The spatial organization of the brain’s object and face representations in the temporal lobe is critical for understanding high-level vision and cognition but is poorly understood. Recently, exciting progress has been made using advanced imaging and physiology methods in humans and nonhuman primates, and the combination of such methods may be particularly powerful. Studies applying these methods help us to understand how neuronal activity, optical imaging, and functional magnetic resonance imaging signals are related within the temporal lobe, and to uncover the fine-grained and large-scale spatial organization of object and face representations in the primate brain.

**Key words:** face perception; object recognition; single unit; local field potential; optical imaging; fMRI
thermore, V1 data demonstrated that the fMRI signal obtained from the intrinsic blood oxygenation level-dependent (BOLD) contrast is slightly better correlated with an indirect measure of synaptic activity [local field potentials (LFPs)] than with multi-unit spiking activity (MUA) (Logothetis et al., 2001).

Here, we are not so much interested in the relationships between these techniques per se, but in how the combination of these techniques can help us to clarify the functional organization of regions in the temporal lobe. Our understanding of this functional organization is growing steadily, but is still not sufficient to address the big question of how the spatial organization of the temporal lobe relates to its functional role in visual object recognition. One hypothesis is that there are domain-specific systems containing neurons and circuitry for classes of complex objects such as faces, bodies, etc. Another hypothesis is that there is alphabet of small columns where neurons inside each column prefer a particular feature and the active combination of columns represents each object. A third possibility is that neurons that participate in representing the same object are fully intermixed, perhaps organized into multiple interleaved feature maps.

Here, we will focus on how the combination of electrophysiological and imaging techniques can help unravel the functional organization of the temporal lobe. We will first briefly describe what each technique in isolation has revealed and then we will illustrate the power of their combination. In comparing across techniques, we will address two central questions. First, what are the processes that contribute to the signal measured with each of these methods? Although positive correlations have been found between the different methods, these correlations are smaller than expected given the reliability of the data. This opens the possibility that different methods measure different signals, for example, a differently weighted contribution of the input and the output of neurons. Second, what is the degree and spatial scale of clustering of neurons with similar properties in the temporal lobe? The answers from different techniques (e.g., optical imaging vs fMRI) and different designs (e.g., use of different stimuli) are qualitatively similar, but quantitatively different.

**Electrophysiology in monkeys**

Gross et al. (1969, 1972) gave an early, qualitative description of neural receptive fields and response properties of neurons in primate inferior temporal (IT) cortex, suggesting cells with relatively large receptive fields that are selective for complex visual stimuli like faces and hands. Decades later, these findings are still valid, although significant research has quantified and qualified these response properties of IT neurons. Many IT neurons are activated strongly by relatively complex stimuli (Perrett et al., 1982; Desimone et al., 1984), but these stimuli are often moderately complex visual features or object fragments rather than images of whole objects (Kobatake and Tanaka, 1994). Furthermore, although this stimulus selectivity is largely tolerant to image transformations like changes in position and size (Ito et al., 1995; Wallis and Rolls, 1997), single neurons nevertheless retain a surprising degree of sensitivity for such transformations (Op de Beeck and Vogels, 2000; DiCarlo and Maunsell, 2003). Computational work and population analyses of neuronal selectivity suggest that a representation of moderately complex features that is largely tolerant for metric transformations like size and position may be an effective population code allowing for invariant object recognition (Hung et al., 2005; DiCarlo and Cox, 2007). Absolute invariance and selectivity for whole objects are not necessary.

The functional organization of IT cortex is another question that has attracted much interest. Are there feature, object, or category columns/maps in IT cortex analogous to orientation columns in V1 or motion-direction maps in MT? Early single-unit electrophysiological studies suggested some clustering (Desimone and Gross, 1979; Gochin et al., 1991), but nevertheless the similarity in response properties between nearby neurons was not very high. Electrophysiology studies have also investigated regional variations in stimulus preference on a larger scale. Most notably, Baylis et al. (1987) described some regional variation in the preference for simple visual stimuli, more complex visual stimuli, and faces, but this variation was relatively modest. Fujita et al. (1992) showed the existence of a finer columnar organization with respect to the simplest visual stimuli that activate neurons efficiently (the “critical features”). A critical feature for a neuron was determined by simplifying the effective stimulus little by little without changing evoked responses of the neuron (the reduction technique). The resulting moderately complex features were similar across nearby neurons and neurons with similar preferences seemed to be organized in a columnar manner in IT cortex. The connectivity within IT cortex (between areas TEO and TE) and afferent input to IT cortex also suggests a columnar organization at a size of a few hundred micrometers (Saleem et al., 1993; Tanigawa et al., 2005).

One question is why the reduction technique proved so useful for finding clustering of neurons with similar properties, while at the same time it is not uncommon to find nearby neurons with a totally different selectivity. One possibility is that the reduction technique partially removes the effects of the intrinsic circuit (e.g., interneurons and horizontal connections) that determines how an IT neuron responds to the afferent input, and so the effect of the afferent input would be greater using the reduction technique than when using more complex stimuli. From this respect, the reduction technique might inform us what the common input is of the neurons in an IT column. Two other methods have also been successful in reducing the variability between IT neurons. First, recent studies have focused on LFPs, which might be related more to synaptic activity than to spiking output. Second, measurements of multiunit activity (instead of single-unit activity) average out the variability across neurons in the local neuronal population. The latter two methods have recently provided evidence for clustering over a distance <500 µm for multunit activity and over a distance of >1 mm for LFPs (Kreiman et al., 2006). Thus, there appear to be multiple scales and strengths of organization depending on the methodology used and signal that is measured.

**Optical imaging in monkeys**

Activation in neural tissues elicits changes in optical properties of the tissue. Optical imaging, more specifically optical intrinsic signal imaging (OISI), is a technique to visualize these changes (intrinsic signals) from an exposed cortical surface with CCD cameras. Typically, a wavelength of light ~610 nm is used for measurements of the intrinsic signals in IT cortex. The major source of the changes at this wavelength is the absorption changes reflecting increase of deoxyhemoglobin in capillaries due to increased neural activity. Other sources involved in OISI, such as an increase in absorption due to increases in blood volume in capillaries or changes in light scattering due to microstructural changes in neural tissues, do not dominate at this wavelength.

The presentation of any visual stimulus induces increases of intrinsic signals over an area of several millimeters (global signals), but local signals that spread only for ~0.5 mm are detected.
when the intrinsic signals for each stimulus condition is di-vided by the mean of intrinsic signals elicited by many stimuli (Wang et al., 1996, 1998) or by removing the global signals with spatial filtering (Tsunoda et al., 2001). The stimulus specific local signals (activity spots) are considered as supporting evidence for the existence of feature columns in IT cortex. Correlation between spiking activity of single cells and spot activity was examined in two ways. First, OISI with the critical stimulus features that were predetermined from spiking activity of single neurons revealed optical activity spots at the location where the single neurons were recorded (Wang et al., 1996, 1998). Second, selectivity of neural firing in activity spots was well correlated with selectivity of local optical signals at the spots (Tsunoda et al., 2001). Thus, spiking activity of neurons is well correlated in stimulus selectivity and in spatial localization with the local optical signals. It should be pointed out that, because intrinsic signals reflect population (spiking) activity of neurons, the best correlation between the local signals and neuronal firing is obtained at the level of MUA rather than single-unit activity (SUA).

Optical imaging with object images revealed that each object image is represented by the combined activity of multiple activity spots, and each spot represents a visual feature of the complex object image (Tsunoda et al., 2001). Finally, optical imaging with faces revealed that different viewing angles of a face were continuously and systematically mapped on the IT cortex (Wang et al., 1998). This might be part of a general tendency that similar visual features are mapped in nearby locations in IT cortex (Tanaka, 2003).

**Human and monkey fMRI**

Most fMRI studies in humans are based on an intrinsic signal, the BOLD contrast, which originates in a mismatch between the blood flow and oxygen extraction of the tissue. This mismatch leads to a net decrease in the concentration of deoxyhemoglobin in veins and capillaries around neuronal activity, which locally increases the magnetic resonance (MR) signal because deoxyhemoglobin acts as an MR contrast agent. In this respect, BOLD fMRI exhibits a striking contrast to OISI. OISI is sensitive to the early increase of deoxyhemoglobin in capillaries, whereas BOLD fMRI is more sensitive to the following oversupply of oxyhemoglobin (causing a decrease in the deoxyhemoglobin concentration) in capillaries and veins.

Even though in many human fMRI studies voxel sizes of 3 mm or larger are used, which are much larger than the size of the feature columns identified in monkey IT with invasive techniques, such studies have revealed surprisingly strong selectivity for comparisons between object classes such as faces versus objects and scenes (Kanwisher et al., 1997), (headless) bodies versus other object categories (Downing et al., 2001), and scenes/buildings versus objects and faces (Epstein and Kanwisher, 1998; Ishai et al., 1999). Similarly, in monkeys BOLD and contrast agent-dependent cerebral blood volume (CBV) images acquired at a resolution of 1.25 mm revealed patches in monkey IT cortex that responded more strongly to faces or body parts than to objects (Tsao et al., 2003; Pink et al., 2005; Op de Beeck et al., 2008b). These findings are exciting, not the least because they were unexpected from previous electrophysiology experiments that had not shown large (>1 mm) cortical regions with a strong preference for specific object categories and with a consistent location across subjects.

Recent studies have used high-resolution fMRI (HR-fMRI) as the ultimate solution to better link fMRI to high-resolution opti-

cical imaging and electrophysiology data. Data suggest that regions that look like one large homogeneous blob (~1 cm diameter) at standard fMRI resolution can be distinguished into separate regions (Schwarzlose et al., 2005) or even more heterogeneous clusters of voxels (Grill-Spector et al., 2006) at higher fMRI resolution. The difficulty of high-resolution fMRI, however, is that using smaller voxels decreases the signal-to-noise ratio (SNR), and increasing the resolution comes at the risk that (weak) activation may be missed. SNR improvement can be gained via parallel imaging, using high field scanners, or both. For parallel imaging, the sensitivity is improved by using multiple smaller radiofrequency coils (RF) coils. High magnetic fields increase the overall signal-to-noise, and the BOLD signal is also larger at high field. This makes the application of spin-echo (SE) sequences more practicable at high field (Goense et al., 2008). SE sequences have two advantages, compared with the commonly used gradient echo EPI (GE) sequence. SE sequences are less susceptible to signal loss from susceptibility artifacts from the air cavity that plague fMRI of the temporal lobe. Most importantly, SE sequences are thought to provide more localized activations from the capillary bed rather than from larger veins. For example, although ocular dominance columns and orientation columns in cat or human V1 are often difficult to discern with GE imaging, other fMRI methods (SE-BOLD imaging as well as cerebral blood volume/flow imaging) have higher specificity and allow for submillimeter resolution (Duong et al., 2003; Kim et al., 2004; Zhao et al., 2005; Yacoub et al., 2008). Nevertheless, it is yet unknown what is the maximal possible spatial resolution with HR-fMRI. This HR-fMRI limit depends on several factors: (1) signal-to-noise limitations, (2) the point spread function of the BOLD signal (which differs between SE and GE imaging), (3) the spatial scale of the neural clustering, and (4) the spatial scale of the separation between clusters having similar properties.

Many important functional questions remain unresolved given the currently available fMRI evidence. What factors underlie the large-scale selectivity to specific categories (e.g., faces and headless bodies) observed with fMRI? Is the selectivity stronger for these categories than for other categories? Is the spatial scale of representation (clustering of neurons with similar properties) larger and thus more easily measured with fMRI? Does processing of these categories rely on different computations? Is category selectivity related to other characteristics of functional organization? How do these category-selective regions come about? fMRI studies have addressed some of these questions. First, it has been suggested that the strong selectivity for stimuli like faces might be related to their form or shape (Haxby et al., 2000), to specific processing related to faces (i.e., face recognition requires recognition at the exemplar level more than objects) (Gauthier, 2000), to semantic attributes in humans (Chao et al., 1999), and/or to eccentricity biases (Hasson et al., 2002) (regions that prefer faces also have a preference for foveal vs peripheral representations). Furthermore, selectivity in the ventral stream takes more than a decade to reach the adult-like state (Golarai et al., 2007). However, it is yet unclear how the strong selectivity for some object classes is related to these properties or their combination (Op de Beeck et al., 2008a). At the very least, a recent electrophysiological study suggests that the object category structure encoded by the firing rates of IT neurons corresponds to a level of visual form beyond a randomly selected set of moderately complex features (Kiani et al., 2007).
Correspondence between fMRI, optical imaging, and electrophysiology

Previous work in other areas of the brain, most notably primary visual cortex, has focused on the relationship between fMRI, optical imaging, SUA and MUA, and LFPs. fMRI, single-neuron activity, and LFPs tend to correspond rather well in typical situations (Logothetis et al., 2001; Mukamel et al., 2005). This is possibly not that surprising, given that the input to a cluster of neurons is a significant determinant of their output. However, a few studies have been able to disentangle input and output of V1 neurons by manipulating the temporal characteristics of the stimuli. Synthetic activity and LFPs are less sensitive to adaptation (more sustained response to constant stimulation) and the frequency of flicker (nonbaseline response to rapidly flickering stimulus) compared with the spiking output of V1 neurons. In these cases, the fMRI signal seems to correlate more with LFPs than with the spiking output (Logothetis et al., 2001; Lauritzen, 2005; Viswanathan and Freeman, 2007). Dissociation between fMRI/LFP and spiking activity can also be induced with pharmacological manipulations that specifically affect the spiking output (Rauch et al., 2008). However, the correlation between spikes and LFP may depend on the degree of intercorrelations between the firing of neurons, and there is still some controversy about the exact relationship between these signals (Nir et al., 2008; Viswanathan and Freeman, 2008).

Here, we are mainly interested in the correspondence between these signals across spatial locations. A first point of consideration is the resolution and spatial spread of each technique. MUA and SUA sample the output of neurons, whereas LFP is a weighted signal dominated by dendritic input (Logothetis, 2003). MUA and SUA have highest spatial resolution, whereas LFPs are correlated across larger distances, often larger than feature columns or ocular dominance or orientation columns in V1. In V1, the spatial spread of LFP signals is of the order of a few millimeters (Juergens et al., 1999; Goense and Logothetis, 2008), whereas in IT the LFP spans a larger region, ~5–8 mm (Kreiman et al., 2006). In IT, LFPs were correlated across sites with a much larger distance than multiunit activity, and in addition LFP selectivity could not be predicted very well from multiunit activity averaged across sites with varying diameter (Kreiman et al., 2006). This confirms that LFP represents also in IT a different signal than the MUA. Because the spatial extent of LFP is apparently larger than the spatial specificity of the local optical imaging signals, it is unclear how the specificity of the LFP relates to the specificity of the hemodynamic signals (optical imaging and fMRI).

Because optical imaging and fMRI both measure the hemodynamic response induced by neural activity, the two techniques should show similar results, as indeed they do in V1, in which correspondence was found between orientation columns measured by optical imaging and contrast agent-enhanced fMRI (Fukuda et al., 2006). This demonstration involved very high-field scanning (9.4 T) and contrast agent-enhanced CBV-weighted fMRI in cats, as well as analysis techniques that allowed removal of potentially existing global nonselective signals. As mentioned above, GE-BOLD fMRI might lack sufficient spatial specificity to show a similar correspondence. Similar experiments are not straightforward in primate IT, because typically the effective fMRI resolution is lower (~2 mm) and the spots seen in optical imaging are not larger than in cat V1. At current fMRI resolutions, we expect to get a weighted average of the spots/feature columns within a voxel, which may fall apart as the resolution increases. The reported fMRI selectivity for unfamiliar shapes in monkey IT cortex (Op de Beeck et al., 2008b) might reflect such subsampling of very selective feature columns, but higher-resolution scanning is needed to test this hypothesis.

Many unknowns exist about the degree of correspondence between the invasive techniques and fMRI, but some correspondence is definitely present. For example, in monkeys it has been shown that fMRI-identified face-selective patches in the temporal lobe (six per hemisphere) contain a striking majority of neurons that are face-selective (Tsao et al., 2006). Functional connectivity experiments have already revealed the connectivity of the six face patches (Moeller et al., 2008), and electrophysiological experiments might pinpoint the unique step in face recognition performed by each of the different face patches.

However, many questions remain. First of all, faces may be a special category of stimuli, and we can question how these findings translate to other object categories. Furthermore, the properties of the fMRI-identified face patches have not been exhaustively compared with the properties of patches and columns as defined electrophysiologically through a dense sampling of single-unit activity both inside and outside the fMRI-identified face patches. Although the currently available evidence is very suggestive that face-selective patches have different properties than feature columns, a definitive answer awaits a better understanding of the functional and anatomical properties of both patches and columns. Finally, it is nontrivial to compare results across species. Homology of regions across species beyond early sensory areas is complex and often debatable (Kaas, 2008); the spatial scale of structures may differ across species (Adams et al., 2007). Furthermore, performing equivalent experiments across species is difficult even with fMRI. For example, should the comparison of face-selective responses across species use the same stimuli (e.g., human faces) or conspecific stimuli (human faces for humans and monkey faces for macaques)?

Conclusion

We are far from a complete understanding of the fine-scale and large-scale spatial organization of the cortical regions important for object and face recognition. Nevertheless, we have already experienced the power of a combined application of multiple neuroscientific techniques that measure functional organization at different spatial scales.

Ultimately, we apply these methods in the temporal lobe to understand how the brain recognizes and categorizes objects. The spatial organization of object representations is only one part of this general question, and it is the part for which the combined application of these neuroscientific techniques is most useful. However, neighboring single neurons often have clearly different selectivity properties and functions. Such differences are necessarily confounded, and underestimated, by techniques that pool signals over larger numbers of neurons, such as optical imaging and fMRI. Thus, the “gold standard” for understanding neural mechanisms remains single-unit electrophysiology. Nevertheless, the coarser-scale techniques can reveal large-scale patterns of organization that would otherwise go undetected, thereby helping single-unit electrophysiologists to target their electrodes better. In addition, the noninvasive techniques allow comparison between human and nonhuman brains, thereby providing a strong test of human–monkey correspondences and even allowing human studies to lead targeted electrophysiological experiments.

There are several outstanding questions that will only be solved by the combined application of all these methods, direct in the same study or indirect in different studies/laboratories. These
questions center on the relationships between the different methods, the factors that underlie spatial organization in the temporal lobe at multiple scales, and how this multiple-scale organization emerges in the light of plasticity during development and adulthood. At a very basic level, we still do not know why we see clustering for particular functional properties (e.g., does the clustering of face cells in patches contribute to the speed and efficiency of face recognition?). Answering questions like this one will bring us closer to a full understanding of how the spatial organization of the temporal lobe relates to its functional role in visual object and face recognition.

References


Op de Beeck et al. • Spatial Organization in the Temporal Lobe


