIs), which were selected according to their preferential activation by a specific stimulus category regardless of the attentional state (see Table 2 for Talairach coordinates of each region). Note that this statistical test is different than the one used for Figure 3. In order to relate to attentional modulation in each of these ROIs, we compared the activity during epochs that were identical in terms of their physical content but differed in the task performed by the subjects (i.e. attending the pictures vs. attending the arrows). Statistical significance was verified by a paired t-test calculated between the ‘attend pictures’ and ‘attend arrows’ conditions, for each arrow size in each ROI.
Results

Figure 4a shows the activation profile of a region in the collateral sulcus, which was defined by its preferential activation to buildings over faces regardless of the attentional demands (all buildings>all faces). This region exhibited significant signal enhancement when attending the building images compared to the arrows, for both the ‘large’ and ‘small arrows’ conditions. In addition, this region also exhibited signal enhancement when attending the face images which was significant only for the ‘small arrow’ condition, and a significantly greater fMRI signal (PSC) for the large arrows compared to the small arrows when presented alone (paired t-test).

Figure 4b shows the activation profile of regions within the pFs and LO, which were defined by their preferential activation to faces compared to buildings (all faces>all buildings) and separated according to anatomical criteria. The former focus was situated in the vicinity of the posterior fusiform gyrus (pFs), which is anterior and lateral to area V4/V8 and extends into the inferior temporal sulcus and the latter (LO) was situated ventrally and posteriorly to area MT/V5 and extended into the posterior infero-temporal sulcus. Note that despite the similarity in the face selectivity in both pFs and the face-related LO, two differences were evident in the activation profile in these two regions. First, pFs exhibited substantial signal enhancement when attending the face stimuli, while LO exhibited a weaker level of such modulation, which was evident only for the ‘large arrow’ condition. Second, LO exhibited stronger activation than the pFs for the ‘arrows alone’ condition both for the large and small arrows (two-way ANOVA: significant main effect of ROI: F=5.2 p<10\(^{-2}\); but no effect of arrow size: F=0.4 p<0.6 and no significant interaction between these factors: F=0.03 p<0.9).

Figure 4c shows the activation profile of area MT/V5 that was defined by its preferential activation to the rotating arrows when they were presented alone (all arrows alone>blank). Interestingly, this area exhibited an opposite attentional effect to that found in object-related cortex (CoS, pFs, LO). That is, the activity was stronger when subjects attended the arrows compared to when they attended the pictures. The enhanced MT/V5 activation provides indirect evidence that subjects indeed attended the arrows during the ‘attend arrows’ conditions.
Finally, we examined to what extent manipulating purely spatial (‘spotlight’) attention while maintaining object shape can differentially affect VOT and LO activation (Figure 5). This was done by comparing the activation during all the conditions in which subjects attended the large arrows (blue-green) vs. the conditions in which they attended the small arrows (yellow-red), the conditions in which the arrows appeared alone were not included in the statistical analysis. The white dotted line, marked on both hemispheres, is an estimated border separating object-related regions from lower visual areas. This border was obtained by contrasting objects vs. texture patterns in a separate localizer experiment that was run during a different scanning session. Details of this experiment can be found in (Levy et al., 2001). This was done for 7 out of 8 subjects.

Note that the early retinotopic areas (e.g. left of the dotted white line on the right hemisphere) showed a clear attention-related eccentricity map, that is center-biased regions were activated more strongly when attention was directed to the small foveal arrows (red), while areas which have a peripheral bias were preferentially activated by attention to the large peripheral arrows (green). Importantly, this ‘attentional eccentricity’ map, particularly in the peripheral regions, extended outside the retinotopic areas mainly into LO but only slightly into VOT. Face-related pFs did not show a specific attentional eccentricity bias, indicating that its activation was not substantially modulated by spatial shifts in attention when attention was directed to the arrow-shapes.

Discussion

Interaction between eccentricity mapping and attentional modulation

Recently we have found evidence for topography organization according to eccentricity bias that extends beyond the classical retinotopic areas into high-order ventral stream visual areas (Levy et al., 2001; Malach et al., 2002). In addition, we have shown a clear association between category specific activation and eccentricity. The present results are relevant to the question of how the eccentricity biases in occipito-temporal cortex emerge. One can envision several different alternatives: it could be that the
association between category-related activation and eccentricity was created due to slow ecological needs imposed during evolution, and therefore this association is now innate or hardwired. Alternatively it could be that such needs are imposed during development, so that this association is being acquired rather than innate. Compatible with this notion are findings showing that representations within high-order object-related areas can be modified following learning (Gauthier et al., 1999) and also findings showing that similar to faces, written letters and words activate areas, which are associated with central representation (Hasson et al., 2002).

Finally it could be that the eccentricity-category association is not a result of any long-term processes but rather is a result of real-time spatial attentional shifts. According to this alternative, whenever one encounters a stimulus, which requires detailed scrutiny (e.g. a face or a word), a small central ‘spotlight of attention’ is used and hence visual areas, which have a central bias, are immediately recruited. In contrast, whenever stimuli that require large-scale feature integration are presented (e.g. buildings), one uses a large, peripheral ‘spotlight of attention’ and this recruits areas, which are associated with peripheral field representation. Spotlight attention effects were indeed discovered in retinotopic visual areas (Tootell et al., 1998; Brefozynski and DeYoe, 1999; Martinez et al., 1999; Somers et al., 1999) and also in high order object-related areas (Downing et al., 2001). Moreover, increased activity was found in retinotopic visual areas when subjects directed their attention to particular location in a visual scene even in the absence of a visual stimulus (Kastner et al., 1999)

Thus, if the source of the eccentricity-category association was due to shape-independent spatial attention, then in the current experiment we would expect to find that whenever subjects attended the small arrows, central-biased regions would be recruited regardless of the identity of the underlying stimulus (i.e. a face or a building) while peripheral regions would be activated when attending the large peripheral arrow. However the results show that shape selectivity maps within ventral occipito-temporal cortex (pFs and the CoS) as well as within LO were largely maintained regardless of the attentional task (i.e. attention to pictures in Figure 3a compared to attention to arrows in Figure 3b) or the spatial focus (Figure 5).
This general preservation of the stimulus selectivity maps under different selective attentional states show that shape-invariant space-based attentional demands could not account for the category-topography association found in high-order object-related areas.

On the other hand, shifting attention from the arrows to the face/building pictures enhanced the activation in the VOT and extended this activation further anteriorly (compare Figures 3a and b). This modulatory effect provides further support for the existence of object-based attentional mechanism (Duncan, 1984; Vecera and Farah, 1994; Desimone and Duncan, 1995) within those areas. Such object-based attentional modulations were shown by O’craven et al. (O’Craven et al., 1999) in the fusiform face area, FFA (Kanwisher et al., 1997), and in the parahippocampal place area, PPA (Epstein and Kanwisher, 1998), when stimuli (faces and buildings) occupied the same location in space.

The fact that VOT activation extended further anteriorly during selective attention to pictures of objects is nicely compatible with the notion of a posterior-anterior hierarchical axis of object representation (Lerner et al., 2001; Avidan et al., 2002). Thus, it appears that the anterior part of the VOT further departs from the physical retinal stimulation and is more tightly coupled to the perceptual or attentional state of the observer compared to its posterior part.

**A combined eccentricity/object attentional effect**

In the present study we found that while the eccentricity-related attentional effects extended into high-order object related cortex, the VOT itself was only slightly affected by such spatial attention effects. This result can be explained by the fact that VOT was not sensitive to the arrow shapes (see CoS and pFs, Figure 4).

Taken together, these results are compatible with an organization principle of occipito-temporal object-related cortex, which can be described as a combination of two selectivity dimensions: sensitivity to object shape on the one hand and the eccentricity map on the other hand (Levy et al., 2001; Malach et al., 2002). Similarly, attentional modulation in these regions appears to require both shape selectivity and eccentricity.
Results

From this analysis we can hypothesize, that attending to a combination of object shape and eccentricity- i.e. attending to face images in the visual field center vs. its periphery, should reveal stronger eccentricity-bias attentional effects in the ventral occipito-temporal cortex.

**Differential attention effects in the dorsal and ventral visual pathways**

The neural substrates mediating object and spatial processing of both macaque and human visual systems appear to be segregated into two anatomically distinct and functionally specialized streams: the ventral and dorsal pathways respectively (Ungerleider and Mishkin, 1982). In recent years several neuroimaging studies were able to demonstrate this segregation in the human visual cortex by directing subjects' attention to different aspects of identical visual stimuli (Haxby et al., 1991; Haxby et al., 1994; Kohler et al., 1995). Our results also demonstrate a clear differentiation of dorsal from ventral stream areas, as a consequence of different selective attentional tasks (Figure 2). The preferential activation for the arrows within dorsal regions, which are primarily engaged in processing spatial information and motion, can be attributed to the task performed with these stimuli, which involved spatial information.

Interestingly, while attentional modulation in object-related areas was manifested in stronger activation when subjects attended the pictures compared to when they attended the arrows, area MT/V5 exhibited a reversed pattern of modulation. That is, enhanced activation when attending the arrows compared to attending the pictures, for both the face and the building stimuli (Figure 4c). These findings are consistent with previous imaging findings showing attentional modulation in this area (Beauchamp et al., 1997; O’Craven et al., 1997; Rees et al., 1997; Tootell et al., 1998).

The differential attention effects between the dorsal and ventral visual pathways demonstrate nicely the concept of correlation between stimulus selectivity and attentional modulation. Such correlations were described by Corbetta et al. who showed, using PET, that attention to different features in an image such as its color, shape, or velocity activated segregated visual areas, each specializing in processing the corresponding visual feature (Corbetta et al., 1990).
Accordingly, ventral stream areas, which have strong stimulus selectivity for objects, were strongly modulated by attending the object pictures, while dorsal stream areas, which exhibited stronger stimulus selectivity for the arrows were preferentially activated by attending to them. Such correlation can also account for the in-homogeneity of the shape related attentional modulations found within the object-related areas. While LO exhibited only mild shape related attentional modulation, VOT exhibited more robust effects. This can be attributed to the stronger activation by the arrow shape found in LO compared to the VOT.

Correlation between stimulus selectivity and attentional modulation will naturally emerge if attentional enhancement was targeted to neurons according to their shape-selectivity. Thus, cortical areas in which there is a marked difference between the neural activity generated by complex objects compared to the arrow in either way (e.g. VOT compared to MT/V5) would also show strong attentional modulation when shifting attention between the arrow and the pictures. In contrast, regions having a more balanced activation to the two types of stimuli (LO) will show weaker modulation effects.
References


Results

Figure 1

Figure 1: Experimental design of attention experiment
Stimuli were presented in short epochs and included faces and buildings. A subjective arrow-head (either small or large), which was composed of 3 dots, was superimposed on each image. Dots were enlarged here for demonstration purpose and were red in the original experiment. Subjects were instructed to covertly name the arrows’ direction in one set of epochs (attend arrows, left) and to perform a categorization task in the other set (attend pictures, right) while maintaining fixation throughout the experiment. Identical stimuli were presented in the two sets of epochs. The arrows also appeared alone during a separate condition.
Figure 2: Attention to pictures versus attention to arrows

Averaged activation map of the general attention effect (8 subjects, GLM statistics) presented on two unfolded hemispheres of one subject. The map was created by contrasting the ‘attend arrows’ (red to yellow colors) and ‘attend pictures’ (green to blue colors) conditions. Importantly, identical visual stimuli were used in both conditions. Note that while more dorsal areas tended to be colored in yellow and red more ventral regions tended to be colored in green and blue, thus indicating the different attentional modulation within these regions. Abbreviations: LO, lateral occipital; pFs, posterior fusiform; CoS, collateral sulcus; VOT, ventral occipito-temporal. The statistical criterion for the highlighted voxels was $p<10^{-6}$ uncorrected.
Figure 3
Results

Figure 4

(a) all buildings > all faces

(b) all faces > all buildings

(c) arrows alone > blank

CoS

pFs

LO

MT/VS

% Signal Change

Attend Pictures (faces)
Attend Arrows
Attend Pictures (build.)
Attend Arrows

L.A., S.A.
Figure 3: Preservation of stimulus selectivity maps under different attentional states.
a. Stimulus selectivity maps of the ‘attend pictures’ conditions: Averaged activation map of the stimulus selectivity (8 subjects, GLM statistics) presented on two unfolded hemispheres of one subject. The map was created by contrasting the faces (green to blue colors) and buildings (red to yellow colors) while subjects were attending these stimuli. Note the strong face selectivity within face-related LO and the pFs and the strong building selectivity within the CoS. The inset shows an enlarged map of the occipito-temporal regions.
b. Stimulus selectivity maps of the ‘attend arrows’ conditions: Same maps as in Figure 3a but created by contrasting the faces and buildings stimuli during the ‘attend arrows’ condition. Importantly, the statistical criterion for the activated voxels of both panels of this figure was identical (p<10^-5 uncorrected). Note that the stimulus selectivity in occipito-temporal regions was largely maintained across the two different attentional states (Figure 3a vs. Figure 3b). However activity during the ‘attend pictures’ condition was extended into more anterior regions, as indicated by the white arrows placed on the right hemisphere in Figure 3a. Some differences between the two selective attentional states also exist in dorsal regions. Selective activation for buildings in the vicinity of the intraparietal sulcus was substantially reduced during the ‘attend arrows’ condition, and in addition activity was extended also to the vicinity of the post-central sulcus. This is indicated by the white arrow placed on the right hemisphere in Figure 3b, showing the anterior edge of activation in the ‘attend pictures’ conditions.

Figure 4: Activation profiles in the different ROIs of the Attention experiment
a. Activation profile of the collateral sulcus obtained by searching for building selective voxels (all buildings>all faces). Red, green and gray indicates the stimulus category: faces, buildings and arrows-alone respectively. Dark and light colors indicate the task performed by the subjects: ‘attend pictures’ or ‘attend arrows’ respectively. Attentional modulation is manifested in stronger signal during the ‘attend pictures’ compared to the ‘attend arrows’ condition. This region exhibited attentional modulation for the building stimuli for both arrow size and also substantial modulation for the face stimuli in the ‘small arrow’ condition. Asterisks denote significance level as calculated between the ‘attend pictures’ compared to the ‘attend arrows’ in each of the picture conditions, or between the large and small arrow conditions when they were presented alone (paired t-test, *p<0.05, **p<0.005, ***p<0.0005), error bars indicate ± standard error of the mean (SEM). Abbreviations: L.A., large arrow, S.A., small arrow.
b. Activation profiles of areas pFs (top) and LO (bottom) obtained by searching for face selective voxels (all faces>all buildings). Note that pFs exhibited significant attentional modulation for the face stimuli for both arrow sizes while LO exhibited this effect only for the ‘large arrow’ condition. Conventions as in Figure 4a.
c. Activation profile of area MT/V5 (arrows alone>blank). Area MT/V5 exhibited a reversed attentional modulation compared to the object selective regions, i.e. its activity was stronger when subjects attended the arrows compared to when they attended the pictures of either faces or buildings. Conventions as in Figure 4a.
Figure 5: Attention-related eccentricity mapping

Averaged attention-related eccentricity map (8 subjects, GLM statistics) presented on both hemispheres of one subject. The map was obtained by contrasting the conditions in which subjects attended the small arrows (red to yellow) with the conditions in which they attended the large arrows (green to blue) across the face and building stimuli. A clear attention-related eccentricity map was obtained within lower, retinotopic visual area (e.g. left of the white dotted line in the right hemisphere). This map also extended to the more dorsal lateral occipital (LO) object area but was only slightly extended into VOT (right of the white dotted line in the right hemisphere). Statistical criterion for highlighted voxels: $p<10^{-5}$ uncorrected.
Table 1: Behavioral data of Attention experiment

Reaction time and proportion of correct responses for each experimental condition of the Attention experiment averaged across 5 subjects as measured outside the scanner. Values represent the mean±std.

<table>
<thead>
<tr>
<th>Experimental Condition</th>
<th>Reaction time (msec)</th>
<th>Proportion of correct responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faces Large Arrow Attend Faces</td>
<td>615±57</td>
<td>0.90±0.06</td>
</tr>
<tr>
<td>Faces Large Arrow Attend Arrows</td>
<td>667±49</td>
<td>0.98±0.03</td>
</tr>
<tr>
<td>Faces Small Arrow Attend Faces</td>
<td>611±64</td>
<td>0.96±0.03</td>
</tr>
<tr>
<td>Faces Small Arrow Attend Arrows</td>
<td>641±70</td>
<td>0.95±0.03</td>
</tr>
<tr>
<td>Houses Large Arrow Attend Houses</td>
<td>680±47</td>
<td>0.82±0.11</td>
</tr>
<tr>
<td>Houses Large Arrow Attend Arrows</td>
<td>660±52</td>
<td>0.99±0.02</td>
</tr>
<tr>
<td>Houses Small Arrow Attend Houses</td>
<td>642±71</td>
<td>0.81±0.09</td>
</tr>
<tr>
<td>Houses Small Arrow Attend Arrows</td>
<td>619±88</td>
<td>0.99±0.02</td>
</tr>
<tr>
<td>Large Arrows Alone</td>
<td>583±72</td>
<td>0.99±0.03</td>
</tr>
<tr>
<td>Small Arrows Alone</td>
<td>527±48</td>
<td>1.00±0.00</td>
</tr>
</tbody>
</table>

Table 2: Talairach coordinates

Talairach coordinates (Talairach and Tournoux, 1988) for LO, pFs, CoS and MT/V5 as derived from all subjects. Values represent the mean±std in mm.

<table>
<thead>
<tr>
<th></th>
<th>Left</th>
<th>Right</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>LO</td>
<td>-47±4</td>
<td>-69±6</td>
</tr>
<tr>
<td>pFs</td>
<td>-38±2</td>
<td>-48±9</td>
</tr>
<tr>
<td>CoS</td>
<td>-25±2</td>
<td>-46±8</td>
</tr>
<tr>
<td>MT/V5</td>
<td>-46±4</td>
<td>-67±4</td>
</tr>
</tbody>
</table>
Chapter 3
General Discussion

Inspired by the notion that object recognition and perception largely depend on the interplay between feedforward, bottom-up information on the one hand, and high-level, top-down information on the other, this thesis attempted to assess the relative contribution of each of those information types to object-related activation in the human brain. To that end I first delineated high-order object related areas, which are more related to the actual perception of a stimulus, from lower areas, which are more tuned to its physical content. I then explored the principles of object representation, within these high order areas, by applying a novel method in an attempt to bypass the spatial limits of fMRI, and study object-related areas at the neuronal population level. I also examined the role of top-down attentional modulation in constructing these representations. Finally, I had a unique opportunity to investigate the relationship between brain activation and conscious perception directly, in an individual who suffers from a specific face recognition deficit (prosopagnosia).

3.1 Hierarchical organization of the visual cortex

The way object recognition processes are accomplished by the visual system is still far from being fully understood. Yet, it is generally accepted that this sequence of processing utilizes an anatomical-physiological hierarchy of visual areas having multiple feedforward, feedback and cross-area pathways (Felleman and Van Essen, 1991). The fMRI method, unlike single unit recordings, provides the opportunity to study the responses of multiple cortical areas simultaneously during the course of one experiment. Thus this method provides a powerful tool for studying the functional and anatomical properties of hierarchical processing within the entire constellation of human visual cortex. This can be achieved by tracking the differential responses in each visual area to parametric manipulations of an identical set of stimuli along one specific axis.
Following this rational we used a single, well defined visual property, that of image contrast, to follow the transformation in image representation along the entire constellation of ventral stream, human visual areas. Contrast offers a distinctive dichotomy: while perceptually, object recognition is highly invariant to changes of this parameter beyond a minimal threshold level, retinal responses are drastically modulated by changes in the contrast level. Consequently, the contrast response function can be used as a tool to explore the extent to which activation in a given visual area is determined by the physical aspect of the stimulus (i.e. contrast) and the extent to which this response is related to subject’s perceptual performance. We found that the contrast response profile of visual areas changes along the cortical hierarchy (i.e. V1, V2, Vp, V4/V8 and finally LO and pFs), moving from strong contrast dependence in early visual areas, to contrast invariance of varying degree in high order object areas. Importantly, this transformation was gradual and did not involve abrupt transition along particular visual areas.

From a broader perspective of object recognition processes, this transformation towards contrast invariance is yet another example of a visual process enabling object constancy (Grill-Spector et al., 1999; Gross, 1972; Sary et al., 1993). We have also shown that the fMRI signal within high order object-related areas, unlike that of early visual areas, was substantially correlated to subjects’ explicit recognition performance.

These results nicely reflect the hierarchical trend in the human visual cortex, in which cortical responses gradually depart from the physical aspects of the visual stimulus and become correlated with perceptual experience (see also Lerner et al., 2001).

Interestingly, analysis of frontal cortical regions did not show a significantly enhanced invariance compared with that found in object-related areas, so it appears that the contrast invariance effect reached its highest degree already at the level of occipito-temporal object-related cortex. In addition, the contrast invariance manifested in object-related regions did not depend on specific attention to the object stimuli, and was also found when subjects performed a distracting, attention-demanding foveal task. Both these
findings suggest the possibility that contrast invariance may be accomplished in a feedforward manner, and that top-down information may not be a necessary component in constructing this effect. Accordingly, contrast invariance could be obtained due to high sensitivity to low contrast and a non-linear saturation effect in the neuronal response. Such enhanced contrast sensitivity in higher order visual areas might result from the large receptive field size, characteristic of neurons in these areas. This simply follows from the assumption that spatial summation of inputs will increase sensitivity in successive visual areas (Sclar et al., 1990). Thus, the gradual increase in receptive-field size along the ventral stream (Amir et al., 1993; Tootell et al., 1997; Van Essen, 1985), reaching its highest level in occipito-temporal regions (Grill-Spector et al., 1998) may therefore account for the substantial contrast invariance observed in these high-order regions.

It is interesting to discuss the results in the framework of a recent theory of visual perception termed, Reversed Hierarchy Theory (RHT), proposed by Hochstein and Ahissar (Ahissar and Hochstein, 1997) and S. Hochstein, personal communication. This theory suggests that early explicit perception which is generalized and more holistic in its nature, (vision at a glance), begins at the top of the classical visual hierarchy, but depends on implicit information processing along this hierarchy. In contrast, low level details (vision with scrutiny), are initially not accessible for conscious perception, reflecting the dissociation between early explicit perception and low-level vision. Such details become available once explicit mechanisms cascade in the reverse direction along the same hierarchy, thus, incorporating into conscious perception detailed information available at lower levels.

Accordingly, in our experiment subjects were required to perform a high level task of object recognition, an operation that can be accomplished, according to RHT, by high order object-related areas. However, had the task been more low level in its nature, e.g., contrast discrimination, it would have been accomplished by lower visual areas, following information flow in the reverse direction of the hierarchy. In that respect, it is worth mentioning that unlike the case in our study, in situations where both the task and
stimuli were tailored for optimally activating areas such as primary visual cortex, activity in this region, rather than in high order ones, was more correlated with performance (Boynton et al., 1999; Huk and Heeger, 2000; Ress et al., 2000). One straightforward prediction of RHT is related to the dynamics of activity expected in low-level cortical areas. It is predicted that the initial activity in these regions would be stimulus driven. However, once vision with scrutiny is required, then following the activation of feedback along the reverse hierarchy chain, activity patterns in V1 would change. Indeed, such findings were reported in single unit recording studies (Lamme and Roelfsema, 2000; Roelfsema et al., 1998). Predictions regarding the temporal dynamics of recognition processes can also be made from theoretical models of object recognition (Ullman, 1996), in which recognition is achieved by the combination of bottom-up and top-down information processing and the simultaneous exploration of multiple alternatives in each of these directions. Thus, for example simultaneous neuronal activity is expected to be found in both ascending and descending visual pathways before recognition is accomplished. However, lacking sufficient temporal resolution, the fMRI method is currently not optimized for addressing such issues, at least not in a direct way. It could be that studies combining fMRI, which offers good spatial resolution, with other methods such as MEG and ERP, which benefit from higher temporal resolution, are the key for understanding such dynamic properties of object processing in the human visual system.

### 3.2 Models of cortical organization for object representation

While in the first study, described above, I investigated the entire constellation of ventral stream hierarchy, taking advantage of the large coverage offered by the fMRI method, in the second one I focused on principles of object representation within the high order object-related areas, located at the top of this hierarchy. This issue cannot be studied directly using fMRI, due to the limited spatial resolution (millimeter range) of this method. Thus in order to circumvent this limitation I applied the fMR-adaptation approach which enables to tag neuronal activation at a sub-voxel resolution (Grill-Spector and Malach, 2001).
Crucially, I first showed that the fMR-adaptation level is linked to neuronal activation, and therefore can be used as a measure for the level of activity of individual cortical neurons, thus bypassing the spatial resolution limits of the fMRI method. This finding enabled me to address the issue of object representation within high order visual areas at the neuronal level.

One of the fundamental issues regarding visual object recognition and representation in humans is whether all kinds of objects are processed by the same neural mechanism, or whether instead some object classes are handled by distinct processing “modules”. This latter concept, of specialized and relatively isolated cortical modules for specific cognitive tasks has been put forward most forcefully by Fodor (Fodor, 1983). Fodor suggested domain specificity as a central criterion that must be met by a module, i.e. a module should be specialized in processing a particular domain of content (e.g. faces) and not others.

The strongest indication for a modular recognition system comes from the case of faces. Findings from neuropsychology (Damasio et al., 1990), cognitive psychology and ERP studies (Bentin et al., 1996; McCarthy et al., 1999), as well as brain imaging (Clark et al., 1996; Halgren et al., 1999; Kanwisher et al., 1997; Puce et al., 1995) suggest the existence of a face-related region in the human visual cortex located within the ventral processing stream. This region was suggested to be a face-processing module and was termed the fusiform face area (FFA) (Kanwisher et al., 1997). Face selective cells were also found in monkeys’ IT and STS (Baylis et al., 1987; Desimone et al., 1984; Gross et al., 1972; Perrett et al., 1982). Interestingly, however, other object categories such as building images (Aguirre et al., 1998; Epstein and Kanwisher, 1998), tools (Martin et al., 1996), animals (Martin et al., 1996), human body (Downing et al., 2001b), words and letter strings (Hasson et al., 2002b; Puce et al., 1996) and even chairs (Ishai et al., 1999), have been reported to generate specific and differential activation patterns within occipito-temporal regions, and even within the putative face-selective one (FFA).
The main feature of the FFA, which led to its characterization as an independent module, was the low level of its fMRI signal obtained for non-face object stimuli. However, a weak fMRI signal may not be a simple reflection of global, sub-optimal, neuronal activation, but may also result from a variety of heterogeneous neuronal activity patterns. I therefore studied the characteristics of these low fMRI signals, using the fMR-adaptation approach, and discovered that indeed they result from highly selective albeit small neuronal populations, which reside within category-selective object areas. These results suggest that the functional specialization within ventral stream areas is not absolute, and representations of various object categories coexist within these regions. Thus our findings undermine the existence of a modular organization in the human visual cortex, at least in a strict sense, since they show that alleged cortical modules such as the FFA are involved in processing other object categories and moreover they contain neurons that are optimized for these non-face categories. Supporting evidence for this view also comes from a study by Levy et al. (Levy et al., 2002), in which images of heads shown from front and back views consistently activated highly overlapping, yet slightly displaced regions in the fusiform gyrus, thus suggesting a graded topography rather than a modular one, in which representations of related objects slide smoothly along the fusiform gyrus.

What are the possible models of object representation that might account for these findings of a heterogeneous mosaic of object representations on the one hand with anatomically consistent peaks of selective activation for specific object categories on the other hand?

Recent studies of the macaque visual cortex (Fujita et al., 1992; Tanaka, 1996) have revealed a systematic columnar organization in the TE region, such that cells with relatively similar response properties are clustered in a columnar fashion perpendicular to the cortical surface, whereas cells in different columns respond to different visual features. Accordingly, one possible interpretation of the data from human object-related cortex is that the heterogeneous mosaic of object representations occurs at a columnar level, so that shape-selective columns of various object categories interdigitate within the same
cortical region. Consistently with this view, the peaks of category related activation could result from the fact that exemplars of the same category tend to share similar visual features, while images from different object categories tend to be visually different. Hence, if similar features are represented by neighboring neurons, then objects having similar shapes could activate the same object-related cortical regions. According to this view the heterogeneous mosaic of object representations (Avidan et al., 2002b) may be the manifestation of the fact that some visual features are shared by more than one object category.

In that respect it is important to note that object categories differ substantially not only in their visual appearance but also in their use (e.g. faces are used for social interaction, buildings are used for navigation). Thus it may be that object representations are clustered not according to their physical visual characteristics, but rather according to their most common usage. Another important parameter, which can explain some of the differences between various object categories are expertise effects which were indeed shown to modify activation at least in face-related cortex (Gauthier et al., 2000).

Another possibility is that cortical modules do exist within object-related regions although not in a strict sense as explained above. This scheme is somewhat problematic since obviously there is not enough cortex to support all putative object categories. Thus, a more reasonable suggestion is that certain categories such as faces, tools, letters and buildings are unique in having a specialized neuronal representation.

A more reduced model of modular organization suggests that only faces constitute a specialized object category, given their ecological and sociological significance (Farah et al., 1998; Kanwisher, 2000; Moscovitch and Moscovitch, 2000; Moscovitch et al., 1997). However, in either case one would still have to elucidate the principles of representation of those object categories, which are not represented in specialized modules and to account for the uniqueness of those that are.

While these models are interesting and are based on ample imaging findings as well as on physiological evidence, they seem to provide only a partial answer to the puzzle of object representation.
A completely different perspective is that objects are represented in a distributed fashion across the entire constellation of object related regions (Haxby et al., 2001; Ishai et al., 1999) and see also (Edelman et al., 1998). According to this view the representation of a face or an object is reflected by a distinct pattern of response across a wide expanse of cortex in which both large and small-amplitude responses carry information about object appearance. The main advantage of this model compared with those presented above, is that it predicts how all objects might evoke distinct patterns of response in object-related areas and, thereby, provides an explicit account for how this cortex can produce unique representations for a virtually unlimited number of categories. One shortcoming regarding the work of Haxby and colleagues (Haxby et al., 2001) and see also (Chao et al., 2002) is that despite the fact that they stress the importance of both large and small-amplitude fMRI responses and suggest that both signals carry information about object appearance, they do not make the critical distinction regarding the underlying neuronal populations, which produce these low fMRI signals (Avidan et al., 2002b) and see Chapter 2.2. Such a distinction is essential for understating object representation in the neuronal level. Nevertheless, the major caveat of their approach is that it does not account for the anatomically consistent peaks of category-related activation found within occipito-temporal object areas. In that respect, although powerful, a fully distributed representation in which the entire occipito-temporal cortex is involved in the representation of all object categories seems a too extreme approach. In contrast, my findings (Avidan et al., 2002b) support the existence of a distributed representation but in a more limited fashion, in which specific category-related activation does exist.

Finally, Malach et al. (Levy et al., 2001; Malach et al., 2002) have recently suggested an alternative scheme for object representation, which seems to account for the complexity of the results presented above in a more comprehensive manner. This scheme was based on their findings of topography in high order object areas, in which eccentricity bias, magnification factor and specific object shapes are linked in an orderly manner. Specifically, they showed that face-related regions are associated with a foveal bias, while building-related regions are associated with a peripheral visual field bias (see Figure 1). They
hypothesized that these category-eccentricity associations may reflect differential needs of recognition processes for high cortical magnification (associated with central vision) or large-scale integration (associated with peripheral vision). Consequently, each object category, based on its resolution needs, would be mapped onto this global, unifying organization scheme. However, this organization scheme is compatible with either a fully modular cortical organization on the one hand and a completely distributed representation on the other and therefore does not allow resolving these two opponent views.

It is important to note that the alternative schemes, for object representation within occipito-temporal cortex presented above, are not necessarily mutually exclusive. Moreover it is clear that the visual cortex is not a homogenous entity and thus different principles may dominate the representation in some visual areas but not in others. For example early visual areas are more dominated by a part-based representation while a more holistic representation exists in high order areas. Further study is required in order to join some or all of these organization principles into one coherent approach. Finally it is also important to note in that respect that the visual cortex constantly interacts with other brain regions such as prefrontal cortex where a more abstract, semantic-based representation may exist and this interaction could also dynamically affect object representation within the occipito temporal regions.

Figure 1: Relationship between face and building-related activation and visual field eccentricity. Preferential activation to faces (orange) and buildings (blue) shown on a ventral view of an inflated brain (a) and on an unfolded right hemisphere (b). c. Borders of the face and building-related activation from (b) are superimposed on an eccentricity map. Note the association between face-related activation and central visual field bias (yellow), and the association between building-related activation and peripheral visual field bias (green). Adopted from Malach et al., 2002
3.3 Attentional modulation of object recognition

How did the associations between category selectivity and eccentricity biases, described in the previous section, emerge?

One possibility is that these associations were constructed due to slow ecological needs, imposed during evolution, and therefore they are now innate or hardwired. A different hypothesis is that the stimulus selectivity and receptive field locations of visual neurons are learned rather than innate. Thus, the category-eccentricity associations result from a long-term exposure to the statistics of our everyday perceptual experience. After all, during daily life, we usually fixate our eyes on faces, while for perception of places or buildings we rely more heavily on peripheral vision. This possibility is compatible with findings showing that similar to faces, written letters and words also activate areas, which are associated with a central bias (Hasson et al., 2002b).

However, such a finding is also compatible with an alternative explanation: It could be that these associations are not formed due to slow, long-term processes, as suggested above, but rather they are the outcome of real-time spatial attention demands. So that whenever one encounters a stimulus, which requires detailed scrutiny (e.g. a face or a word), a small central ‘spotlight of attention’ is used and hence visual areas, which are centrally biased, are immediately recruited. In contrast, whenever stimuli that require large-scale feature integration are presented (e.g. buildings), one uses a large, peripheral ‘spotlight of attention’ and this recruits areas, which are associated with peripheral field representation. This possibility is consistent with spotlight-like attentional effects described not only in early retinotopic visual areas (Brefochnski and DeYoe, 1999; Martinez et al., 1999; Somers et al., 1999; Tootell et al., 1998) but also in high order object-related ones (Downing et al., 2001a).

I therefore examined the role of top-down, moment-to-moment attentional demands in constructing the category-eccentricity associations found within occipito-temporal object-related regions. I also characterized the nature of the attentional mechanisms found in these regions by comparing the relative
modulatory impact of spatial attention and attention to object shape (Avidan et al., 2002c), see Chapter 2.3. I found that shape-invariant space-based attentional demands could not account by themselves for the category-topography association found in high-order object-related areas. Still, spatial attention had some effect on the activation in these regions particularly in the more dorsal lateral occipital object area (LO). On the other hand I found that attention to object shape caused shape-specific modulatory effects, which were expressed in the spreading of activation into the more anterior parts of the ventral occipito-temporal (VOT) cortex. This latter finding is nicely compatible with the notion of a posterior-anterior hierarchical axis of object representation discussed above (section 3.1). Particularly, it appears that the anterior part of the VOT further departs from the physical retinal stimulation and is more tightly coupled to the perceptual or attentional state of the observer compared with its posterior part. These results are also consistent with recent findings (Epstein et al., 2003) showing that LO and PPA are not sensitive to the effects of spatial stimulation per se (i.e. center vs. periphery), but yet their responses can be modulated by tasks posing differential spatial attentional demands. Taken together, these results are compatible with an organization principle of occipito-temporal object-related cortex, which can be described as a combination of two dimensions: shape selectivity on the one hand and eccentricity mapping on the other (Levy et al., 2001; Malach et al., 2002). Likewise, top-down attentional modulation in these regions reflects an on-going interaction between these two dimensions.

### 3.4 Correlations between fMRI signal and object perception.

One of the main advantages of the fMRI method is that it enables to study directly the relationship or correlation, between brain activation in a specific area and explicit behavioral performance of human subjects. Such correlations have been taken as evidence for the contribution of that area to the perceptual state under investigation. Following that rational, many studies showed correlation between activation in occipito-temporal object-related areas and subjects’ performance, thus further stressing the role of these brain regions in object recognition.
While such correlations are interesting and telling, a detailed inspection of their characteristics is somewhat puzzling since both signal increase and decrease were found as a function of behavioral improvement (for a review of this issue, see section 1.2.3.1). Thus, I found that the fMRI signal in occipito-temporal areas was substantially correlated to the recognition performance of the subjects and was increased as performance was improved (See chapter 2.1). Similar findings were also reported under different behavioral manipulation (Bar et al., 2001; Grill-Spector et al., 2000; James et al., 2000; Kleinschmidt et al., 2002). However, it turns out that under different experimental paradigms, such as fMR-adaptation, this correlation does not hold (Grill-Spector and Malach, 2001), and following repeated stimulus presentation, which poses no difficulty to object recognition, a substantial signal decrease is obtained (Avidan et al., 2002b; Grill-Spector et al., 1999; Kourtzi and Kanwisher, 2000). Moreover, under classical priming manipulations, which, by definition, involve behavioral improvement, a reversed correlation is found, that is- a signal decrease as a function of behavioral improvement (Buckner et al., 1998; Chao et al., 2002; van Turennout et al., 2000).

This inconsistency between studies may be attributed, at least partially, to the kind of stimuli used in the two sets of experiments. Thus, in studies showing a positive correlation between fMRI signal and performance, the behavioral manipulation involves crossing the recognition threshold by either reducing stimulus contrast (Avidan et al., 2002a; Kleinschmidt et al., 2002), limiting stimulus’ exposure time (Bar et al., 2001; Grill-Spector et al., 2000), or by gradually unmasking the stimuli (James et al., 2000). In contrast, classical priming studies, showing a negative correlation, use suprathreshold stimuli that are easy to recognize, and behavioral improvement is then expressed in faster reaction time or naming. It could be that under subthreshold conditions, signal enhancement corresponds to the buildup of a new representation for the stimuli. In that case, neuronal populations would be increasingly recruited as the relevant parameter is manipulated across recognition threshold concomitantly with recognition performance, leading to the positive correlation between the two measures. Suprathreshold stimuli, on the other hand, already have a cortical representation, so that signal reduction could indicate sharpening and
tuning of this representation at the network level (Li et al., 1993; Miller et al., 1991). Importantly, however, such a mechanism was never demonstrated yet in single unit recording.

The nature of the complex and dynamic relationship between fMRI signal and recognition performance is far from being resolved. As presented above, one approach to tackle this issue is to measure brain activity while deliberately manipulating subjects’ conscious recognition and perception. I encountered a unique opportunity to examine the relationship between brain activation and conscious recognition from yet another perspective, that of specific behavioral object recognition deficit. This complementary approach is important since it enables not only to identify brain areas, which are correlated with recognition processes, but also to learn more about the necessity and sufficiency of those brain areas for object recognition.

We studied an individual (YT) who suffers from prosopagnosia, a specific and severe impairment in face identification. While this syndrome is usually acquired after a bilateral lesion in the vicinity of the fusiform gyrus, this person suffers from congenital prosopagnosia, without evident structural brain lesion. Using fMRI we tested the functional brain activity profile for face and non-face stimuli both in YT and in a control group of 12 subjects (Hasson et al., 2002a) and see Appendix of this thesis. We found that, despite being markedly impaired in face identification, YT had a grossly normal pattern of face-related activation. These findings suggest that YT’s profound prosopagnosia is not a result of decreased activation or lack of selectivity in occipito-temporal face-related regions.

This case study reveals an intriguing discrepancy between behavioral impairment in face identification on the one hand, and an apparently normal face-related activity on the other, especially in the FFA. The results support the hypothesis that while necessary, occipito-temporal face-related activity, at least as measured with fMRI, might not be sufficient on its own to allow normal face identification, and is rather more related to face detection. The exact role of these regions in either identifying or detecting other object categories remains to be established, probably with further studies of individuals suffering from complementary recognition deficits (e.g. (Moscovitch et al., 1997)). In any case, such findings may pose
important constrains on models of human object recognition. Thus, for example one valid interpretation of these findings is that the actual recognition or identification of faces (and maybe of other objects as well), is not completed at the level of high order, visual object-related areas such as the FFA, and activation in more anterior regions, where a more abstract or semantic representation may exist, is required for successful identification to be accomplished (e.g. Bar et al., 2001). It could be that in congenital prosopagnosia, the streaming of information from the FFA into those more anterior regions is disrupted, and this leads to the severe behavioral impairment in face processing. Current advances in imaging techniques such as Diffusion Tensor Imaging (DTI), which enables to track fibers connecting different brain regions, may be useful in order to test this hypothesis.
References


References


Appendix

Face-selective activation in a congenital prosopagnosic subject

Uri Hasson¹, Galia Avidan², Leon Y. Deouell³, Talma Hendler⁴, Shlomo Bentin⁵, Rafael Malach¹

Affiliation
¹ Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel.
² The Interdisciplinary Center for Neural Computation, Hebrew University of Jerusalem, Jerusalem, Israel.
³ Department of Psychology and Helen Wills Neuroscience Institute, University of California, Berkeley, USA.
⁴ Imaging Department, Wohl Institute for Advanced Imaging, Tel Aviv Sourasky Medical Center, Israel.
⁵ Department of Psychology, Hebrew University Jerusalem, Israel.
⁶ University, and Institute des Sciences Cognitives, CNRS, France.

Submitted
Abstract

Congenital prosopagnosia is a severe impairment in face identification manifested from early childhood in the absence of any evident brain lesion. We used fMRI to study the brain activity elicited by faces in such a patient (YT) in an attempt to shed more light on the nature of the brain mechanisms subserving face identification. The face-related activation pattern of YT in the ventral occipito-temporal cortex was similar to that observed in a control group on several parameters: anatomical location, activation profiles and hemispheric laterality. In addition, using a modified vase-face illusion we found that YT's brain activity in the face-related regions manifested clear global grouping processes that showed no significant difference compared to normal controls. However, subtle differences in the degree of selectivity between object and faces were observed in the lateral occipital cortex. These data suggest that face-related activation in ventral occipito-temporal cortex, although necessary, might not be sufficient by itself for normal face identification.

Introduction

Recent neuroimaging studies have identified a region within the human fusiform gyrus, which appears to be preferentially activated by images of faces compared to many other object categories (Halgren et al., 1999; Haxby, Hoffman, & Gobbini, 2000; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999; Kanwisher, McDermott, & Chun, 1997). This selectivity for faces in the fusiform gyrus is in agreement with previous evidence based on electrophysiological (event-related potentials; ERP) recordings from both the surface of the occipitotemporal cortex (N200; for a comprehensive review see (Allison, Puce, Spencer, & McCarthy, 1999) and from the posterior-inferior temporal scalp (N170, (Bentin, Allison, Puce, Perez, & McCarthy, 1996; George, Evans, Fiori, Davidoff, & Renault, 1996).

The selectivity for faces, as observed with fMRI, is preserved under various experimental manipulations of both the task and the configuration of the face stimuli. For example, similar levels of activation were found
in the fusiform face-related area (FFA) for either grayscale photographs or line drawings of human faces, cartoon faces, cat faces and different viewpoints of the same face (Tong, Nakayama, Moscovitch, Weinrib, & Kanwisher, 2000). In addition, similar levels of activation were observed when subjects attended faces in a demanding “one-back” memory task and during passive viewing (Tong et al., 2000). Similarly, the N170 is as distinctive of schematically drawn faces as of photographs of natural faces (Sagiv & Bentin, 2001), and insensitive to task manipulations and/or attention factors (Carmel & Bentin, 2002). Thus, these brain activations are sensitive to faces, rather than to low-level visual features of the stimuli.

The functional role of the FFA in face processing is still a matter of ongoing debate. One possibility is that activity in the FFA is associated with within-category face identification (but see Gauthier, Skudlarski, Gore, & Anderson, 2000). Support for this view comes from studies that have shown differential pattern of activation for familiar and unfamiliar faces (George et al., 1999; Henson, Shallice, & Dolan, 2000; Henson, Shallice, Gorno-Tempini, & Dolan, 2002; Katanoda, Yoshikawa, & Sugishita, 2000). However, other studies did not show such an effect (Dubois et al., 1999; Leveroni et al., 2000; Nakamura et al., 2000). These latter studies argue against the involvement of the FFA in face identification, stressing its role in face detection (i.e. distinction of face from non-face stimuli), and in the structural encoding of face parts into a coherent whole.

An additional approach for examining the functional role of the FFA and other face-related areas in the brain is to study individuals who suffer from prosopagnosia, a specific impairment in face identification (Bodamer, 1947; for a recent review see De Renzi, 1997). This syndrome is usually acquired after a bilateral lesion in the vicinity of the fusiform gyrus (Farah, 1995). Indeed, examining two patients whose prosopagnostic deficit was acquired during adolescence, Marotta, Genovese, & Behrmann, (2001) found alterations in the face-related activity in the fusiform gyrus. Specifically, although faces activated regions
in the posterior fusiform gyri of these two patients, the activity was posterior relative to that found in normal subjects.

While studies of acquired prosopagnosia reveal important information about face processing, their ability to unveil the function of essential regions in the normal brain is limited by several factors. First, natural lesions are rarely circumscribed to a particular region of interest. The lesions found in patients suffering from acquired prosopagnosia vary substantially in etiology, size and location and are not strictly confined to the FFA (Schweich & Bruyer, 1993). In fact, there is a considerable debate regarding the necessary and sufficient lesion that would result in prosopagnosia (see Farah, 1995, for review). Second, brain lesions might have a distant effect that, through connectional disruptions, goes beyond the specific locus of a lesion (diaschisis). Last, when functional imaging is performed in brain damaged patients, particular caveats arise, depending on the method of measurement. For example, in measuring ERPs, the conductivity of the brain is altered in the region of encephalomalacia, effecting the scalp distribution of potentials (e.g., Aboud, Bar, Rosenfeld, Ring, & Glass, 1996; see Deouell, Hämaläinen and Bentin, 2000 for further discussion). In fMRI, either the underlying disease (e.g., atherosclerosis) or the mechanism of brain injury (e.g., closed head injury or infection) may alter blood flow or disrupt the normal neuro-vascular coupling (e.g., Pineiro, Pendlebury, Johansen-Berg, & Matthews, 2002). Thus, interpretation of hemodynamic alterations as reflecting neural changes is risky.

An approach that might bypass these limitations is to study individuals who suffer from congenital prosopagnosia without evident structural brain lesion. Bentin, Deouell, & Soroker (1999) recently reported such a person, YT, a healthy 39 year-old businessman with no history of neurological disease, and no evidence of anatomical lesion, but who complained of having severe problems in recognizing faces since early childhood. When formally tested, YT was able to identify only 24 out of 670 famous faces (mixed with 580 faces of unknown individuals), while a control group of 24 subjects, matched with YT for age and education, identified an average of 391 faces from this set. In contrast to his face identification deficit, YT
Appendix was able to easily determine the gender, age and the emotional state on the basis of a person’s face, had no difficulty in recognizing objects other than faces, and exhibited normal or above normal performance in holistic and analytic visual, memory and cognitive tests (for details see Bentin et al., 1999). Bentin et al. further found that the N170, which is normally larger in response to faces than to other objects, is similarly large for face and non-face stimuli in YT.

Using fMRI we now tested the functional brain activity profile for face and non-face stimuli both in YT and in a control group of 12 subjects. In particular, we aimed at comparing the patterns of activity elicited by faces in the ventral occipito-temporal areas (VOT) of YT and the control subjects. Since face-related regions are part of a complex network of occipito-temporal object-related areas, we also looked at areas, which are preferentially activated by non-face stimuli. In addition we tested for holistic integration processes in YT using a modified Rubin face-vase illusion (Hasson, Hendler, Ben Bashat, & Malach, 2001).

Methods

Control Subjects

Thirteen control subjects (ages 26-49) participated in either one or two experiments: Twelve (7 females) participated in Experiment 1 and 10 (5 females) in Experiment 2. All subjects, including YT, provided written informed consent. The Tel-Aviv Sourasky Medical Center Ethic Committee approved the experimental procedure.

MRI Setup

Subjects were scanned in a 1.5T Signa Horizon LX 8.25 GE scanner equipped with a standard birdcage head coil. Blood oxygenation level dependent (BOLD) contrast was obtained with gradient-echo echo-planar imaging (EPI) sequence (TR = 3000, TE = 55, flip angle = 90°, field of view 24 × 24 cm², matrix size 80 × 80). The scanned volume included 17 nearly axial slices of 4mm thickness and 1mm gap. T1-
weighted high resolution (1×1×1mm) anatomical images and 3D spoiled gradient echo sequence were acquired on each subject to allow accurate cortical segmentation, reconstruction and volume-based statistical analysis.

Visual stimuli and experimental design
An interleaved short block design was used in both experiments. Each epoch lasted 9 sec, followed by a 6-sec blank screen. A central red fixation point was present throughout the experiments. The stimuli were generated on a PC, and projected via LCD projector onto a tangent screen located in the scanner. During both experiments subjects were instructed to identify whether two consecutive images were identical or not (1-back memory task). One or two consecutive repetitions of the same image occurred in each epoch.

Experiment 1
The visual stimuli used in the first experiment (Figure 1) included line drawings of faces, buildings, common man-made objects and geometric patterns. Nine images of the same type were presented in each epoch; each image was presented for 800 msec and was followed by a 200 msec blank screen. Each experimental condition was repeated 7 times, in pseudo-random order. The experiment started with 27 sec and ended with 9 sec of a blank (fixation only) screen. The experiment lasted 450 sec.

Experiment 2
The second experiment (Figure 7) included a modified version of the Rubin-face and Rubin-vase stimuli, as well as line drawings of front-faces and goblets, which served as an independent localizer of the face-related regions. Contours of the Rubin-face were produced by tracing edges of face photographs. By duplicating each profile outline, a vase that shares the exact outline with the profiles was constructed. The illusion was modified by biasing the perception to one perceptual state or the other (vase or profiles). The biasing was accomplished by coloring one object in a uniform color and placing it over a striped background (see figure 7-A). In order to prevent subjects from seeing the complementary perceptual
interpretation each figure was presented for 200 msec only, and was followed by a masking grid that remained on the screen for 800 msec. The crucial point to note is that the Rubin-vase and Rubin-face have similar local features but give rise to different global perceptual states. The experiment started with 21 sec and ended with 12 sec of a blank (fixation only) screen. The experiment lasted 507 sec (for more details see Hasson et al., 2001).

Data Analysis

fMRI data were analyzed using the “BrainVoyager” software package (Brain Innovation, Maastricht, Netherlands) and with complementary in-house software. Data of each subject from each scan were analyzed separately. Preprocessing of functional scans included 3D-motion correction and filtering out of low frequencies up to 5 cycles per experiment (slow drift). The first three images of each functional scan were discarded. Statistical analysis was based on the General Linear Model (Friston et al., 1995). The GLM analysis was performed independently for the time course of each individual voxel. Each experimental condition (except for blank) was defined as a separate predictor, using a box-car shape and a hemodynamic lag of 3 sec for each predictor. Only voxels at the significance level of $p<10^{-5}$ (uncorrected) were included in the statistical maps. Clusters of six or more contiguous voxels were considered significant. Percent signal change for each subject in each experiment was calculated as the percent activation from a blank baseline.

The cortical surface was reconstructed from the 3D-spoiled gradient echo scan. The procedure included segmentation of the white matter using a grow-region function, the smooth covering of a sphere around the segmented region, and the expansion of the reconstructed white matter into the gray matter. The obtained activation maps were superimposed on the inflated cortex and the Talairach coordinates (Talairach & Tournoux, 1988) were determined for the center of each ROI (region of interest).
Selectivity measure and bootstrap analysis: To assess the degree of selectivity in various activated areas, we calculated the difference between two categories of stimuli (e.g. faces-house). The null hypothesis in our analysis is that YT is not different from the group of controls in this measure of selectivity. However, whereas there are established methods for comparing two groups, there is no clear method of rejecting the null hypothesis comparing a single subject to a small group, which is not necessarily normally distributed. We therefore used a non-parametric bootstrap analysis, which examined the probability of finding the observed difference between YT and the group’s mean. The analysis consisted of the following procedure, repeated $10^4$ times: 1) A group of 12 was selected randomly, with replacement, from among the 13 participants (including the controls and YT) and the mean was calculated; 2) A single “target” subject was randomly selected from among the 13 subjects and 3) the difference between the random “target” (stage 2) and the mean of the random group (stage 1) was noted. The cumulative frequency distribution of these differences was plotted, and the place of the real observed difference between YT and the actual control group was found in this distribution. This place represents the probability of finding the observed difference by randomly selecting a subject and a group from the population.

“Internal localizer” test: To obtain an unbiased statistical test within a scan, we took advantage of the short-block presentation and adopted a procedure, which we termed the “internal localizer” approach (Lerner, Hendler, & Malach, 2002). In this procedure a subset of the epochs served to localize regions of interest, while another subset, not used in the statistical localization tests, was used to evaluate the activation level. More specifically, for each localizer test (e.g. "faces" vs. "buildings") 2 statistical tests were conducted. In each test, four different epochs served as anatomical localizers, while the rest of the epochs were not included in the test, and the level of activation during these epochs was measured separately. The data obtained from all epochs, which were ignored in the statistical test (but consisted of the same type of stimuli), were averaged within each subject. These data were averaged across subjects.
and are presented by the diamonds in Figures 4, 6. Note that this approach has the advantage that the localizer test is done on epochs which were in the same scan in which the activation level was measured (rather than the more common separate localizer scan) - thus minimizing inaccuracies due to head motion. However, the measured activity in the other epochs is unbiased since these epochs were not included in the statistical localizer test.

Laterality index

A laterality index was calculated for all subjects. For each subject we defined face-related regions by contrasting face stimuli with building stimuli. A weighted-average of number of face-related voxels in the left vs. right hemisphere was calculated for each subject:

$$\frac{\text{RH-LH}}{\text{RH} + \text{LH}}$$

Where RH and LH stand for the number of face-related voxels in the right and left hemispheres, respectively. The laterality index ranges between 1 and –1, positive values indicating a bias to the right hemisphere, and negative values indicating bias to the left.

**Results**

Experiment 1: Localization of face-related-regions

First, we compared YT’s BOLD response to face and non-face stimuli with that of a control group of 12 healthy subjects. The stimuli were line drawings of faces, buildings, man-made objects, and geometric patterns (Figure 1). These were presented in a short-block design while subjects performed a one-back memory task (see methods for details). To locate activity elicited selectively by faces and by buildings, we conducted a statistical test searching for voxels that were preferentially activated by faces compared to buildings and vice versa. The results for each subject in the control group and for YT are presented on an inflated brain from a ventral view in the right and left panels of figure 2 respectively. A lateral view and unfolded view of the same data is shown for YT and a representative subject (SN) in figure 3.
In agreement with previous reports (Halgren et al., 1999; Hasson et al., 2001; Haxby, Hoffman, & Gobbini, 2000; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999; Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997; Puce, Allison, Asgæi, Gore, & McCarthy, 1996), in all control subjects faces activated the posterior fusiform gyrus (Figure 2) and the lateral occipital cortex (Figure 3). The face-related activity in the fusiform gyrus corresponds to the FFA (Kanwisher et al., 1997). Building-related activity was found in the vicinity of the collateral sulcus (blue foci in Figure 2).

YT exhibited a pattern of activation (left panel of figure 2 and 3) similar to that observed in the control group. That is, there was a clear focus of face-related activity in the vicinity of the fusiform gyrus. Note that despite substantial inter-subject variability, we found preferential activation for faces in all subjects including YT, which was invariably located laterally to the activation for building images in the collateral sulcus (see also Malach, Levy, & Hasson, 2002). A similar face-related activation map was found while contrasting the face stimuli with the object stimuli (not shown, for similar contrast see also experiment 2). Moreover, YT’s face-related regions were located within the range of face-related Talairach coordinates of the control group, both in the fusiform gyrus and in the lateral occipital cortex (Table 1).

Activation profile of face-related-regions

In order to assess the degree of selectivity of the face-related regions we measured the level of activation elicited in these areas by faces, buildings, objects and patterns in YT and in the control subjects. Each face-related region was defined using the “internal localizer” approach (see methods for details). The results are shown in Figure 4. In agreement with pervious findings, in both the lateral occipital and fusiform face-related foci, the activation to faces in the control group (blue diamonds) was significantly greater than the activation to buildings (paired t-test p<10^{-6}). YT’s activation was similar to that of the control group; the level of activation for each category, within each region of interest, was within the range
of one standard deviation of the mean from the control group, hence, not statistically different (except for
the activation elicited by patterns in the right FFA).

To further estimate the similarity between YT and the control group we also calculated a selectivity
measure for faces compared to objects in each of these areas (see Method). This measure calculated for
YT was within the range of one standard deviation of the mean of the control group in the right and left
FFA and in the right LO but not in left LO. In order to estimate the statistical significance of these findings
we used the bootstrap method, which assessed the probability of observing the difference found between
YT and the control group by chance (see Methods). Corresponding with the pattern described above, this
probability was high (indicating insignificant difference) for the right FFA (0.4), left FFA (0.44), and right
LO (0.4), but low (approaching significance) in the left LO (p<0.053).

Laterality index
Many studies have shown that face-related activity in normal subjects shows a right hemisphere bias (e.g.
Kanwisher et al., 1997). We therefore calculated a laterality-index for each subject, which compared the
number of right vs. left hemisphere face-related voxels. The index ranges between 1 and −1, where
positive values indicate a bias to the right hemisphere, and negative values indicate a bias to the left (see
Methods for details). This laterality index was calculated for both the fusiform and the lateral-occipital
face-related regions (Figure 5). YT’s laterality index in the fusiform gyrus was biased toward the left.
However, such left bias was also found in four additional control subjects, with no apparent deficit in face
identification. Moreover, the left bias of two control subjects (BB and HB) exceeds that of YT. Using the
bootstrap method, the probability of finding YT’s index compared to the group’s by chance was 0.17 (i.e.
not significant). Thus, YT’s left lateralization within the fusiform gyrus cannot be reliably taken as a
functional marker for his behavioral deficit. In contrast to the fusiform gyrus, YT’s laterality-index in the
lateral-occipital cortex is biased toward the right hemisphere, and is within the range of the control group.
Activity elicited by non-face stimuli

The FFA is part of a complex network of areas specifically responsive to objects within the human occipito-temporal cortex (Grill-Spector et al., 1998; Ishai et al., 1999; Malach et al., 1995). Thus, it could be that YT’s behavioral deficit would be manifested in the activity of other object-related regions within the occipito-temporal cortex. We therefore measured the level of activation in the collateral sulcus (buildings vs. faces) and lateral-occipital cortex (objects vs. faces) in YT and in the control group (Figure 6). In agreement with previous reports (Epstein, Harris, Stanley, & Kanwisher, 1999; Kourtzi & Kanwisher, 2000; Levy, Hasson, Avidan, Hendler, & Malach, 2001; Malach et al., 1995) in the control group the activity in the collateral sulcus was stronger for buildings compared to faces (paired t-test p<10^{-5}), while the activity in the object-related region in the lateral-occipital cortex was stronger for objects compared to faces (paired t-test p<10^{-4}).

YT’s activation profile within the collateral sulcus was again similar to that of the control group, with levels of activity within the range of one standard deviation of the mean of the control group (Figure 6a). The stimulus selectivity measure for buildings compared to faces in this region was not significantly different than that of the control group (p<0.41, for both the right and left collateral sulcus). The selectivity measure for objects compared to faces in the lateral occipital cortex revealed a more complex pattern (Figure 6b). While YT’s left object-related region was not significantly different from that of the control group (p<0.22), the right object-related region in YT, as opposed to the control group, showed only weak preference for objects, a trend that was nearly statistically significant (p<0.056).

To summarize the results so far, although YT is severely impaired in face identification, the activity elicited by faces in the fusiform gyrus and lateral occipital regions was within the normal range of the control group in: the anatomical location (figure 2-3), Talairach coordinates (table 1), gross selectivity (figure 4) and laterality-index (figure 5). Whereas no differences were found between YT and the control group in the FFA region, subtle differences may still exist in the lateral occipital cortex, as a tendency was found
towards reduced selectivity for objects vs. faces in the right object-related LO, and for faces vs. objects in the left face-related LO. Of course, activity per se does not imply that the perceptual process in the activated areas was normal. Experiment 2 was designed to partially address this issue.

Experiment 2: Holistic vs. local processing in face-related regions

One of the hallmarks of face-identification is its reliance on holistic processing; that is, grouping the face components into a global facial configuration. It is possible that YT’s apparently normal face-related activity was induced by processing face parts (e.g. eyes, nose, and lips), while his performance impairment results from a failure to integrate those parts into a whole.

To explore this possibility, we used a modified version of the Rubin face-vase illusion, which was previously used by Hasson et al. (2001). In the Rubin face-vase illusion, the same local contours create two different visual percepts, depending on the figure-ground segmentation. Thus, if local features mediate face-selective activation, we would expect that activation in face-related regions would be similar in the Rubin-vase and Rubin-face perceptual states since the local feature structure is similar in the two images. In contrast, if global processes, which go beyond the local contours, drive the face-related activation, we would expect higher activation in these regions to the Rubin-face perceptual state compared to the Rubin-vase perceptual state. Our previous study showed that face related activation in normal subjects is modulated by global grouping processes, and is not induced solely by representation of local stimulus features. In the current study we used the same paradigm to test whether YT’s face-related activation would also reveal such holistic processes.

Face-related regions were independently localized by contrasting line drawings of faces with goblet images (figure 7-A). As in the previous experiment face-related regions in YT were found in two locations: the lateral occipital cortex and the fusiform gyrus (Figure 7-B). These regions overlapped the regions identified in Experiment 1 (Lateral occipital cortex: Right, 42, -74, -0. Left, -47, -73, -1. Fusiform gyrus:
Right, 33, -54, -14. Left, -31, -56, -19). This overlap, which was found despite the use of an entirely new set of object stimuli, provides an independent replication of the normal localization of face regions in YT, as observed in the previous experiment.

The holistic versus local processing was assessed by comparing the MR signal during the Rubin-face and Rubin-vase perceptual states. The mean MR signal in the control group for the Rubin-face and the Rubin-vase stimuli in the lateral occipital and fusiform regions are presented in Figure 7-C (data from Hasson et al., 2001). Despite the similarity between these stimuli, the face selective regions were activated more by the Rubin-faces compared to the Rubin-vases. This effect occurred both in the fusiform gyrus (paired t-test, p<0.01), and in the lateral occipital region (paired t-test, p<0.01). YT’s activation for each category was again within the range of one standard deviation of the mean of the control group in both regions, as was the difference between the two categories (p=0.45 for the FFA, p<0.26 for LO). Thus, similar to the control group, activation in YT’s face-related regions were affected by the holistic configuration of faces and not only by their local features.

**Discussion**

In the present study we report the case of a congenital prosopagnosic person, YT, who is markedly impaired in face identification, despite having a grossly normal pattern of face-related activation as measured by fMRI. Moreover, preferential activation during the Rubin-face compared to the Rubin-vase perceptual states suggests that global grouping processes contribute to his face-related activation. These findings suggest that YT’s profound prosopagnosia is not a result of decreased activation or lack of selectivity in occipito-temporal face-related regions, nor can it be easily accounted for by holistic processing impairment. Therefore we suggest that occipito-temporal face-related activity, at least as measured with fMRI, is not sufficient on its own to allow normal face identification.
YT reveals an intriguing discrepancy between behavioral impairment in face identification on the one hand, and an apparently normal face-related activity on the other, especially in the FFA. How could this discrepancy be interpreted? Within the framework of a visual perception model that distinguishes between the detection and structural encoding of the visual percept, on the one hand, and its within-category identification, on the other (e.g., Marr, 1982), the present results support the hypothesis that the FFA and perhaps the lateral occipital cortex are more involved in the former than in the latter process. This conclusion is congruent with previous fMRI and electrophysiological findings.

In normal subjects, FFA activity was hardly affected by face inversion, despite a drastic reduction in face identification performance (Kanwisher, Tong, & Nakayama, 1998). In contrast, in the same study, the activation in the FFA was drastically reduced when two-tone “Moony-faces” were inverted. The inversion of Moony-faces causes a substantial reduction in the ability to recognize these images as faces, probably because of disruption in the integration of the inverted Moony-face features into a coherent percept of a face. Thus, these data support the view that the FFA is involved in structural encoding of face-stimuli as such, rather than with within-category identification. Similarly, ERP studies identified the scalp N170 as a component with prominent selectivity for faces, and have shown that while being preferentially elicited by face parts, this component is not sensitive either to the familiarity of the face (Bentin & Deouell, 2000), or to its internal configuration (Bentin et al., 1996).

The hypothesis that the FFA is involved in detection and structural encoding of faces is fully compatible with YT’s behavioral profile, since he can easily distinguish between face and non-face stimuli, despite his impairment in face identification. Our finding of holistic figure-ground segregation in YT’s face–related regions is also consistent with his performance, which exhibited normal holistic processing in a series of visual tasks (Bentin et al., 1999). Moreover, a recent behavioral study of another congenital prosopagnosic person, BC, (Duchaine, 2000), showed that similar to YT, while BC was severely impaired
in face identification, he performed normally in tasks requiring the reconstruction of visual configurations. Although there are reports of congenital prosopagnosics that did not develop configurative processing (e.g., case AV reported by de Gelder & Rouw 2000, and case EP reported by Nunn, Postma, & Pearson 2001), the case of BC demonstrates that prosopagnosia associated with apparently normal configurative processing is not peculiar to YT.

In a recent imaging study with two prosopagnosic patients, whose deficit was acquired during adolescence, Marotta et al. (2001) also found face-related activation in the vicinity of the fusiform gyrus, but as opposed to control subjects the face-related activation in the two acquired patients was found in posterior parts of the fusiform gyrus, and not in its more anterior parts, which correspond to the FFA. In another fMRI study Hadjikhani and de Gelder did not find any face selective activation within occipito-temporal cortex, and particularly in the FFA, in two acquired prosopagnosic patients (Hadjikhani & De Gelder, 2002). The lack of FFA activation in acquired prosopagnosic patients might imply that the FFA is necessary for face identification. Our finding of a normal FFA activation in YT is a complementary finding suggesting that the activation in the FFA is not sufficient for face identification. Moreover, it points to different mechanisms of impairments in the acquired and in the congenital cases. Hadjikhani and de Gelder also tested the activation pattern in a congenital prosopagnosic patient (AV). Whereas YT showed clear face-selective and object-selective activations with normal spatial distribution in ventral occipito-temporal region, in AV faces and objects produced similar level of activation in the regions corresponding to the FFA and LO. This discrepancy might point to variable mechanisms of prosopagnosia even in congenital cases. As mentioned above, AV, unlike YT, had impaired configurative processing (de Gelder & Rouw, 2000). YT therefore presents a minimal case, in displaying pure deficit in face identification, with preserved selective activation in the FFA, providing critical information regarding the sufficiency of these brain activations for face identification.
In agreement with our current finding of normal face-related MR activation in YT, Bentin et al. (1999) found a distinct N170 for faces in YT, which was roughly similar in strength and distribution to that of the control group. However, they also found that while for control subjects the N170 was significantly larger for faces than for objects, for YT it was elicited to the same extent by both categories. This finding was recently replicated in 2 other congenital prosopagnosia cases (Sagiv, Barnes, & Robertson, 2000). In the current study we found a strong trend toward reduced selectivity for objects in the right object-related LO, and for faces in the left face-related LO. These results may reflect the same abnormality that yielded YT’s non-selective N170 results. Taken together, the ERP and the fMRI findings suggest that the selectivity for faces in his lateral occipital regions is less prominent than in normal people. Accordingly, a possible source of congenital prosopagnosia is an inefficient distinction between the encoding of faces and other categories. This might lead to a failure in development of dedicated face-specific processing or storage strategies, which are necessary for face identification (Bentin et al., 1999).

Although our current finding suggests that the FFA is involved in face detection and configuration rather than in identification per se, we do not rule out the possibility that the FFA also contributes to face identification. Indeed, it is quite possible that the activation of the FFA (as a face “detector”) is a necessary step in normal face identification, which is then followed by the operation of higher cortical regions (e.g. Bar et al., 2001). That is, the FFA activity might be necessary, but is not sufficient nor a final stage of face identification. In agreement with this possibility, a recent study by Tippet et al. reported patients with preserved face perception, but with impairment in learning to recognize new (previously unfamiliar) faces (Tippett, Miller, & Farah, 2000). The authors termed this selective deficit prosopamnesia, and argued for a division of labor between neural systems for learning and neural systems for representation of familiar faces. It could be that YT has all the information needed to recognize faces (reflected in the normal pattern of fMRI activation in the FFA), but lacks the ability to adequately stream face information to higher order, face-specific knowledge-base representation (cf. Bentin et al., 1999). When this deficiency is congenital, the result is prosopagnosia rather than prosopamnesia.
Finally, it should be noted that the current fMRI spatial resolution is in the order of millimeters, while the functional units involved in face processing (as observed by optical imaging) might be in the range of hundreds of microns (Wang, Tanaka, & Tanifuji, 1996; Wang, Tanifuji, & Tanaka, 1998). Therefore, it remains a possibility that despite the preserved overall activation by faces, the representation of faces in the FFA itself is disrupted in a way that renders face identification inefficient, and this distortion is not captured in the overall fMRI signal (for a discussion of this issue see Avidan, Hasson, Hendler, Zohary, & Malach, 2002).

Conclusion

Studying individuals who lack specific cognitive abilities is a powerful approach for establishing the relation between brain activity and cognitive roles. The fact that YT’s brain has not suffered trauma or neurological disease that may complicate the analysis of hemodynamic or electrical measurements makes this case particularly informative. The finding of selective activation in the ventral occipital temporal cortex for faces vs. non-face stimuli in a subject who lacks the ability to recognize familiar faces, suggests that this preferential activation is not sufficient on its own for normal face identification. In addition, the conjunction of fMRI and ERP data from the same subject suggests that adequate selectivity in the lateral parts of the occipito temporal cortex, manifested as early as 170 ms, may be essential for face identification.
References


Figure 1

A. Examples of the four stimulus categories used in the experiment: faces, buildings, common objects, and patterns.

B. An interleaved short block presentation design was used in the experiment. Each epoch lasted 9 sec, followed by a 6-sec blank. Nine images of the same type were presented in each epoch.
Figure 2

Figure 3
Figure 2: Ventral view of category-related activation in YT and normal control subjects.
Enlarged on the left is the activation obtained in YT’s brain for faces (faces>buildings; orange) and buildings (buildings>faces; blue), the data is presented on an inflated brain shown from a ventral view. On the right is the same activation obtained for the 12 control subjects. Despite substantial inter-subject variability, the preferential activation for faces was consistently located laterally to the building-related activation in all subjects and also in YT. Specifically note the clear face-related activation in the vicinity of the Fusiform gyrus.

Figure 3: Lateral and unfolded view of category-related activation in YT and one representative subject.
The activation obtained in YT’s (left) and one representative subject (SN – right), for faces (faces>buildings; orange) and buildings (buildings>faces; blue), the data is presented on a lateral view of an inflated brain and unfolded view of the same hemisphere.
**Figure 4**

Activation profiles of face-related voxels in the lateral-occipital cortex and the fusiform gyrus. The data was sampled using the ‘internal localizer’ approach (see Methods for details). Each graph shows the activation profile obtained for YT and 12 control subjects for each stimulus category presented in the object mapping experiment (faces, building, objects, and pattern stimuli). Bars indicate the percent signal change for YT, blue diamonds indicate the mean percent signal change across the control group, and error bars indicate ± 1 standard deviation. Note that the activation obtained for YT for each category in each face-related region is within the range of one standard deviation of the mean of the control group.
**Figure 5**

A laterality index, which compares the number of face-related voxels in the right vs. left hemisphere was calculated for each subject (see Methods for details). Top chart: laterality index in the Fusiform gyrus, bottom chart: laterality index in the Lateral-occipital cortex. Each blue diamond represents the laterality index of a single control subject, the red diamond represents YT. The laterality index ranges between 1 and –1, positive values indicate a bias to the right hemisphere, and negative values indicate bias to the left. Note that YT’s laterality index is within the distribution of the control subjects.
Figure 6: activation profiles of object-related regions.

Activation profiles for object-related voxels in the lateral-occipital cortex and the building-related voxels in the collateral sulcus (see figure 2 for details). Note that the activation obtained for YT for each category in each object-related region is within the range of one standard deviation of the mean of the control group.
Figure 7

A. Examples of the stimuli shown in the experiment: Localizer stimuli (top row) - to independently localize face-related regions we included line drawings of front-face, which were contrasted with goblet images. Rubin vase-face stimuli - to reveal holistic aspects of face activation we employed a modified version of the Rubin vase-face illusion, presenting a Rubin-face (left) and Rubin-vase (right).

B. Localization of face-related activation for YT: face-related activation obtained by contrasting the front-face stimuli with the goblet stimuli.

C. Activation profiles in the lateral-occipital face related region (left), and the posterior fusiform face-related region (right). Each graph shows the percent signal change obtained for YT and 12 control subjects for the uniform vase and the uniform-profile stimuli. Bars indicate data from YT, blue diamonds indicate the mean percent signal change across the control group, and error bars indicate ± 1 standard deviation. Note that in YT, as in the control group, face-related regions showed significantly higher activation to the Rubin-face compared to the Rubin-vase condition, despite the presence of similar local features in the two conditions.

Figure 7: Rubin Face-Vase Experiment
Table 1: Talairach coordinates
Talairach coordinates for the Lateral-occipital and Fusiform gyrus, derived from YT and from 12 control subjects in exp. 1. Values represent the mean ± SD in mm. Note that YT coordinates are within the range of the Talairach coordinates of the control group.

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R L</td>
<td>R L</td>
<td>R L</td>
</tr>
<tr>
<td>Lateral occipital</td>
<td>YT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>43 -43</td>
<td>-69 -66</td>
<td>3 -1</td>
</tr>
<tr>
<td>Controls</td>
<td>39±4</td>
<td>-42±5</td>
<td>-69±7</td>
</tr>
<tr>
<td>Fusiform gyrus</td>
<td>YT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>35 -31</td>
<td>-52 -53</td>
<td>-15 -19</td>
</tr>
<tr>
<td>Controls</td>
<td>36±5</td>
<td>-38±4</td>
<td>-48±8</td>
</tr>
</tbody>
</table>