

Object Perception: Physiology

Accepted Dec. 26th, 2008

Object perception or object recognition is the process in which visual input is assigned a meaningful interpretation that is available to perceptual awareness. It is fundamental to our ability to interpret and act in the world. Object perception is thought to occur through computations across a hierarchy of processing stages in visual cortex, named the *ventral visual pathway*. This pathway begins in the primary visual cortex, area V1 in the occipital lobe, and ascends to regions in lateral occipital cortex and ventral occipito-temporal cortex. Damage (such as lesions, stroke, disease) to higher level visual areas in this pathway leads to specific deficits in object perception, such as the inability to recognize objects (object agnosia) and/or inability to recognize faces (prosopagnosia) while not affecting other visual abilities, such as determining the motion of objects, or their contrast. As such, higher level regions in the ventral stream are thought to be necessary for conscious object perception.

Before the advent of noninvasive neuroimaging methods such as functional magnetic resonance imaging (fMRI) in the mid 1990's, knowledge about the function of the ventral stream was based on single unit electrophysiology measurements in monkeys and on lesion studies. These studies showed that neurons in the monkey inferotemporal (IT) cortex respond selectively to shapes and complex objects, and that lesions to the ventral stream can produce specific deficits in object recognition. However, it is difficult to make

inferences on the relation between responses of single neurons in the monkey's brain and its perception of objects. Further interpreting lesion data is complicated because lesions are typically diffuse, they usually disrupt both a cortical region and its connections and they are not replicable across patients.

With the advent of neuroimaging there has been an enormous advancement of our understanding of the neural basis of object perception in humans. The first set of fMRI studies of object and face perception in humans identified the regions in the human brain that respond selectivity to objects and faces. Then a series of studies demonstrated that activation in object-selective and face-selective regions correlates with success at recognizing object and faces, respectively, providing striking evidence for the involvement of these region in perception. Once researchers found which regions in cortex are involved in object perception the focus of research shifted to examining the nature of representations and computations that are implemented in these regions to understand how they enable efficient object recognition in humans.

Object-Selective Regions in the Human Brain

Object-selective regions in the human brain consist of a constellation of regions in the lateral and ventral occipito-temporal cortex, called the lateral occipital complex (LOC). These regions are defined functionally. Using fMRI scientists scan subjects' brains while they view different kinds of visual stimuli and determine which parts of the brain respond selectively to certain stimuli over other stimuli. Scientists found a constellation of regions in the lateral occipital cortex that respond more strongly when subjects view pictures of

objects (e.g. cars, abstract sculptures, etc) than when they view scrambled images of these objects, textures or patterns. The LOC includes a posterior region, LO, located adjacent and posterior to a region involved in processing visual motion (area MT) and a ventral region, pFus/OTS which overlaps the occipito temporal sulcus (OTS) and the posterior part of the fusiform (Fus) gyrus (**Fig. 1**).

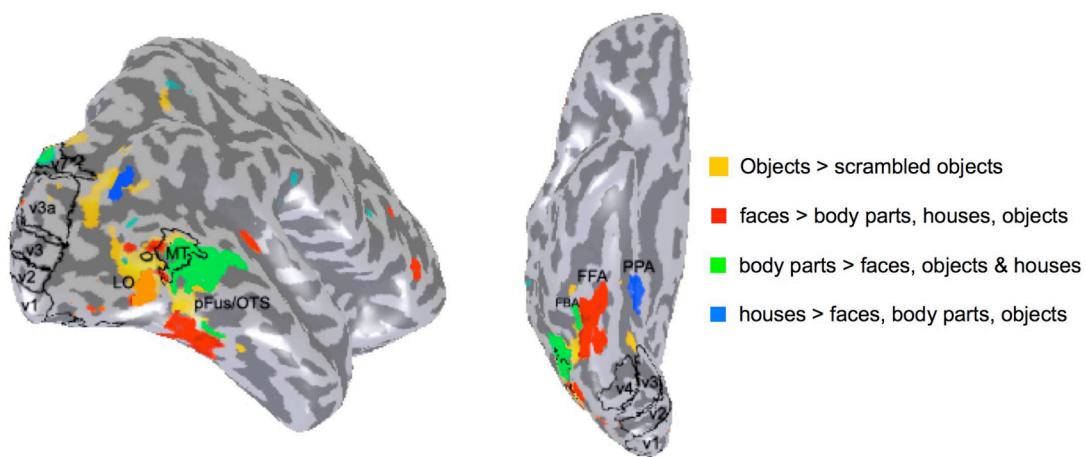


Figure 1: Object (yellow), Face (red), bodypart (green) and house (blue) selective regions presented on an inflated cortical surface of a right hemisphere of a representative subject. Dark grays indicate sulci and lighter gray indicate gyrii. *Left:* lateral view. *Right:* ventral view. Orange indicates regions that are both face and object-selective. *LO:* lateral occipital. *pFUS/OTS:* posterior fusiform/occipito-temporal sulcus. *MT:* mid temporal motion selective region. *FBA:* fusiform bodypart area; *FFA:* fusiform face area. *PPA:* parahippocampal place area. Boundaries of early visual areas *V1*, *V2*, *V3*, *V3a* and *V4* are shown in black.

LOC responds similarly to many kinds of objects and object categories, including novel objects. It is thought to be in the intermediate or high-level stages of the ventral stream visual hierarchy after early visual areas (V1, V2, V3, and V4) because it responds to specific visual stimuli (objects and shapes; whereas early visual areas respond to all visual stimuli including texture patterns) and shows lesser sensitivity to low-level visual information such as stimulus contrast, stimulus position and stimulus size compared to

early visual regions. There are also object-selective regions in the dorsal stream, which are located on the dorsal part of the occipital lobe, including a region lateral to V3a, and regions overlapping the intraparietal sulcus (IPS) in the parietal lobe. However, these dorsal object-selective regions do not correlate with object perception and may be involved in computations related to visually guided actions towards objects. A review of dorsal object-selective cortex is beyond the scope of this entry.

The Role of the LOC in Object Perception

Although the LOC is activated strongly when subjects view pictures of objects this does not by itself prove that it is the region in the brain that ‘performs’ object recognition. There are many differences between objects and scrambled objects (or objects and textures). Objects have a shape, surfaces, contours, they are associated with a meaning and semantic information and they are generally more interesting than textures. Each of these factors may affect the higher LOC response to objects than control stimuli. Nevertheless, several studies show that activation in the LOC is correlated with object perception rather than low-level features in the visual stimulus.

Converging evidence from multiple studies revealed an important aspect of coding in the LOC: it responds to object shape, not low-level or local visual features. These studies used stimuli that were controlled for low-level visual information across objects and nonobject controls and showed that LOC responds only when there is a global shape in the stimulus even when the low-level information is identical. These experiments have shown that LOC responds more strongly to: (1) objects defined by luminance compared

to nonobjects generated from luminance (including bars, patterns and scrambled objects), (2) objects generated from random dot stereograms than formless random dot stereograms, (3) objects generated from structure-from-motion relative to random motion and (4) objects generated from textures than flat texture patterns made of the same basic components. Further, LOC's response to objects is also similar across object formats (e.g. gray-scale photographs, black and white line drawings of objects, and object silhouettes) and it responds selectively to objects delineated by both real and illusory contours. Overall, these data suggest that LOC's activation to objects exhibits *cue invariance*. That is, it is insensitive to the specific visual cues that define an object.

Several recent studies also showed that the LOC responds to shape rather than surfaces or local contours. LOC's responses are higher to global shapes than surfaces or planes, and higher to global shapes than scrambled contours of the same shapes. For example, occluding bars presented in front of an object (which resembles the real life situation of seeing objects behind blinds) or stimuli with incomplete contours that still define a clear global shape (e.g. drawing an object's contour in which some of the strokes are missing), do not significantly reduce LOC's response as compared to unoccluded shapes or shapes with complete contours. These data underscore the findings LOC's responses are driven by shape rather than low-level visual information that generates form or local features in the visual stimulus.

Importantly, LOC activations are correlated with subjects' object recognition performance. High LOC responses correlate with successful object recognition, and low

LOC responses correlate with trials in which objects are present, but are not recognized.

The correlation between object perception and brain activation in the LOC has been demonstrated using various methods. Some researchers have monotonically varied a parameter that affects the ability of subjects to perceive objects and have tracked the correlation between changes in brain activation and subjects' recognition performance. Parameters that have been manipulated include image presentation duration, the amount of visible object behind occluding blinds, object contrast, and the degree of coherence of objects presented in noise. Such studies have shown that pre-recognition activation in the LOC is lower than when objects are recognized, and that the level of activation increases monotonically until objects are recognized. Other studies used a different approach in which they showed subjects pictures close to *recognition threshold* (the border between seeing and not seeing), and compared brain activation in trials where subjects recognized objects to brain activation in trials where objects were present but not recognized. These experiments showed that LOC responses are higher in trials in which subjects detected objects than when they did not detect objects and LOC's responses were highest when objects were correctly identified.

Neural Mechanisms of Invariant Object perception

One of the reasons that object perception is difficult because there is significant variability in the appearance of an object (e.g. various views of a car look very different) which necessitates considerable generalization, yet the visual system needs to retain specificity to discriminate between objects that are similar (e.g. different cars contain the same parts arranged in a similar configuration). There are many factors that can affect the

appearance of objects including: (i) the object being at different locations relative to the observer, (ii) the 2-dimensional projection of a 3-dimensional object on the retina varies considerably due to changes in its viewpoint relative to the observer (iii) differential illumination which affect objects' color, contrast and shadowing and (iv) occlusion. Nevertheless, humans are able to recognize objects in a tenth of a second across large changes in their appearance. This ability is referred to as *invariant object recognition*.

How does the LOC deal with the variability in objects' appearance? One view suggests that invariant object recognition is accomplished because the underlying neural representations are invariant to the appearance of objects. Thus, there will be similar neural responses even when the appearance of an object changes considerably. Other theories suggest that invariance may be generated through a sequence of computations across a hierarchically organized processing stream in which the level of invariance increases from one level of the processing to the next. For example, at the lowest level neurons code local features and in higher levels of the processing stream neurons respond to more complex shapes and are less sensitive to object transformations.

fMRI studies of invariant object perception found differential sensitivity across the ventral stream to object transformations such as size, position, illumination and viewpoint. Intermediate regions such as LO show higher sensitivity to image transformations than higher-level regions such as pFus/OTS. That is, LO shows sensitivity to object size, position, illumination and rotation, whereas pFUS/OTS shows lesser sensitivity to object size and position, but still shows considerable sensitivity to

object view and illumination. Notably, accumulating evidence from many studies suggests that at no point in the ventral stream neural representations are entirely invariant to object transformations. These results support an account in which invariant recognition is supported by a pooled response across neural populations that are sensitive to object transformations. One way in which this can be accomplished is by a neural code that contains independent sensitivity to object identity and transformation.

One object transformation that has been extensively studied is sensitivity to object position. Position invariance is thought to be accomplished in part by an increase in the size of neural receptive fields along the visual hierarchy. That is, as one ascends the visual hierarchy, neurons respond to stimuli across a larger part of the visual field. Findings from monkey electrophysiology suggest that even though neurons at the highest stages of the visual hierarchy respond to stimuli across a large part of the visual field, these neurons retain some sensitivity to object position and size. In humans, both the LOC and category-selective regions in the ventral stream (see below) respond to objects presented at multiple positions and sizes. However, the amplitude of response to object varies across different retinal positions. LO, pFUS/OTS as well as category-selective regions respond more strongly to objects presented in the contralateral vs. ipsilateral visual field. Some regions, such as LO, also respond more strongly to objects presented in the lower visual field than upper visual field. Responses also vary with eccentricity: LO, fusiform face-selective regions and word-selective regions in occipito-temporal sulcus respond more strongly to centrally than peripherally presented stimuli. In contrast,

place-selective regions respond more strongly to peripherally than centrally presented stimuli.

These studies of sensitivity to object position indicate that neural populations in the human ventral stream retain some degree of sensitivity to position. How then is position invariant recognition accomplished if position information is present in the neural response? Current theories suggest that this is accomplished by maintaining separable information about object identity and object position across the response of a neural population that codes both identity and position. First, maintaining separable information about object position and identity may allow preserving information about the structural relationships between object parts. (E.g. keeping information such as “the head is above the torso” and “the torso is above the legs”). Second, it may provide a robust way for generating position invariance by using a population code. That is, the pooled response across a population of neurons may generate position invariant perceptions. For example, consider a population of neurons. Each neuron’s response is modulated (that is, its response varies) according to the position of the object in the visual field. However, at each given position in the visual field each neuron’s response is higher for one object over other objects. Since the higher response for an object is consistent across positions it is possible to determine the object identity independent of position. Third, separable object and position information may allow concurrent localization and recognition of objects. That is, recognizing both *what the object is* and also determining *where it is*. This possibility challenges the prevailing hypothesis that position and identity information are represented in two parallel processing in which object identity (‘what’) information is

thought to be represented in the ventral processing stream and object position ('where') is thought to be represented in the dorsal processing stream.

Category Selective Regions in the Human Ventral Stream

In addition to the LOC, researchers found several additional regions in the ventral stream that show preferential responses to specific object categories (**Fig. 1**). Searching for regions with categorical preference was motivated by reports that suggested that lesions to the ventral stream can produce very specific deficits - such as the inability to recognize faces, or the inability to read words, while other visual (and recognition) faculties are preserved. By contrasting activations to different kinds of objects researchers found ventral regions that show higher responses to specific object categories such as a region in the left occipito-temporal sulcus that responds more strongly to letters than textures (the "visual word form area", VWFA), several foci that respond more strongly to faces than other objects (including the "fusiform face area", FFA) regions that respond more strongly to houses and places than faces and objects (including a region in the parahippocampal gyrus, the "parahippocampal place area", PPA) and regions that respond more strongly to body parts than faces and objects (including a region near MT called the "extrastriate body area", EBA and a region in the fusiform gyrus, "the fusiform body area", FBA). Nevertheless, many of these object-selective and category-selective regions respond to more than one object category.

Findings of category selective regions initiated a fierce debate about the principles of functional organization in the ventral stream. Are there regions in the cortex that are

specialized for any object category? Is there something special about computations relevant to specific categories that generate specialized cortical regions for these computations? In explaining the pattern of functional selectivity in the ventral stream, four prominent views have emerged.

Limited category-specific modules and a general area for all other objects

Nancy Kanwisher and co-workers suggested that ventral temporal cortex contains a limited number of modules specialized for the recognition of special object categories such as faces (in the FFA), places (in the PPA), and body parts (in the EBA and FBA). The remaining object-selective cortex (LOC), is a general-purpose mechanism for perceiving any kind of visually presented object or shape. The underlying hypothesis is that there are few “domain-specific modules” that perform computations that are specific to these classes of stimuli beyond what would be required from a general object recognition system. For example, faces, like other objects need to be recognized across variations in their appearance (a domain-general process). However, given the importance of face processing for social interactions, there are aspects of face processing that are unique. Specialized face processing may include identifying faces at the individual level (e.g. John vs. Harry), extracting gender information, gaze, expression etc. These unique face-related computations may be implemented in specialized face-selective regions.

Process maps

Michael Tarr and Isabel Gauthier proposed that object representations are clustered according to the type of processing that is required, rather than according to their visual attributes. It is possible that different levels of processing may require dedicated computations that are performed in localized cortical regions. For example, faces are usually recognized at the individual level (e.g. “That is Bob Jacobs”), but many objects are typically recognized at the category level (e.g. “That is a horse”). However, expert recognizers particularly excel at recognizing specific exemplars of their expert category (e.g. bird experts can easily distinguish between a “willow sparrow” and a “house sparrow”). These researchers hypothesized that expert processing uses processing similar to face processing because it requires fine-grained visual discriminations between items that have similar parts and configuration. Isabel Gauthier, Michael Tarr and colleagues further showed that the FFA responds more strongly in birds and car experts when they viewed images of their expert category compared to other common objects. Therefore, they suggested that the FFA is involved in fine-grain discrimination between exemplars of any object category and this processing is automatically recruited in experts.

Distributed object-form topography

James Haxby and coworkers posited an ‘object form topography’ in which occipito-temporal cortex contains a topographically organized representation of shape attributes. The representation of an object is reflected by a distinct pattern of response across all ventral cortex, and this distributed activation produces the visual perception. James Haxby and colleagues showed that the response to a given category could be determined

by the distributed pattern of activation across all ventral-temporal cortex. Further, they showed that it is possible to predict what object category subjects viewed even when regions that show maximal activation to a particular category were excluded. Thus, this model suggests that ventral temporal cortex represents object category information in an overlapping and distributed fashion. One of the appealing aspects of this distributed representation is that it allows representation of a large number of object categories. Second, this model posits that both weak and strong signals in the ventral processing stream convey useful information about object category.

Topographic representation

Rafael Malach and colleagues suggested that eccentricity biases underlie the organization of ventral regions because they found a correlation between category preference (higher response to one category over others) and eccentricity bias (higher response to a specific eccentricity than other eccentricities). They showed that regions that prefer houses to objects, also respond more strongly to peripheral than central (foveal) visual stimulation. In contrast, regions that prefer faces or letters respond more strongly to centrally presented stimuli than peripherally presented stimuli. Rafael Malach and colleagues proposed that the correlation between category selectivity and eccentricity bias is driven by spatial resolution needs. Thus, objects whose recognition depends on analysis of fine details are associated with central visual field representations, and objects whose recognition requires large-scale integration are associated with peripheral visual field representations.

Presently, there is no consensus in the field about which account best explains ventral stream functional organization. Much of the debate centers on the degree to which object proceeding is constrained to discrete modules or involves distributed computations across large stretches of the ventral stream. The debate is both about the spatial scale on which computations for object recognition occur and about the fundamental principles that underlie specialization in the ventral stream.

In sum, neuroimaging research and neurophysiological research in animals in the past decade has advanced our understanding of object representations in the human brain. These studies have identified the functional organization of the human ventral stream, showed the involvement of ventral stream regions in object recognition and have laid fundamental stepping stones in understanding the neural mechanisms underlying invariant object recognition.

- Kalanit Grill-Spector

Related Topics

Cortical Organization

Face Perception

Face Perception: Physiological

Perceptual Development: Visual Object Permanence and Identity

Perceptual Expertise

Object Perception

Recognition

Vision: Cognitive Influences

Visual Processing: Extra-Striate Cortex

Word Recognition

Suggested Reading

Andresen DR, Vinberg J, Grill-Spector K. (2009) *The representation of object viewpoint in the human visual cortex*. NeuroImage. Epub 2008 Nov 25th.

Gauthier I, Skudlarski P, Gore JC, Anderson AW. (200) *Expertise for cars and birds recruits brain areas involved in face recognition*. Nat Neurosci. 3(2):191-7.

Golarai G, Ghahremani DG, Whitfield-Gabrieli S, Reiss A, Eberhardt JL, Gabrieli JD, Grill-Spector K. (2007) *Differential development of high-level visual cortex correlates with category-specific recognition memory*. Nat Neurosci. 10(4):512-28.

Grill-Spector, K, Kushnir, T, Edelman S, Avidan G, Itzhak Y and Malach R. (1999)

Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron, 24(1):187-20

Grill-Spector K, Kushnir T, Hendlar T, and Malach R (2000). *The dynamics of object-selective activation correlate with recognition performance in humans.* Nature Neuroscience, 3(8):837-43.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. (2001) *Distributed and overlapping representations of faces and objects in ventral temporal cortex.* Science. 293(5539):2425-30.

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

Kriegeskorte N, Mur M., Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K,7 and Bandettini PA (2008) *Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey,* Neuron 60, 1126–1141.

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

Li N, DiCarlo JJ (2008) *Unsupervised natural experience rapidly alters invariant object representation in visual cortex.* Science 321(5895):1502-7.

Parr LA, Hecht E, Barks SK, Preuss TM, Votaw JR.(2009) *Face Processing in the Chimpanzee Brain.* Curr Biol. 2008 Dec 17. [Epub ahead of print]

Sayres R, Grill-Spector K. (2008) *Relating retinotopic and object-selective responses in human lateral occipital cortex.* J Neurophysiol. 100(1):249-67

Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006). *A cortical region*

consisting entirely of face-selective cells. Science. 311(5761):670-4.

Vinberg J and Grill-Spector K. (2008) *Representation of Shapes, Edges, and Surfaces*

Across Multiple Cues in the Human Visual Cortex. J Neurophysiol. 99(3):1380-

93.