5 Understanding Visual Representation by Developing Receptive-Field Models

Kendrick N. Kay

Summary

To study representation in the visual system, researchers typically adopt one of two approaches. The first approach is *tuning curve measurement*, in which the researcher selects a stimulus dimension and then measures responses to specialized stimuli that vary along that dimension. Stimulus dimensions can range from low-level dimensions, such as contrast, to high-level dimensions, such as object category. The second approach is *multivariate pattern classification*, in which the researcher collects the same type of data as in the tuning-curve approach but uses these data to train a statistical classifier that attempts to predict the dimension of interest from measured responses. This approach has recently become quite popular in functional magnetic resonance imaging (fMRI).

In this chapter, we argue that the tuning curve and classification approaches suffer from two critical problems: first, these approaches presuppose that individual stimulus dimensions can be cleanly isolated from one another, but careful consideration of stimulus statistics reveals that isolation is in fact quite difficult to achieve; second, these approaches provide no means for generalizing results to other types of stimulus. We then describe *receptive-field estimation*, an alternative approach that addresses these problems. In receptive-field estimation, the researcher measures responses to a large number of stimuli drawn from a general stimulus class and then develops receptive-field models that describe how arbitrary stimuli are mapped onto responses. Although receptive-field estimation is traditionally associated with electrophysiology, we review recent work of ours demonstrating the application of this technique to fMRI of primary visual cortex. The success of our approach suggests that receptive-field estimation may be a promising direction for future fMRI studies.

Conventional Approaches for Studying Visual Representation

What Is the Goal in Studying Visual Representation?

The primate visual system is composed of several dozen distinct areas, each of which plays a unique role in the processing of visual input. The standard way to characterize the role played by a given visual area is to detail the properties, or dimensions, of the stimulus that modulate activity in that area (Van Essen and Gallant, 1994). For example, it is well established that activity in primary visual cortex is modulated by simple low-level stimulus dimensions such as orientation and spatial frequency (Lennie and Movshon, 2005). In contrast, activity in inferior temporal cortex is thought to be modulated by complex dimensions that are far removed from the raw visual input, such as object category and object position (Op de Beeck, Haushofer, and Kanwisher, 2008). We stipulate that the goal in studying visual representation is to determine what stimulus dimensions modulate activity in each visual area. (Most researchers would probably accept this definition.)

The Tuning-Curve Measurement Approach

The simplest and most common approach for studying visual representation is tuning curve measurement (figure 5.1a). This approach has its roots in classic electrophysiological studies (Hubel and Wiesel, 1959; Campbell, Cooper, and Enroth-Cugell, 1969) and is often used in functional magnetic resonance imaging (fMRI) (Wandell, 1999; Grill-Spector and Malach, 2004). In the tuning-curve approach, the researcher first selects a stimulus dimension believed to be relevant to a given brain area. The researcher then designs specialized stimuli that vary along the dimension of interest and measures responses to these stimuli. Finally, the researcher builds a tuning curve model that links different points along the dimension of interest to responses from each unit (e.g., neuron, voxel, or region-of-interest). The main objective of the tuning-curve approach is to demonstrate that responses in a given brain area are modulated by the dimension of interest.

The tuning-curve approach covers a wide range of studies (figure 5.2). For example, consider an fMRI study in which voxel responses are averaged across a region-of-interest and then two or more experimental conditions are contrasted, such as faces versus houses (Epstein and Kanwisher, 1998; Ishai et al., 1999). This type of study implicitly uses a simple tuning curve model that assigns a separate value to each point along the dimension of interest (for example, a value of 5 could be assigned to “face” and a value of 2 could be assigned to “house”). As another example, consider retinotopic mapping studies in which responses of individual voxels to a large number of contrast-defined images are measured (Wandell, Dumoulin, and Brewer, 2007). Some of these studies use relatively sophisticated tuning curve models, such

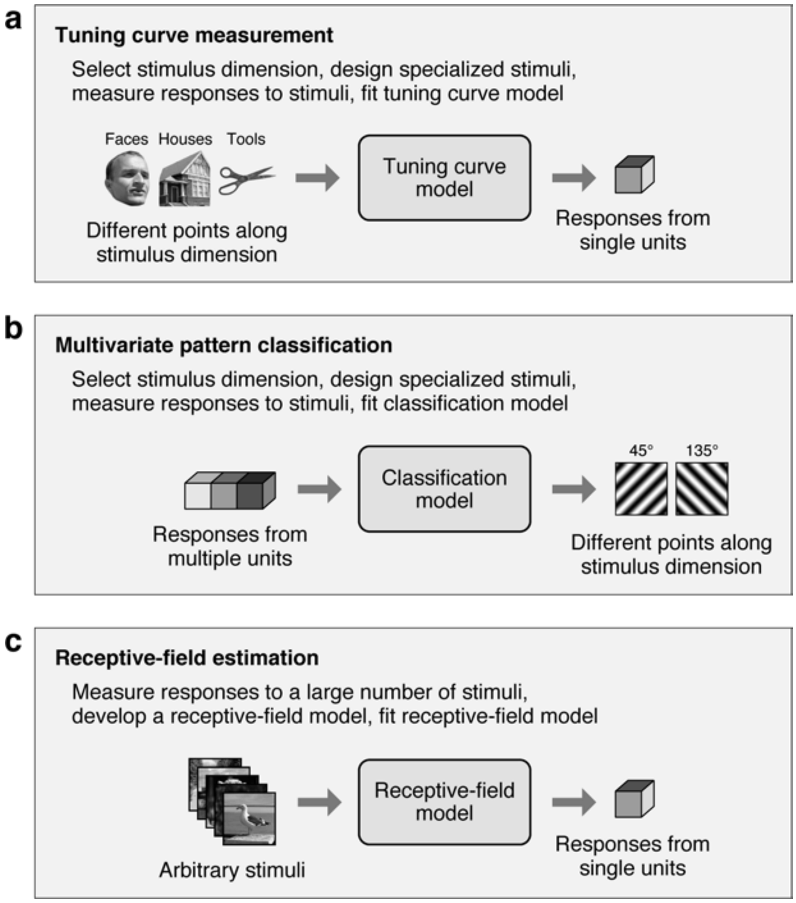
a
Tuning curve measurement
Select stimulus dimension, design specialized stimuli, measure responses to stimuli, fit tuning curve model
Faces Houses Tools
Tuning curve model
Different points along stimulus dimension
Responses from single units
b
Multivariate pattern classification
Select stimulus dimension, design specialized stimuli, measure responses to stimuli, fit classification model
Classification model
45° 135°
Responses from multiple units
Different points along stimulus dimension
c
Receptive-field estimation
Measure responses to a large number of stimuli, develop a receptive-field model, fit receptive-field model
Receptive-field model
Arbitrary stimuli
Responses from single units

Figure 5.1
Different approaches for studying visual representation. (a) Tuning curve measurement. This approach involves measuring responses to stimuli that vary along a specific dimension and then building a tuning curve model that links different points along the dimension of interest to responses from each unit (e.g., neuron, voxel, or region-of-interest). The tuning curve model is usually a simple model that associates a separate value with each point along the dimension of interest, but can be more sophisticated (see figure 5.2). The main objective of tuning curve measurement is to demonstrate that the dimension of interest modulates responses in a given brain area. (b) Multivariate pattern classification. This approach involves measuring responses to stimuli that vary along a specific dimension and then building a classification model that uses responses from multiple units to predict which point along the dimension of interest is present. Like the tuning-curve approach, the classification approach seeks to demonstrate that the dimension of interest modulates responses in a given brain area. However, the classification approach enjoys greater statistical power because responses from multiple units are simultaneously taken into account. (c) Receptive-field estimation. This approach involves measuring responses to a large number of stimuli drawn from a general stimulus class and then building receptive-field models that describe how arbitrary stimuli are mapped onto responses from each unit. Unlike tuning curve models, receptive-field models formalize stimulus dimensions such that the dimensions can be computed for arbitrary stimuli (see figure 5.4). The objective of receptive-field estimation is to develop models that explain as much variance in responses as possible.

Q

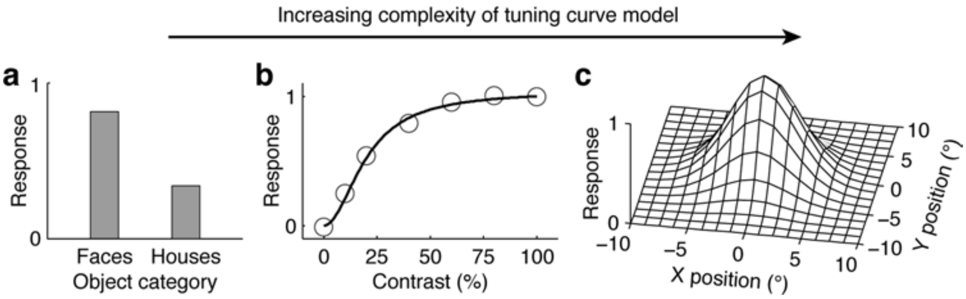

Figure 5.2
Tuning curve models can vary widely in complexity. (a) Model of object category tuning. Suppose we measure responses to objects drawn from two categories, faces and houses. In this case, the dimension of interest is defined on a nominal scale and we can construct a simple tuning curve model that assigns a separate value to each category (Epstein and Kanwisher, 1998; Ishai et al., 1999). (b) Model of contrast tuning. Suppose we measure responses to an image presented at different levels of contrast. In this case, the dimension of interest is defined on a ratio scale and we can construct a slightly more sophisticated tuning curve model that takes a contrast value and passes it through a nonlinear function to generate a predicted response (Albrecht and Hamilton, 1982; Boynton et al., 1999; Carandini and Sengpiel, 2004). (c) Model of spatial tuning. Suppose we measure responses to contrast-defined images that vary in contrast across the visual field (see figure 5.3). In this case, we can construct a sophisticated tuning curve model that takes a spatial pattern of contrast, multiplies this pattern with a two-dimensional Gaussian function, and then sums over the result to generate a predicted response (Larsson and Heeger, 2006; Thirion et al., 2006; Dumoulin and Wandell, 2008).

as a model that takes a spatial pattern of contrast (e.g., a binary image where 0 represents zero contrast and 1 represents full contrast) and filters this pattern with a two-dimensional Gaussian function in order to generate a predicted response (Larsson and Heeger, 2006; Thirion et al., 2006; Dumoulin and Wandell, 2008).

The Multivariate-Pattern Classification Approach

A recently developed approach for studying representation is multivariate pattern classification (figure 5.1b). This approach was initially used in fMRI to investigate the representation of object categories in ventral temporal cortex (Haxby et al., 2001; Cox and Savoy, 2003), but has since been applied to many other types of study, including studies of low-level stimulus dimensions such as orientation (Haynes and Rees, 2005; Kamitani and Tong, 2005) and electrophysiological studies (Hung et al., 2005; Tsao et al., 2006).

The initial steps in multivariate pattern classification are identical to those in tuning curve measurement: the researcher selects a stimulus dimension, designs specialized stimuli that vary along that dimension, and measures responses to these stimuli. However, the classification approach analyzes the resulting data in a different way. In the first stage of the analysis, a subset of the data is used to train a classification model that uses responses from multiple units to predict which point along the dimension of interest is present. For example, one might imagine training a support vector machine that uses responses from a set of 100 voxels to predict which of two grating orientations is present. In the second stage of the analysis, a separate subset of the data is used to evaluate the accuracy of the classification model. Using a separate subset controls for overfitting and ensures an unbiased estimate of accuracy.[1]

Multivariate pattern classification and tuning curve measurement are similar in that both approaches attempt to demonstrate that a dimension of interest modulates responses in a brain area by building a model that relates different points along the dimension of interest to observed responses. However, in the tuning-curve approach, the model is directed from the dimension of interest to the observed responses, whereas in the classification approach, the model is directed from the observed responses to the dimension of interest. Another difference concerns the number of units involved. The tuning-curve approach builds a separate model for each unit, whereas the classification approach builds a single model that incorporates responses from multiple units. The ability to incorporate responses from multiple units provides the classification approach with increased statistical power compared to the tuning-curve approach (Haynes and Rees, 2005; Kamitani and Tong, 2005).

Problems with Conventional Approaches

Although the tuning curve and classification approaches can reveal valuable insight into representation, they face two critical problems. The first is that response modulations presumed to be caused by the dimension of interest could in fact be caused by some other dimension correlated with the dimension of interest. For example, suppose we are interested in the dimension of object category and we measure responses in a given brain area to images of animals, buildings, and tools. If we find selectivity for buildings, can we conclude unequivocally that the brain area is tuned for object category? No, because it is possible that the brain area is actually tuned for some other dimension correlated with object categories. For instance, buildings might have greater power at vertical orientations compared to animals and tools, and the brain area might simply be tuned for vertical orientations.

The usual strategy for dealing with the problem of correlated dimensions is to design stimuli such that unwanted dimensions are controlled for. For example, when designing stimuli that depict objects from different categories, it is typical to equalize the size and position of the objects (for example, Kiani et al., 2007; Kriegeskorte et al., 2008). However, careful consideration of stimulus statistics reveals that it is actually quite difficult to design stimuli that perfectly isolate a single stimulus dimension; rather, it is common for a set of stimuli to vary along multiple dimensions (figure 5.3). Thus, in general, efforts to control stimuli can reduce the severity of the problem of correlated dimensions but cannot completely eliminate the problem.

Q

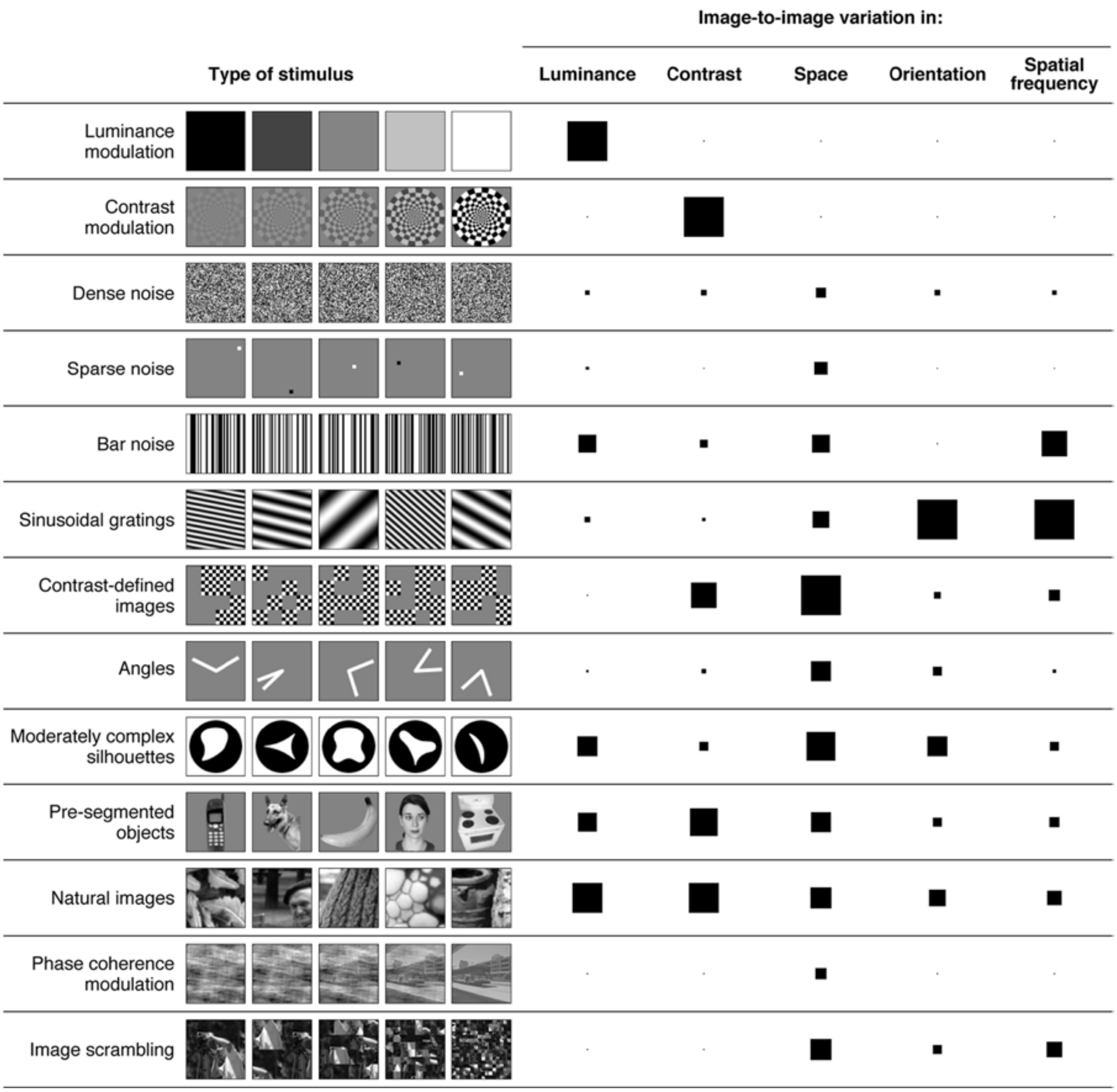

Figure 5.3
Stimuli typically vary along multiple stimulus dimensions. The tuning curve and classification approaches involve designing specialized stimuli that probe specific stimulus dimensions. However, a fundamental problem with this strategy is that in general it is not possible to cleanly separate different stimulus dimensions from one another. Thus, an effect that is presumed to be caused by a certain dimension may actually be caused by other, unconsidered dimensions. To illustrate, in this figure we analyze a variety of stimulus types with respect to several basic dimensions. For each stimulus type, we quantify the amount of image-to-image variation along the dimension of luminance (mean of image pixels), contrast (standard deviation of image pixels), space (standard deviation of image pixels within each element of an 8 × 8 grid), orientation (average spectral power within each of eight orientation bins), and spatial frequency (average spectral power within each of nine spatial frequency bins). (For full details on methods, please see the appendix.) The area of each square indicates the amount of image-to-image variation, and the squares have been scaled such that the maximum square size in each column is the same. The results demonstrate that different stimulus types typically do not isolate single dimensions but instead probe multiple dimensions simultaneously.

The second problem faced by the tuning curve and classification approaches is that these approaches investigate stimulus dimensions without providing a formal description of how to compute the dimensions for arbitrary stimuli. This lack of formalization makes it difficult to take results obtained using one type of stimulus and to generalize them to other types of stimulus. For example, suppose we are interested in the dimension of curvature and we measure responses while parametrically varying the angle formed by two line segments (Pasupathy and Connor, 1999; Hegde and Van Essen, 2000; Ito and Komatsu, 2004). This type of stimulus is convenient because we can simply define curvature as the magnitude of the angle formed by the line segments. However, this definition is specific to stimuli consisting of two line segments, and it is unclear how to generalize results to other types of stimulus.

The Receptive-Field Estimation Approach

What Is a Receptive Field?

The concept of a receptive field was introduced by electrophysiologists in the mid–twentieth century (Hartline, 1938; Kuffler, 1953; Hubel and Wiesel, 1959) and continues to play a central role in our understanding of the visual system. The term "receptive field" is often used to refer to the region of the visual field within which stimuli evoke responses from a given neuron. Other times, the term is used to refer to the specific linear spatiotemporal filter that characterizes the functional behavior of a given neuron (for example, the receptive field of a retinal ganglion cell is approximately a center-surround filter). In both cases the core function of a receptive field is to characterize the circumstances under which a given unit responds to visual stimuli. We therefore propose the following more general definition: a receptive field is any computational model that describes how arbitrary stimuli are transformed into responses from a given unit. Notice that this generalized definition is applicable to any visual area and to any unit of measurement (e.g., neuron, voxel, region-of-interest).

Receptive-field models provide a formal description of how stimulus dimensions are linked to brain responses. For example, consider a receptive-field model that applies a Gabor filter to the stimulus in order to generate a predicted response. This model formalizes the dimensions of orientation, spatial frequency, and contrast such that they can be computed for arbitrary stimuli, and it integrates these dimensions into a single description of how stimuli are mapped onto responses (figure 5.4).

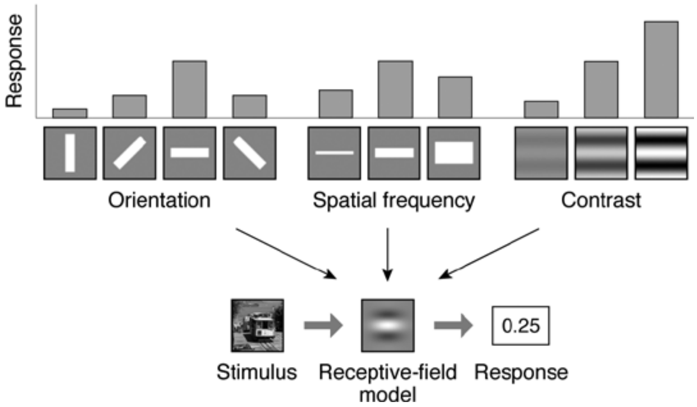

Figure 5.4
Receptive-field models formalize and integrate stimulus dimensions. Suppose we measure tuning curves for the dimensions of orientation, spatial frequency, and contrast. Although these tuning curves provide useful information, it remains unclear how to predict responses to stimuli that differ from those used to measure the tuning curves. Now consider a receptive-field model that applies a Gabor filter to the stimulus in order to generate a predicted response. This simple model performs two vital functions. One, the model formalizes the dimensions of orientation, spatial frequency, and contrast such that they can be computed for arbitrary stimuli. Two, the model integrates the dimensions into a single description of how stimuli are mapped onto responses.

What Is Receptive-Field Estimation?

Receptive-field estimation is an approach to studying visual representation that focuses on developing and testing receptive-field models, and has been used in many electrophysiological studies over the years (see Felsen et al., 2005; Rust et al., 2005; Touryan, Felsen, and Dan, 2005; Bonin, Mante, and Carandini, 2006; David, Hayden, and Gallant, 2006; Nishimoto, Ishida, and Ohzawa, 2006; Rust et al., 2006; Schwartz et al., 2006; Sharpee et al., 2006; Butts et al., 2007; Cadieu et al., 2007; Chen et al., 2007; Mante, Bonin, and Carandini, 2008; Pillow et al., 2008). In essence, receptive-field estimation treats visual representation as a regression problem where the goal is to construct a model that uses stimuli to explain variance in observed responses (Wu, David, and Gallant, 2006).

In receptive-field estimation (figure 5.1c), the researcher first measures responses to a large number of stimuli drawn from a general stimulus class. The researcher then develops one or more receptive-field models and uses a subset of the data to estimate the free parameters of these models. Finally, the researcher uses a separate subset of the data to evaluate the accuracy of the models. Using a separate subset controls for overfitting and ensures that models with different numbers of free parameters can be compared fairly.

The prototypical example of receptive-field estimation is white-noise reverse correlation, a procedure in which white noise is used to drive a neuron and the correlation between each pixel and the response of the neuron is computed (Jones and Palmer, 1987a; Chichilnisky, 2001). This procedure in effect fits a linear receptive-field model in which the predicted response is taken to be a weighted sum of pixels. Note, however, that receptive-field estimation is not limited to linear models nor to simple, mathematically convenient stimuli such as white noise; for example, nonlinear receptive-field models have been developed using responses to complex natural images (Prenger et al., 2004; Touryan, Felsen, and Dan, 2005; David, Hayden, and Gallant, 2006).

Receptive-Field Estimation Addresses Problems with Conventional Approaches

Receptive-field estimation addresses each of the two problems that affect the tuning curve and classification approaches. First, consider the problem of correlated dimensions. In receptive-field estimation, there is no need to construct stimuli that isolate individual stimulus dimensions. Rather, the researcher is free to use stimuli that vary along a variety of dimensions. To decide which of several dimensions best explains responses in a given brain area, the researcher formalizes each dimension into a receptive-field model and finds the model with the highest accuracy. Notice that this strategy is effective even if there exist correlations between dimensions.

Next, consider the problem of generalization. Unlike tuning curve and classification models, receptive-field models formalize stimulus dimensions and provide complete specifications of the mapping between stimulus and response. Thus, receptive-field models are not tied to any particular type of stimulus and can in principle predict responses to arbitrary stimuli. Of course, how well *in practice* a given receptive-field model generalizes to novel stimuli is contingent on the stimuli and the amount of data used to estimate the model and the extent to which the brain area under consideration manifests nonlinearities not captured by the model.

Receptive-Field Estimation Assesses the Relative Importance of Stimulus Dimensions

In the tuning curve and classification approaches, stimuli are specifically designed to emphasize a dimension of interest while minimizing the influence of other dimensions. Thus, even if we find that the dimension of interest substantially modulates responses in a given brain area, we do not gain a sense of how important the dimension is relative to other dimensions. However, the issue of importance can be easily addressed under the approach of receptive-field estimation. Here, stimuli are sampled from a general stimulus class (for example, natural images) and are not tailored for any particular stimulus dimension. We can therefore obtain an unbiased assessment of the importance of a given dimension by simply quantifying the amount of variance in responses that the dimension accounts for.

Challenges in Receptive-Field Estimation

The main challenge in receptive-field estimation is the difficulty of developing new receptive-field models. This difficulty stems from the fact that formalizing stimulus dimensions is not a trivial task: although certain low-level dimensions such as contrast are well understood and can be easily formalized, other dimensions such as object shape are understood only at a conceptual level, and formalization of these dimensions remains a challenging endeavor. To gain ideas for new receptive-field models, it may be useful to examine computational models developed in other fields such as theoretical neuroscience (for example, Olshausen and Field, 1996; Bell and Sejnowski, 1997; Berkes and Wiskott, 2005; Cadieu and Olshausen, 2009; Hyvärinen, Hurri, and Hoyer, 2009; Karklin and Lewicki, 2009) and computer vision (for example, Lowe, 1999; Martin, Fowlkes, and Malik, 2004; Serre et al., 2007; Pinto et al., 2009).

Another difficulty is that only a limited amount of data can be collected in a given experiment, making it difficult to estimate receptive-field models with many free parameters. To compensate for limited data, it is useful to optimize the quality of the data that are in fact collected. This can be accomplished through a variety of means, such as carefully controlling the behavioral and attentional state of the subject; reducing non-neuronal sources of noise such as head motion in fMRI studies; and optimizing in real-time the stimuli used in an experiment (Benda et al., 2007; Yamane et al., 2008; Lewi, Butera, and Paninski, 2009). Another strategy for dealing with data limitations is to incorporate prior knowledge about the brain area under investigation, thereby reducing the amount of information that the data have to convey. This can be accomplished either by reducing the complexity of a model before parameter estimation (for example, restricting a model to a specific region of the visual field) or by using maximum a posteriori methods for parameter estimation (Wu, David, and Gallant, 2006; Paninski, Pillow, and Lewi, 2007).

Application of Receptive-Field Estimation to fMRI

Gabor Wavelet Pyramid Model of Voxels in Primary Visual Cortex

Although receptive-field estimation has been traditionally restricted to electrophysiology, there is no intrinsic reason that this must be the case. Indeed, emerging research indicates the viability of using other measurement techniques such as optical imaging (Baker and Issa, 2005; Mante and Carandini, 2005; Basole et al., 2006) and fMRI (Bartels, Zeki, and Logothetis, 2008; Dumoulin and Wandell, 2008; Kay et al., 2008a; Kriegeskorte et al., 2008; Miyawaki et al., 2008; Naselaris et al., 2009) to develop models of visual representation that are more sophisticated than simple tuning curve or classification models.[2] Here we review recent work of ours demonstrating the application of receptive-field estimation to fMRI (Kay et al., 2008a; see also Naselaris et al., 2009).

Because receptive-field estimation is not a standard approach in fMRI, we started off by targeting a relatively well-understood brain area, primary visual cortex (V1). Electrophysiological studies indicate that there are two major functional classes of neurons in V1, simple cells and complex cells. To a first approximation, a simple cell can be modeled as a single half-wave rectified Gabor filter, and a complex cell can be modeled as the sum of several half-wave rectified Gabor filters (Movshon, Thompson, and Tolhurst, 1978a, 1978c; Daugman, 1980; Adelson and Bergen, 1985; Jones and Palmer, 1987b). We reasoned that if the activity in a V1 voxel reflects the pooled activity of a large number of simple and complex cells, then it should be possible to model a V1 voxel as a population of half-wave rectified Gabor filters (figure 5.5). We term this model the *Gabor model*.[4]

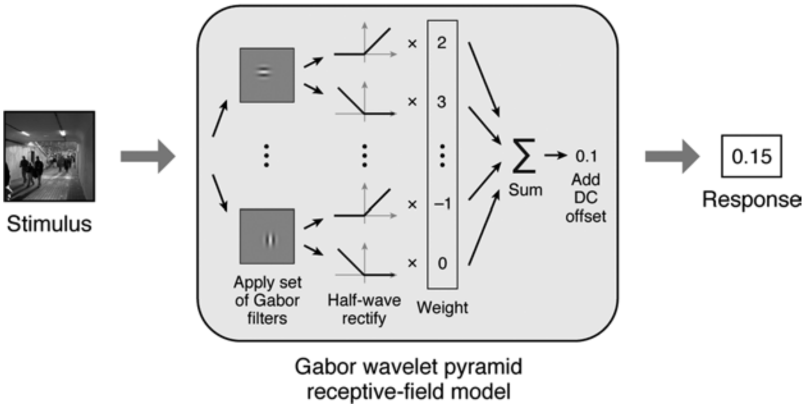
Stimulus
Apply set of Gabor filters
Half-wave rectify
×
2
3
−1
0
Weight
Σ
Sum
0.1
Add DC offset
0.15
Response
Gabor wavelet pyramid receptive-field model

Figure 5.5
Gabor wavelet pyramid receptive-field model. In Kay et al. (2008a), we measured fMRI activity in early visual areas while subjects viewed a large number of grayscale natural images. We then devised a receptive-field model that could potentially characterize the mapping between visual stimuli and voxel responses. In the model, the stimulus image is first filtered with a diverse set of Gabor filters occurring at different positions, orientations, spatial frequencies, and phases. The filter outputs are then half-wave rectified, weighted by a set of free parameters, and summed together.[3] Finally, a DC offset is added, producing the predicted response. This model is based on standard models of V1 neurons (Ringach, 2004; Carandini et al., 2005) and is suitable for characterizing the pooled activity of a large population of V1 neurons.

Accuracy of the Gabor Model

To validate the use of receptive-field estimation in fMRI, we sought to confirm that the Gabor model accurately characterizes voxel responses in V1. To this end we measured fMRI activity (4 T, surface coil, GE-EPI, $2 \times 2 \times 2.5$ mm^3, 1 Hz) in early visual areas while subjects passively viewed a large number of grayscale natural images. For each subject two sets of data were acquired: a training dataset that consisted of 1,750 images presented 2 times each and a validation dataset that consisted of 120 images presented 13 times each. For each voxel, a response timecourse (see Kay et al., 2008b) was estimated and deconvolved from the time-series data, producing an estimate of the amplitude of the response to each distinct image.

We fit the Gabor model to each voxel by applying gradient descent with early stopping to the data in the training dataset (Skouras, Goutis, and Bramson, 1994). (Gradient descent with early stopping imposes a shrinkage prior on model parameters and is an example of a maximum a posteriori method for parameter estimation; see section 2.5.) We then assessed the accuracy of the Gabor model by calculating the amount of variance in the validation dataset that is explained by the model. To obtain a realistic assessment of model accuracy, we expressed this amount as a percentage relative to the amount of variance that a perfect model could in principle explain, given the level of noise in the validation dataset (Sahani and Linden, 2003; David and Gallant, 2005).

We found that in V1 the Gabor model accounts for approximately 70 percent of the explainable variance (figure 5.6). This high value is consistent with our understanding of V1 derived from electrophysiology, and it helps validate the use of receptive-field estimation in fMRI. To gain additional insight into the Gabor model, we also examined results in extrastriate visual areas. Neurons in extrastriate areas are thought to be tuned for features more complex than Gabor-like features (Van Essen and Gallant, 1994; Carandini et al., 2005; Orban, 2008), and we expected that the Gabor model would not perform as well in these areas as it does in V1. Indeed, we found that the accuracy of the Gabor model decreases progressively from V1 to V2 to V3 to V4 (figure 5.6).

Consistency of the Gabor Model with Neuronal Tuning Properties

The Gabor model characterizes a V1 voxel as the sum of a large number of Gabor filters (potentially thousands), each of which represents a population of V1 neurons that share tuning for a particular position, phase, orientation, and spatial frequency (figure 5.7). To determine whether this characterization is accurate, we investigated whether the specific sets of Gabor filters that comprise our V1 voxel models are consistent with existing knowledge of the organization and function of V1 neurons.

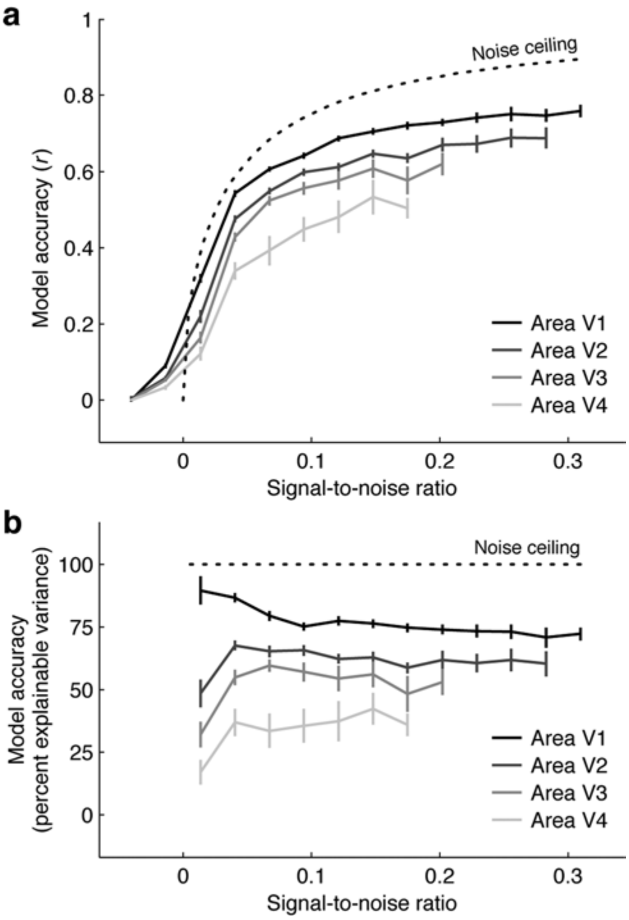

Figure 5.6
Accuracy of the Gabor model. For each voxel, we fit the Gabor model using responses in a training dataset and then assessed how accurately the model predicts responses in a separate validation dataset. (a) Model accuracy as a function of signal-to-noise ratio. In this panel, voxels are binned by signal-to-noise ratio (defined as the ratio between the amount of variance in responses due to the stimulus and the amount of variance in responses due to all other factors). For each bin the median correlation (*r*) between measured and predicted responses is plotted. Error bars indicate ± 1 standard error, and the dotted line indicates the noise ceiling, that is, the theoretical maximum performance that can be achieved given the level of noise in the data. (b) Model accuracy in terms of percent explainable variance. We replot the results shown in panel a, expressing the amount of variance explained by the Gabor model (r^2) as a percentage relative to the amount of variance that a perfect model could in principle explain. In V1, the Gabor model accounts for approximately 70 percent of the explainable variance.

Q

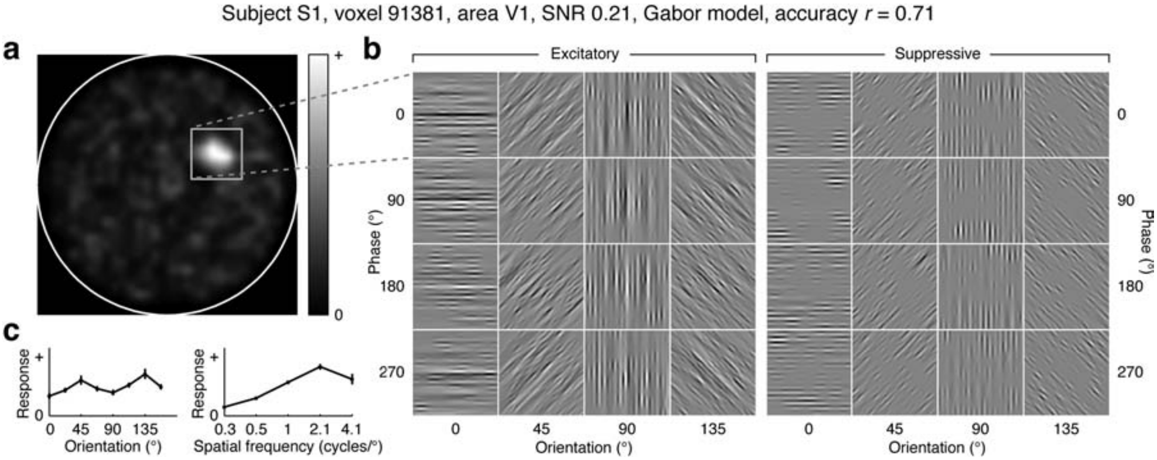

Figure 5.7
Visualization of the receptive field of a representative voxel. (a) Spatial envelope. The receptive field (RF) estimate displayed in this panel was obtained by applying the Gabor model to the full extent of the stimulus (20° × 20°). The intensity of each pixel indicates the sensitivity of the RF to that location in the visual field, the white circle indicates the bounds of the stimulus, and the gray square indicates the estimated RF location. The results show that the RF is spatially localized in the upper-right quadrant of the visual field. (b) Direct visualization of filters. The RF estimate displayed in this panel was obtained by applying the Gabor model to the estimated RF location. Each individual image corresponds to the estimated RF location and depicts filters that have a specific orientation and phase but a variety of positions and spatial frequencies. The root-mean-square intensity of each filter is proportional to the weight associated with that filter. The results show that filters are mainly excitatory and are broadly distributed across orientation, position, and phase. (c) Orientation and spatial frequency tuning curves. To summarize the tuning properties of the RF estimate shown in panel b, orientation and spatial frequency tuning curves were constructed. This was accomplished by computing the predicted response of the RF to sinusoidal gratings varying in orientation and spatial frequency. The results show that selectivity for orientation is somewhat weaker than selectivity for spatial frequency.

We first considered the dimension of space. In V1, nearby neurons are tuned for nearby positions in the visual field, and there exists a large-scale retinotopic mapping of the visual field onto the cortical surface (Van Essen, Newsome, and Maunsell, 1984; Tootell et al., 1988; Wandell, Dumoulin, and Brewer, 2007). Consistent with these observations, we found that the Gabor filters that contribute to a V1 voxel model tend to cluster together (for example, see figure 5.7a) and that the spatial tuning of our V1 voxel models successfully reproduces the retinotopic organization of V1 (see results in Kay et al., 2008a).

Next, we considered the dimension of orientation. Although individual V1 neurons are highly selective for orientation, neurons in V1 are organized such that a full range of orientations is represented over a scale (0.5–1 mm in the macaque; see Hubel and Wiesel, 1974; Blasdel and Salama, 1986) substantially smaller than the size of the voxels in our experiment ($2 \times 2 \times 2.5$ mm^3). Thus, we expect to find

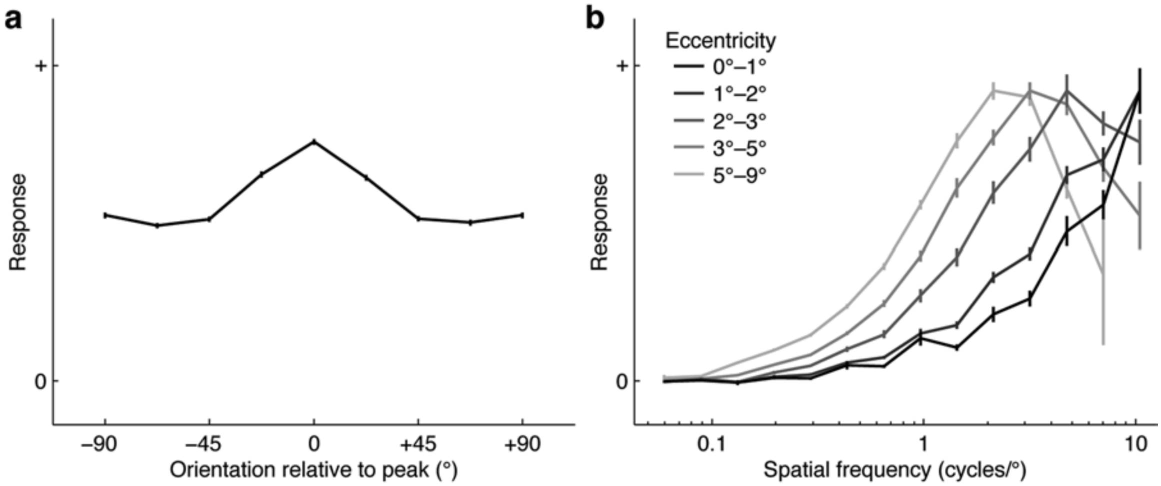

Figure 5.8
Summary of orientation and spatial frequency tuning. We constructed orientation and spatial frequency tuning curves for V1 voxels for which the accuracy (*r*) of the Gabor model was significantly greater than 0 ($p < 0.01$, one-tailed randomization test). (a) Orientation tuning. To summarize results for orientation, we aligned the peaks of the orientation tuning curves and then averaged the tuning curves together. The result is shown, with error bars indicating ± 1 standard error. The fact that the averaged tuning curve is quite broad in shape indicates that voxel orientation tuning is at most a small effect (Haynes and Rees, 2005; Kamitani and Tong, 2005). (b) Spatial frequency tuning. To summarize results for spatial frequency, we grouped voxels according to eccentricity and then averaged the spatial frequency tuning curves of voxels in each group. The resulting tuning curves have been scaled to the same height for display purposes, and error bars indicate ± 1 standard error. Notice that the tuning curves are generally band-pass and that peak spatial frequency decreases as eccentricity increases (Sasaki et al., 2001; Henriksson et al., 2008).

only weak biases in orientation tuning at the voxel level (Haynes and Rees, 2005; Kamitani and Tong, 2005). Orientation tuning curves derived from our voxel models are indeed consistent with this expectation (figure 5.8a).

Finally, we considered the dimension of spatial frequency. Neurons in V1 exhibit band-pass spatial frequency tuning and cover a limited range of spatial frequencies (Schiller, Finlay, and Volman, 1976; Movshon, Thompson, and Tolhurst, 1978b; De Valois, Albrecht, and Thorell, 1982; Foster et al., 1985; Shapley and Lennie, 1985). Thus, even though a V1 voxel contains a wide assortment of neurons, we still expect to find strong band-pass spatial frequency tuning at the voxel level. Furthermore, it is known that neurons in V1 exhibit an overall decrease in preferred spatial frequency as receptive-field eccentricity increases (Schiller, Finlay, and Volman, 1976; De Valois, Albrecht, and Thorell, 1982). Consistent with these several observations, we found that spatial frequency tuning curves derived from our voxel models are generally band-pass and shift toward lower spatial frequencies at peripheral eccentricities (figure 5.8b).

Q

Evaluation of Alternative Models

In order for receptive-field estimation in fMRI to be a useful approach, it must be possible to use fMRI data to discriminate competing receptive-field models. We therefore formulated several alternative models to compare against the Gabor model. Three of the models use the same framework as the Gabor model but involve different types of filters. The *Pixel model* uses individual pixels as filters and thus characterizes the response from a voxel as a weighted sum of half-wave rectified pixel filters. The *Gaussian model* uses two-dimensional Gaussians varying in size and position as filters. The *Fourier model* uses two-dimensional basis functions derived from the discrete Fourier transform as filters (David, Vinje, and Gallant, 2004). The last model that we formulated, the *Energy model*, characterizes the response from a voxel as a weighted sum of the luminance- and contrast-energy of the image (calculated as the half-wave rectified mean and standard deviation of pixel values, respectively). This model is similar to recently proposed models of phase-encoded retinotopic mapping data (Larsson and Heeger, 2006; Thirion et al., 2006; Dumoulin and Wandell, 2008).

We evaluated each of the receptive-field models using the same methods described earlier. To ensure robust model comparison, each model was applied to the specific region of the visual field corresponding to the estimated receptive-field location for each voxel. We observed the following trend in model accuracy for voxels in V1: Pixel < Gaussian < Energy < Fourier < Gabor (figure 5.9). The fact that the Gabor model outperforms alternative models demonstrates that it is possible to use fMRI data to evaluate and discriminate competing receptive-field models. Post-hoc analyses indicate that differences in model accuracy arise primarily from differences in how well each model characterizes voxel spatial frequency tuning (results not shown). This is reasonable, given our earlier observation that voxel spatial frequency tuning is a strong effect (see figure 5.8).

Advantages of Using fMRI for Receptive-Field Estimation

The measurement technique traditionally associated with receptive-field estimation is electrophysiology. What advantages can using fMRI for receptive-field estimation offer? First, fMRI provides simultaneous measurements of activity from multiple brain areas. This enables large datasets to be collected relatively quickly and offers the prospect of using a single dataset to investigate representation in different brain areas. Second, in principle there is no limit to the amount of data that can be collected from a voxel since data can be combined across scan sessions. This is favorable because model accuracy is often limited by the amount of data available for estimation of model parameters. Third, since fMRI is noninvasive, it can be readily applied to human subjects. This could facilitate the investigation of the impact of attention and other cognitive factors on representation.

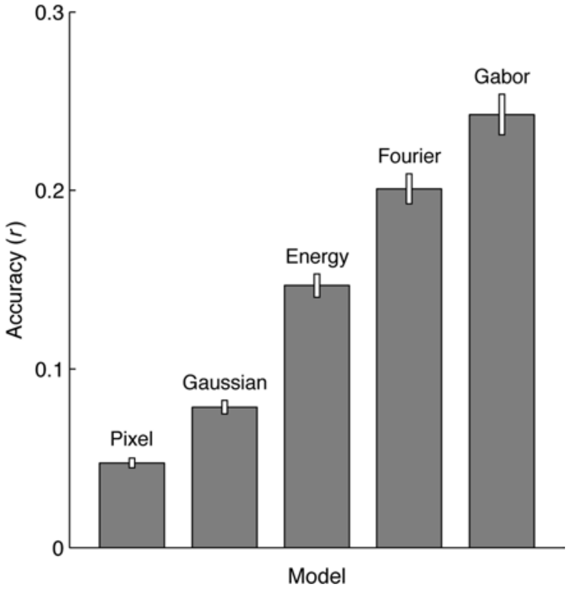

Figure 5.9
Evaluation of alternative models. We formulated several alternative models to compare against the Gabor model. Each model was fit and tested using the same methods used for the Gabor model. In this figure, bar height indicates median accuracy across voxels in V1, and error bars indicate ± 1 standard error. The Gabor model achieves the highest accuracy, consistent with V1 electrophysiology. More generally, these results demonstrate that it is possible to use fMRI data to discriminate competing receptive-field models.

However, fMRI suffers from a critical disadvantage, namely, limited spatial resolution. Despite advances in imaging hardware and techniques, the spatial resolution that can be currently achieved in fMRI while maintaining good coverage and adequate signal-to-noise ratio is relatively low, with voxel sizes on the order of $2 \times 2 \times 2$ mm^3 (at moderate field strength). At this resolution, each voxel pools the activity of hundreds of thousands of neurons, making it difficult to infer functional properties of individual neurons based on fMRI data. Receptive-field models developed in fMRI should therefore be interpreted with respect to what electrophysiology reveals about functional properties at the neuronal level.

The Prospects of Receptive-Field Estimation in Future fMRI Studies

The Case of Ventral Temporal Cortex

Given the feasibility of applying receptive-field estimation to fMRI in V1, we believe that this approach has the potential to improve our understanding of representation throughout the visual system. In this section we speculate on the specific case of

Q

ventral temporal cortex, since this region of the brain and its various subregions (e.g., lateral occipital complex, fusiform face area) are intensely studied by many laboratories.

At first glance, our understanding of ventral temporal cortex seems well developed, since it appears we have already identified object category as the stimulus dimension that primarily modulates responses in this region (Malach, Levy, and Hasson, 2002; Grill-Spector, 2003; Kiani et al., 2007; Op de Beeck, Haushofer, and Kanwisher, 2008). Indeed, current research tends to take for granted the idea that object category is the fundamental stimulus dimension, and instead focuses on the secondary issue of how object categories are topographically organized in the brain (Op de Beeck et al., 2008; Op de Beeck, Haushofer, and Kanwisher, 2008).

However, we contend that our understanding of ventral temporal cortex is in fact quite rudimentary, since object category is a poorly understood stimulus dimension. To illustrate, suppose we construct a tuning curve for contrast by selecting an image, globally scaling the image pixel values to various degrees, and measuring responses to the resulting stimuli. And suppose we construct a tuning curve for object category by selecting different object categories and measuring the average response to objects drawn from each category. Although these two situations are superficially similar, there is a critical difference. In the case of contrast, response modulations can be attributed to a concrete, definitive property of the stimulus (the spread in the distribution of pixel values). But this is not the case for object category, since the critical stimulus property that varies from one category to the next is unknown. Thus, while our understanding of contrast is strong, our understanding of object category is weak.

It is tempting to think that we understand the dimension of object category given the effortlessness with which we, as human observers, recognize objects in our everyday lives. But we must be careful not to confuse this superficial understanding of object category with the in-depth understanding that a formal description of object category would provide. Such a description is exactly what we hope to obtain by applying receptive-field estimation to ventral temporal cortex.

Developing Receptive-Field Models for Ventral Temporal Cortex

There are several approaches that could be used to develop receptive-field models for ventral temporal cortex. One approach is to take existing computational models of object recognition and adapt these models such that they can be fit to responses measured from the brain (for an example, see Cadieu et al., 2007). In this respect, receptive-field estimation can be viewed as a method for incorporating theoretical models into an experimental setting.

A second approach is to start with a high-level theory of visual processing and then attempt to translate the theory into a concrete receptive-field model. For example, selectivity for object category has been hypothesized to reflect semantic properties of objects (Chao, Haxby, and Martin, 1999), specialized processing for certain object categories such as faces (Kanwisher, 2000), form and shape characteristics associated with different object categories (Haxby et al., 2000; Tanaka, 2003), the level at which objects from a given category are processed (Gauthier, 2000; Tarr and Gauthier, 2000), and the eccentricity at which objects from a given category are typically viewed (Malach, Levy, and Hasson, 2002). Translating these theories into receptive-field models and testing the resulting models would be an extremely valuable enterprise.

A final, bottom-up approach for developing receptive-field models is to scrutinize what is already known regarding ventral temporal cortex. For example, studies investigating the dimension of object category typically use single, pre-segmented objects (Haxby et al., 2001; Cox and Savoy, 2003; Hung et al., 2005; Kiani et al., 2007; Kriegeskorte et al., 2008); this simplified setup neglects complexity inherent in real-world natural scenes such as background clutter, multiple objects, and partially occluded objects. Specifying how the dimension of object category can be computed for complex natural scenes would be a useful step toward the development of receptive-field models. As another example, it is known that in addition to object category, object position also modulates responses in ventral temporal cortex (Levy et al., 2001; DiCarlo and Maunsell, 2003; MacEvoy and Epstein, 2007; Sayres and Grill-Spector, 2008; Schwarzlose et al., 2008). Thus, a useful starting point for developing receptive-field models would be to brainstorm potential computational mechanisms that can simultaneously describe tuning along these two dimensions.

Final Thoughts

To be clear, we do not mean to imply that it will be easy to build receptive-field models that accurately characterize responses in ventral temporal cortex, or any other visual area for that matter. Indeed, a major advantage of conventional approaches such as tuning curve measurement is that these approaches are relatively straightforward to carry out and invariably provide some insight into the computations performed by a given area. Nevertheless, we contend that our understanding of visual representation remains fundamentally limited until we develop and test receptive-field models for the various visual areas in the brain.

Acknowledgments

This work was supported by an NDSEG fellowship, the NIH, and UC-Berkeley intramural funds. We thank R. Kiani and N. Kriegeskorte for providing stimuli used in their research. We also thank C. Cadieu, S. David, J. Gallant, K. Gustavsen,

Q

N. Kriegeskorte, T. Naselaris, S. Nishimoto, M. Oliver, B. Pasley, R. Prenger, M. Silver, A. Vu, and J. Winawer for comments on the manuscript.

Appendix: Calculation of Stimulus Statistics for Different Types of Stimulus

In figure 5.3, we depict the amount of image-to-image variation along several stimulus dimensions for a variety of stimulus types. Here we describe the methods used to obtain these results.

Stimuli were prepared as 64 × 64 grayscale images with pixel values in the range 0 (black) to 1 (white). Five hundred samples of each stimulus type were generated, unless otherwise indicated.

- *Luminance modulation* (Rossi, Rittenhouse, and Paradiso, 1996; Kinoshita and Komatsu, 2001; Haynes, Lotto, and Rees, 2004; Peng and Van Essen, 2005; Cornelissen et al., 2006) consisted of a uniform image whose luminance was varied from black to white in 100 equally spaced increments.
- *Contrast modulation* (Albrecht and Hamilton, 1982; Boynton et al., 1999; Avidan et al., 2002; Carandini and Sengpiel, 2004; Kastner et al., 2004) consisted of a radial checkerboard pattern whose contrast was varied from 1 percent to 100 percent in 100 equally spaced increments.
- *Dense noise* (Victor et al., 1994; Reid, Victor, and Shapley, 1997; Chichilnisky, 2001; Olman et al., 2004; Nishimoto, Ishida, and Ohzawa, 2006) was generated by drawing pixel values randomly from a uniform distribution.
- *Sparse noise* (Jones and Palmer, 1987a; DeAngelis, Ohzawa, and Freeman, 1993) was generated by setting a randomly chosen element of a 16 × 16 grid to black or white and setting the other elements to neutral gray.
- *Bar noise* (Lau, Stanley, and Dan, 2002; Touryan, Lau, and Dan, 2002; Rust et al., 2005) consisted of vertical bars (one-pixel wide) whose luminance values were randomly set to black or white.
- *Sinusoidal gratings* (Geisler and Albrecht, 1997; Singh, Smith, and Greenlee, 2000; Albrecht et al., 2002; Mazer et al., 2002; Ringach, 2002) were constructed at full contrast and had randomly chosen orientations, spatial frequencies (in the range 1 to 25 cycles per image), and phases.
- *Contrast-defined images* (Thirion et al., 2006; Miyawaki et al., 2008) consisted of a 4 × 4 grid where each element was randomly set to neutral gray (zero contrast) or filled with an underlying checkerboard pattern (full contrast). The underlying checkerboard pattern consisted of alternating black and white squares defined on a 16 × 16 grid.

- *Angles* (Pasupathy and Connor, 1999; Hegde and Van Essen, 2000; Ito and Komatsu, 2004) consisted of two white line segments placed on a neutral-gray background. Each line segment emanated from the center of the image at a random angle, and had a width of 4 pixels and a length of 29 pixels.
- *Moderately complex silhouettes* (Pasupathy and Connor, 2001, 2002; Brincat and Connor, 2004) were prepared by rendering the 366 images depicted in figure 2 of Pasupathy (2006) at full contrast.
- *Pre-segmented objects* (Haxby et al., 2001; Cox and Savoy, 2003; Hung et al., 2005; Kiani et al., 2007; Kriegeskorte et al., 2008) were prepared by downsampling the 92 images used by Kriegeskorte et al. (2008) and then converting these images to grayscale.
- *Natural images* (Rainer et al., 2001; Smyth et al., 2003; Weliky et al., 2003; David, Vinje, and Gallant, 2004; Olman et al., 2004) consisted of image patches randomly extracted from the photographs used in Kay et al. (2008a). Each image patch was scaled such that pixel values spanned the range 0 to 1.
- *Phase coherence modulation* (Rainer et al., 2001; Dakin et al., 2002; Kayser et al., 2003; Tjan, Lestou, and Kourtzi, 2006; Perna et al., 2008) consisted of a single natural image whose phase spectrum was blended with a random phase spectrum (excluding the DC component). The amount of blending varied from 0 percent to 100 percent in 100 equally spaced increments, and blending was performed linearly with respect to phase angle. After blending, the entire image ensemble was scaled such that pixel values spanned the range 0 to 1.
- *Image scrambling* (Kanwisher, McDermott, and Chun, 1997; Lerner et al., 2001; Rainer et al., 2002; Tsao et al., 2006) consisted of a single natural image that was subjected to various degrees of scrambling. Scrambling was performed by partitioning the image according to a 1 × 1, 2 × 2, 4 × 4, 8 × 8, or 16 × 16 grid and then randomly shuffling the resulting image segments.

Images were quantified with respect to the dimensions of luminance, contrast, space, orientation, and spatial frequency. Luminance and contrast were quantified by computing the mean and standard deviation of image pixels, respectively. For space, orientation, and spatial frequency, the procedure was slightly more complicated. In order to ensure that variations in space, orientation, and spatial frequency do not simply reflect changes in overall image contrast, the images associated with each stimulus type were scaled such that the contrast of each image matched the average contrast of the original, unscaled images. Then, after this contrast-normalization procedure, the dimension of space was quantified by partitioning each image according to an 8 × 8 grid and then computing the standard deviation of image pixels in each of the resulting image segments. The dimensions of

Q

orientation and spatial frequency were quantified by calculating the power spectrum of each image and then computing the mean power in each of eight orientation bins (centered at 0°, 22.5°, ... , and 157.5°) and each of nine spatial frequency bins (1–6, 6–11, ... , and 41–46 cycles per image).

For each stimulus type the amount of image-to-image variation with respect to each of the various dimensions was calculated. This was accomplished by interpreting the quantification of a given dimension as defining a metric space and then computing the average Euclidean distance between pairs of images randomly selected from the given stimulus type. For example, suppose we wish to calculate the amount of image-to-image variation in orientation for natural images. To do this we first quantify orientation for each natural image; this in effect produces a cloud of points residing in an eight-dimensional space. We then compute the average Euclidean distance between pairs of points randomly selected from this cloud.

Notes

[1] We have termed the approach *multivariate pattern classification*, since predicting discrete classes is most common in the literature. However, whether discrete or continuous quantities are predicted is not critical, and our treatment of multivariate pattern classification applies just as well to the case where continuous quantities are predicted (such a case could be termed *multivariate pattern regression*).

[2] Some of these studies involve approaches that are either identical to or closely related to receptive-field estimation; however, not all of the studies can be characterized in that way. A full description of the studies and how they relate to the three basic approaches of tuning curve measurement, multivariate pattern classification, and receptive-field estimation is outside the scope of this paper, but we briefly describe here one notable study (Kriegeskorte et al., 2008). In this study, responses to an assortment of real-world objects were measured and then multivariate dimensionality-reduction techniques (see also Gallant et al., 1996; Op de Beeck et al., 2001; Hegde and Van Essen, 2007; Kiani et al., 2007; Brouwer and Heeger, 2009) were used to visualize and discover the stimulus dimensions important to the various brain areas under consideration. The study also evaluated how well various receptive-field models accounted for the observed results. Receptive-field models were not evaluated with respect to how well they characterize responses from individual brain units (as we propose in this paper), but were instead evaluated with respect to how well they reproduce the similarity structure of the objects (similarity was computed by correlating response patterns obtained for different objects).

[3] Although the model described here uses half-wave rectified Gabor filters, the model in the published study (Kay et al., 2008a) involves computing the square root of the sum of the squares of quadrature-phase Gabor filters. Nevertheless, these two models yield very similar results, and we adopt the former model in order to simplify the presentation.

[4] There are two caveats to our proposed interpretation of the Gabor model. The first caveat is that standard models of V1 neurons are based on the spiking behavior of neurons whereas the blood oxygenation level dependent (BOLD) signal measured in fMRI is coupled to synaptic activity, not spiking activity per se (Lauritzen, 2001; Heeger and Ress, 2002; Bartels et al., 2008; Logothetis, 2008). However, spiking activity is likely to be highly correlated with synaptic activity in the case of simple sensory stimulation (Scannell and Young, 1999; Heeger and Ress, 2002; Kim et al., 2004). It is therefore reasonable to assume that the same stimulus properties that drive spiking activity also drive synaptic activity. The second caveat is that the relationship between neural activity and the strength of the subsequent BOLD response may not be entirely linear (Heeger and Ress, 2002; Logothetis and Wandell, 2004; Lauritzen, 2005). However, nonlinearity does not invalidate the basic interpretation of the Gabor model: under certain reasonable assumptions, a nonlinear relationship between neural activity and the

BOLD response can be incorporated into the Gabor model by simply applying a nonlinear transformation to the output of each filter in the model. Preliminary results indicate that applying a compressive exponent (e.g., 0.5) to filter outputs leads to an increase in the accuracy of the Gabor model for V1 voxels. This is consistent with the existence of a compressive relationship between neural activity and the BOLD response (Logothetis et al., 2001; Logothetis and Wandell, 2004).

References

Adelson EH, Bergen JR. 1985. Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A* 2: 284–299.

Albrecht DG, Geisler WS, Frazor RA, Crane AM. 2002. Visual cortex neurons of monkeys and cats: temporal dynamics of the contrast response function. *J Neurophysiol* 88: 888–913.

Albrecht DG, Hamilton DB. 1982. Striate cortex of monkey and cat: contrast response function. *J Neurophysiol* 48: 217–237.

Avidan G, Harel M, Hendler T, Ben-Bashat D, Zohary E, Malach R. 2002. Contrast sensitivity in human visual areas and its relationship to object recognition. *J Neurophysiol* 87: 3102–3116.

Baker TI, Issa NP. 2005. Cortical maps of separable tuning properties predict population responses to complex visual stimuli. *J Neurophysiol* 94: 775–787.

Bartels A, Logothetis NK, Moutoussis K. 2008. fMRI and its interpretations: an illustration on directional selectivity in area V5/MT. *Trends Neurosci* 31: 444–453.

Bartels A, Zeki S, Logothetis NK. 2008. Natural vision reveals regional specialization to local motion and to contrast-invariant, global flow in the human brain. *Cereb Cortex* 18: 705–717.

Basole A, Kreft-Kerekes V, White LE, Fitzpatrick D. 2006. Cortical cartography revisited: a frequency perspective on the functional architecture of visual cortex. *Prog Brain Res* 154: 121–134.

Bell AJ, Sejnowski TJ. 1997. The "independent components" of natural scenes are edge filters. *Vision Res* 37: 3327–3338.

Benda J, Gollisch T, Machens CK, Herz AV. 2007. From response to stimulus: adaptive sampling in sensory physiology. *Curr Opin Neurobiol* 17: 430–436.

Berkes P, Wiskott L. 2005. Slow feature analysis yields a rich repertoire of complex cell properties. *J Vis* 5: 579–602.

Blasdel GG, Salama G. 1986. Voltage-sensitive dyes reveal a modular organization in monkey striate cortex. *Nature* 321: 579–585.

Bonin V, Mante V, Carandini M. 2006. The statistical computation underlying contrast gain control. *J Neurosci* 26: 6346–6353.

Boynton GM, Demb JB, Glover GH, Heeger DJ. 1999. Neuronal basis of contrast discrimination. *Vision Res* 39: 257–269.

Brincat SL, Connor CE. 2004. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat Neurosci* 7: 880–886.

Brouwer GJ, Heeger DJ. 2009. Decoding and reconstructing color from responses in human visual cortex. *J Neurosci* 29: 13992–14003.

Butts DA, Weng C, Jin J, Yeh CI, Lesica NA, Alonso JM, Stanley GB. 2007. Temporal precision in the neural code and the timescales of natural vision. *Nature* 449: 92–95.

Cadieu C, Kouh M, Pasupathy A, Connor CE, Riesenhuber M, Poggio T. 2007. A model of V4 shape selectivity and invariance. *J Neurophysiol* 98: 1733–1750.

Cadieu CF, Olshausen BA. 2009. Learning transformational invariants from natural movies. In *Advances in Neural Information Processing Systems 21*, ed. D Koller, D Schuurmans, Y Bengio, L Bottou, pp. 209–216. Cambridge, MA: MIT Press.

Campbell FW, Cooper GF, Enroth-Cugell C. 1969. The spatial selectivity of the visual cells of the cat. *J Physiol* 203: 223–235.

Q

Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen BA, Gallant JL, Rust NC. 2005. Do we know what the early visual system does? *J Neurosci* 25: 10577–10597.

Carandini M, Sengpiel F. 2004. Contrast invariance of functional maps in cat primary visual cortex. *J Vis* 4: 130–143.

Chao LL, Haxby JV, Martin A. 1999. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat Neurosci* 2: 913–919.

Chen X, Han F, Poo MM, Dan Y. 2007. Excitatory and suppressive receptive field subunits in awake monkey primary visual cortex (V1). *Proc Natl Acad Sci USA* 104: 19120–19125.

Chichilnisky EJ. 2001. A simple white noise analysis of neuronal light responses. *Network* 12: 199–213.

Cornelissen FW, Wade AR, Vladusich T, Dougherty RF, Wandell BA. 2006. No functional magnetic resonance imaging evidence for brightness and color filling-in in early human visual cortex. *J Neurosci* 26: 3634–3641.

Cox DD, Savoy RL. 2003. Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage* 19: 261–270.

Dakin SC, Hess RF, Ledgeway T, Achtman RL. 2002. What causes non-monotonic tuning of fMRI response to noisy images? *Curr Biol* 12: R476–R477.

Daugman JG. 1980. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Res* 20: 847–856.

David SV, Gallant JL. 2005. Predicting neuronal responses during natural vision. *Network* 16: 239–260.

David SV, Hayden BY, Gallant JL. 2006. Spectral receptive field properties explain shape selectivity in area V4. *J Neurophysiol* 96: 3492–3505.

David SV, Vinje WE, Gallant JL. 2004. Natural stimulus statistics alter the receptive field structure of V1 neurons. *J Neurosci* 24: 6991–7006.

De Valois RL, Albrecht DG, Thorell LG. 1982. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res* 22: 545–559.

DeAngelis GC, Ohzawa I, Freeman RD. 1993. Spatiotemporal organization of simple-cell receptive fields in the cat’s striate cortex. I. General characteristics and postnatal development. *J Neurophysiol* 69: 1091–1117.

DiCarlo JJ, Maunsell JH. 2003. Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *J Neurophysiol* 89: 3264–3278.

Dumoulin SO, Wandell BA. 2008. Population receptive field estimates in human visual cortex. *Neuroimage* 39: 647–660.

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

Felsen G, Touryan J, Han F, Dan Y. 2005. Cortical sensitivity to visual features in natural scenes. *PLoS Biol* 3: e342.

Foster KH, Gaska JP, Nagler M, Pollen DA. 1985. Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. *J Physiol* 365: 331–363.

Gallant JL, Connor CE, Rakshit S, Lewis JW, Van Essen DC. 1996. Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *J Neurophysiol* 76: 2718–2739.

Gauthier I. 2000. What constrains the organization of the ventral temporal cortex? *Trends Cogn Sci* 4: 1–2.

Geisler WS, Albrecht DG. 1997. Visual cortex neurons in monkeys and cats: detection, discrimination, and identification. *Vis Neurosci* 14: 897–919.

Grill-Spector K. 2003. The neural basis of object perception. *Curr Opin Neurobiol* 13: 159–166.

Grill-Spector K, Malach R. 2004. The human visual cortex. *Annu Rev Neurosci* 27: 649–677.

Hartline HK. 1938. The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *Am J Physiol* 121: 400–415.

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: 2425–2430.

Haxby JV, Ishai A, Chao LL, Ungerleider LG, Martin A. 2000. Object-form topology in the ventral temporal lobe. *Trends Cogn Sci* 4: 3–4.

Haynes JD, Lotto RB, Rees G. 2004. Responses of human visual cortex to uniform surfaces. *Proc Natl Acad Sci USA* 101: 4286–4291.

Haynes JD, Rees G. 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci* 8: 686–691.

Heeger DJ, Ress D. 2002. What does fMRI tell us about neuronal activity? *Nat Rev Neurosci* 3: 142–151.

Hegde J, Van Essen DC. 2000. Selectivity for complex shapes in primate visual area V2. *J Neurosci* 20: RC61.

Hegde J, Van Essen DC. 2007. A comparative study of shape representation in macaque visual areas V2 and V4. *Cereb Cortex* 17: 1100–1116.

Henriksson L, Nurminen L, Hyvarinen A, Vanni S. 2008. Spatial frequency tuning in human retinotopic visual areas. *J Vis* 8: 1–13.

Hubel DH, Wiesel TN. 1959. Receptive fields of single neurones in the cat's striate cortex. *J Physiol* 148: 574–591.

Hubel DH, Wiesel TN. 1974. Sequence regularity and geometry of orientation columns in the monkey striate cortex. *J Comp Neurol* 158: 267–293.

Hung CP, Kreiman G, Poggio T, DiCarlo JJ. 2005. Fast readout of object identity from macaque inferior temporal cortex. *Science* 310: 863–866.

Hyvärinen A, Hurri J, Hoyer PO. 2009. *Natural image statistics: A probabilistic approach to early computational vision*. New York: Springer.

Ishai A, Ungerleider LG, Martin A, Schouten JL, Haxby JV. 1999. Distributed representation of objects in the human ventral visual pathway. *Proc Natl Acad Sci USA* 96: 9379–9384.

Ito M, Komatsu H. 2004. Representation of angles embedded within contour stimuli in area V2 of macaque monkeys. *J Neurosci* 24: 3313–3324.

Jones JP, Palmer LA. 1987a. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *J Neurophysiol* 58: 1187–1211.

Jones JP, Palmer LA. 1987b. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J Neurophysiol* 58: 1233–1258.

Kamitani Y, Tong F. 2005. Decoding the visual and subjective contents of the human brain. *Nat Neurosci* 8: 679–685.

Kanwisher N. 2000. Domain specificity in face perception. *Nat Neurosci* 3: 759–763.

Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17: 4302–4311.

Karklin Y, Lewicki MS. 2009. Emergence of complex cell properties by learning to generalize in natural scenes. *Nature* 457: 83–86.

Kastner S, O'Connor DH, Fukui MM, Fehd HM, Herwig U, Pinsk MA. 2004. Functional imaging of the human lateral geniculate nucleus and pulvinar. *J Neurophysiol* 91: 438–448.

Kay KN, David SV, Prenger RJ, Hansen KA, Gallant JL. 2008b. Modeling low-frequency fluctuation and hemodynamic response timecourse in event-related fMRI. *Hum Brain Mapp* 29: 142–156.

Kay KN, Naselaris T, Prenger RJ, Gallant JL. 2008a. Identifying natural images from human brain activity. *Nature* 452: 352–355.

Kayser C, Salazar RF, Konig P. 2003. Responses to natural scenes in cat V1. *J Neurophysiol* 90: 1910–1920.

Kiani R, Esteky H, Mirpour K, Tanaka K. 2007. Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *J Neurophysiol* 97: 4296–4309.

Q

Kim DS, Ronen I, Olman C, Kim SG, Ugurbil K, Toth LJ. 2004. Spatial relationship between neuronal activity and BOLD functional MRI. *Neuroimage* 21: 876–885.

Kinoshita M, Komatsu H. 2001. Neural representation of the luminance and brightness of a uniform surface in the macaque primary visual cortex. *J Neurophysiol* 86: 2559–2570.

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA. 2008. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60: 1126–1141.

Kuffler SW. 1953. Discharge patterns and functional organization of mammalian retina. *J Neurophysiol* 16: 37–68.

Larsson J, Heeger DJ. 2006. Two retinotopic visual areas in human lateral occipital cortex. *J Neurosci* 26: 13128–13142.

Lau B, Stanley GB, Dan Y. 2002. Computational subunits of visual cortical neurons revealed by artificial neural networks. *Proc Natl Acad Sci USA* 99: 8974–8979.

Lauritzen M. 2001. Relationship of spikes, synaptic activity, and local changes of cerebral blood flow. *J Cereb Blood Flow Metab* 21: 1367–1383.

Lauritzen M. 2005. Reading vascular changes in brain imaging: is dendritic calcium the key? *Nat Rev Neurosci* 6: 77–85.

Lennie P, Movshon JA. 2005. Coding of color and form in the geniculostriate visual pathway (invited review). *J Opt Soc Am A Opt Image Sci Vis* 22: 2013–2033.

Lerner Y, Hendler T, Ben-Bashat D, Harel M, Malach R. 2001. A hierarchical axis of object processing stages in the human visual cortex. *Cereb Cortex* 11: 287–297.

Levy I, Hasson U, Avidan G, Hendler T, Malach R. 2001. Center-periphery organization of human object areas. *Nat Neurosci* 4: 533–539.

Lewi J, Butera R, Paninski L. 2009. Sequential optimal design of neurophysiology experiments. *Neural Comput* 21: 619–687.

Logothetis NK. 2008. What we can do and what we cannot do with fMRI. *Nature* 453: 869–878.

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A. 2001. Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412: 150–157.

Logothetis NK, Wandell BA. 2004. Interpreting the BOLD signal. *Annu Rev Physiol* 66: 735–769.

Lowe DG. 1999. Object recognition from local scale-invariant features. Proc of the International Conference on Computer Vision:1150–1157.

MacEvoy SP, Epstein RA. 2007. Position selectivity in scene- and object-responsive occipitotemporal regions. *J Neurophysiol* 98: 2089–2098.

Malach R, Levy I, Hasson U. 2002. The topography of high-order human object areas. *Trends Cogn Sci* 6: 176–184.

Mante V, Bonin V, Carandini M. 2008. Functional mechanisms shaping lateral geniculate responses to artificial and natural stimuli. *Neuron* 58: 625–638.

Mante V, Carandini M. 2005. Mapping of stimulus energy in primary visual cortex. *J Neurophysiol* 94: 788–798.

Martin DR, Fowlkes CC, Malik J. 2004. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Trans Pattern Anal Mach Intell* 26: 530–549.

Mazer JA, Vinje WE, McDermott J, Schiller PH, Gallant JL. 2002. Spatial frequency and orientation tuning dynamics in area V1. *Proc Natl Acad Sci USA* 99: 1645–1650.

Miyawaki Y, Uchida H, Yamashita O, Sato MA, Morito Y, Tanabe HC, Sadato N, Kamitani Y. 2008. Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron* 60: 915–929.

Movshon JA, Thompson ID, Tolhurst DJ. 1978a. Receptive field organization of complex cells in the cat's striate cortex. *J Physiol* 283: 79–99.

Movshon JA, Thompson ID, Tolhurst DJ. 1978b. Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. *J Physiol* 283: 101–120.

Movshon JA, Thompson ID, Tolhurst DJ. 1978c. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J Physiol* 283: 53–77.

Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL. 2009. Bayesian reconstruction of natural images from human brain activity. *Neuron* 63: 902–915.

Nishimoto S, Ishida T, Ohzawa I. 2006. Receptive field properties of neurons in the early visual cortex revealed by local spectral reverse correlation. *J Neurosci* 26: 3269–3280.

Olman CA, Ugurbil K, Schrater P, Kersten D. 2004. BOLD fMRI and psychophysical measurements of contrast response to broadband images. *Vision Res* 44: 669–683.

Olshausen BA, Field DJ. 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381: 607–609.

Op de Beeck HP, DiCarlo JJ, Goense JB, Grill-Spector K, Papanastassiou A, Tanifuji M, Tsao DY. 2008. Fine-scale spatial organization of face and object selectivity in the temporal lobe: do functional magnetic resonance imaging, optical imaging, and electrophysiology agree? *J Neurosci* 28: 11796–11801.

Op de Beeck HP, Haushofer J, Kanwisher NG. 2008. Interpreting fMRI data: maps, modules and dimensions. *Nat Rev Neurosci* 9: 123–135.

Op de Beeck H, Wagemans J, Vogels R. 2001. Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nat Neurosci* 4: 1244–1252.

Orban GA. 2008. Higher order visual processing in macaque extrastriate cortex. *Physiol Rev* 88: 59–89.

Paninski L, Pillow J, Lewi J. 2007. Statistical models for neural encoding, decoding, and optimal stimulus design. *Prog Brain Res* 165: 493–507.

Pasupathy A. 2006. Neural basis of shape representation in the primate brain. *Prog Brain Res* 154: 293–313.

Pasupathy A, Connor CE. 1999. Responses to contour features in macaque area V4. *J Neurophysiol* 82: 2490–2502.

Pasupathy A, Connor CE. 2001. Shape representation in area V4: position-specific tuning for boundary conformation. *J Neurophysiol* 86: 2505–2519.

Pasupathy A, Connor CE. 2002. Population coding of shape in area V4. *Nat Neurosci* 5: 1332–1338.

Peng X, Van Essen DC. 2005. Peaked encoding of relative luminance in macaque areas V1 and V2. *J Neurophysiol* 93: 1620–1632.

Perna A, Tosetti M, Montanaro D, Morrone MC 2008. BOLD response to spatial phase congruency in human brain. *J Vis* 8:15 11–15.

Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, Simoncelli EP. 2008. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454: 995–999.

Pinto N, Doukhan D, DiCarlo JJ, Cox DD. 2009. A high-throughput screening approach to discovering good forms of biologically inspired visual representation. *PLOS Comput Biol* 5: e1000579.

Prenger R, Wu MC, David SV, Gallant JL. 2004. Nonlinear V1 responses to natural scenes revealed by neural network analysis. *Neural Netw* 17: 663–679.

Rainer G, Augath M, Trinath T, Logothetis NK. 2001. Nonmonotonic noise tuning of BOLD fMRI signal to natural images in the visual cortex of the anesthetized monkey. *Curr Biol* 11: 846–854.

Rainer G, Augath M, Trinath T, Logothetis NK. 2002. The effect of image scrambling on visual cortical BOLD activity in the anesthetized monkey. *Neuroimage* 16: 607–616.

Reid RC, Victor JD, Shapley RM. 1997. The use of m-sequences in the analysis of visual neurons: linear receptive field properties. *Vis Neurosci* 14: 1015–1027.

Ringach DL. 2002. Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *J Neurophysiol* 88: 455–463.

Ringach DL. 2004. Mapping receptive fields in primary visual cortex. *J Physiol* 558: 717–728.

Rossi AF, Rittenhouse CD, Paradiso MA. 1996. The representation of brightness in primary visual cortex. *Science* 273: 1104–1107.

Q

Rust NC, Mante V, Simoncelli EP, Movshon JA. 2006. How MT cells analyze the motion of visual patterns. *Nat Neurosci* 9: 1421–1431.

Rust NC, Schwartz O, Movshon JA, Simoncelli EP. 2005. Spatiotemporal elements of macaque V1 receptive fields. *Neuron* 46: 945–956.

Sahani M, Linden JF. 2003. How linear are auditory cortical responses? In *Advances in neural information processing systems 15*, ed. S Becker, S Thrun, K Obermayer, pp. 109–116. Cambridge, MA: MIT Press.

Sasaki Y, Hadjikhani N, Fischl B, Liu AK, Marrett S, Dale AM, Tootell RB. 2001. Local and global attention are mapped retinotopically in human occipital cortex. *Proc Natl Acad Sci USA* 98: 2077–2082.

Sayres R, Grill-Spector K. 2008. Relating retinotopic and object-selective responses in human lateral occipital cortex. *J Neurophysiol* 100: 249–267.

Scannell JW, Young MP. 1999. Neuronal population activity and functional imaging. *Proc Biol Sci* 266: 875–881.

Schiller PH, Finlay BL, Volman SF. 1976. Quantitative studies of single-cell properties in monkey striate cortex. III. Spatial frequency. *J Neurophysiol* 39: 1334–1351.

Schwartz O, Pillow JW, Rust NC, Simoncelli EP. 2006. Spike-triggered neural characterization. *J Vis* 6: 484–507.

Schwarzlose RF, Swisher JD, Dang S, Kanwisher N. 2008. The distribution of category and location information across object-selective regions in human visual cortex. *Proc Natl Acad Sci USA* 105: 4447–4452.

Serre T, Wolf L, Bileschi S, Riesenhuber M, Poggio T. 2007. Robust object recognition with cortex-like mechanisms. *IEEE Trans Pattern Anal Mach Intell* 29: 411–426.

Shapley R, Lennie P. 1985. Spatial frequency analysis in the visual system. *Annu Rev Neurosci* 8: 547–583.

Sharpee TO, Sugihara H, Kurgansky AV, Rebrik SP, Stryker MP, Miller KD. 2006. Adaptive filtering enhances information transmission in visual cortex. *Nature* 439: 936–942.

Singh KD, Smith AT, Greenlee MW. 2000. Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. *Neuroimage* 12: 550–564.

Skouras K, Goutis C, Bramson MJ. 1994. Estimation in linear models using gradient descent with early stopping. *Stat Comput* 4: 271–278.

Smyth D, Willmore B, Baker GE, Thompson ID, Tolhurst DJ. 2003. The receptive-field organization of simple cells in primary visual cortex of ferrets under natural scene stimulation. *J Neurosci* 23: 4746–4759.

Tanaka K. 2003. Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar but slightly different stimulus selectivities. *Cereb Cortex* 13: 90–99.

Tarr MJ, Gauthier I. 2000. FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nat Neurosci* 3: 764–769.

Thirion B, Duchesnay E, Hubbard E, Dubois J, Poline JB, Lebihan D, Dehaene S. 2006. Inverse retinotopy: inferring the visual content of images from brain activation patterns. *Neuroimage* 33: 1104–1116.

Tjan BS, Lestou V, Kourtzi Z. 2006. Uncertainty and invariance in the human visual cortex. *J Neurophysiol* 96: 1556–1568.

Tootell RB, Switkes E, Silverman MS, Hamilton SL. 1988. Functional anatomy of macaque striate cortex. II. Retinotopic organization. *J Neurosci* 8: 1531–1568.

Touryan J, Felsen G, Dan Y. 2005. Spatial structure of complex cell receptive fields measured with natural images. *Neuron* 45: 781–791.

Touryan J, Lau B, Dan Y. 2002. Isolation of relevant visual features from random stimuli for cortical complex cells. *J Neurosci* 22: 10811–10818.

Tsao DY, Freiwald WA, Tootell RB, Livingstone MS. 2006. A cortical region consisting entirely of face-selective cells. *Science* 311: 670–674.

Van Essen DC, Gallant JL. 1994. Neural mechanisms of form and motion processing in the primate visual system. *Neuron* 13: 1–10.

Van Essen DC, Newsome WT, Maunsell JH. 1984. The visual field representation in striate cortex of the macaque monkey: asymmetries, anisotropies, and individual variability. *Vision Res* 24: 429–448.

Victor JD, Purpura K, Katz E, Mao B. 1994. Population encoding of spatial frequency, orientation, and color in macaque V1. *J Neurophysiol* 72: 2151–2166.

Wandell BA. 1999. Computational neuroimaging of human visual cortex. *Annu Rev Neurosci* 22: 145–173.

Wandell BA, Dumoulin SO, Brewer AA. 2007. Visual field maps in human cortex. *Neuron* 56: 366–383.

Weliky M, Fiser J, Hunt RH, Wagner DN. 2003. Coding of natural scenes in primary visual cortex. *Neuron* 37: 703–718.

Wu MC, David SV, Gallant JL. 2006. Complete functional characterization of sensory neurons by system identification. *Annu Rev Neurosci* 29: 477–505.

Yamane Y, Carlson ET, Bowman KC, Wang Z, Connor CE. 2008. A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nat Neurosci* 11: 1352–1360.

Q