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.

Q

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

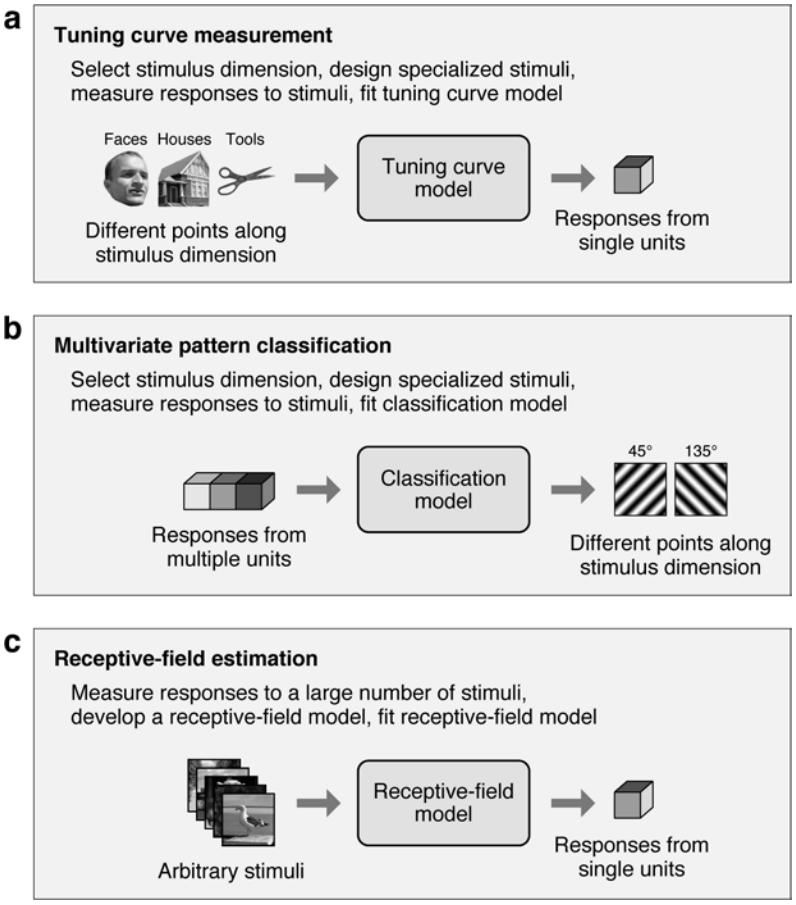


Figure 5.1
Different approaches for studying visual representation. (a) Tuning curve measurement. This approach involves measuring responses to stimuli that vary along a specific dimension and then building a tuning curve model that links different points along the dimension of interest to responses 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.

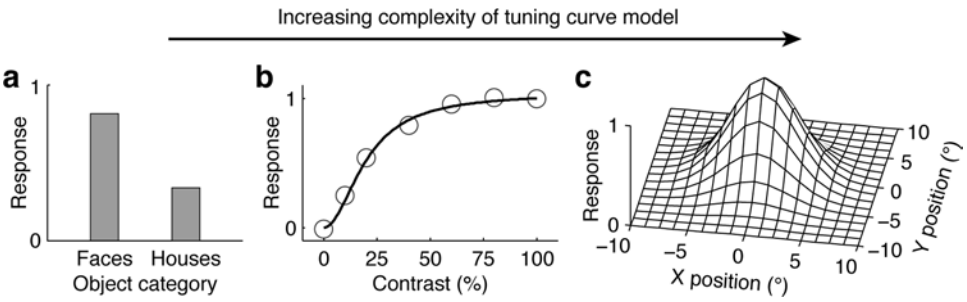


Figure 5.2
Tuning curve models can vary widely in complexity. (a) Model of object category tuning. Suppose we measure responses to objects drawn from two categories, faces and houses. In this case, the dimension of interest is defined on a nominal scale and we can construct a simple tuning curve model 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.

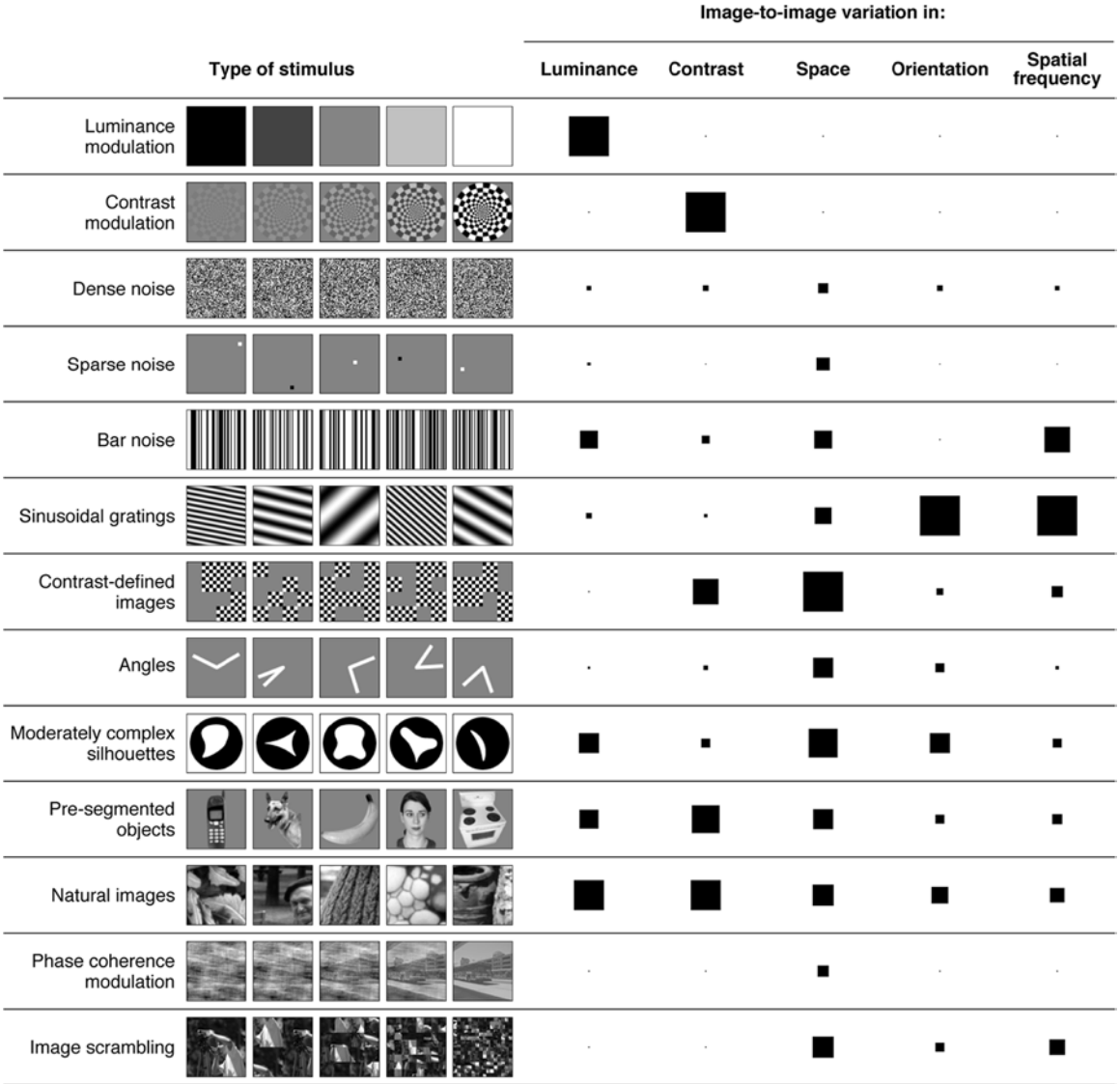


Figure 5.3
Stimuli typically vary along multiple stimulus dimensions. The tuning curve and classification approaches involve designing specialized stimuli that probe specific stimulus dimensions. However, a fundamental problem with this strategy is that in general it is not possible to cleanly separate 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).

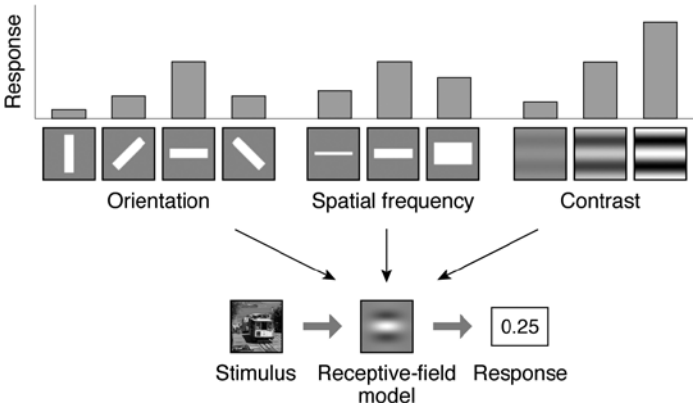


Figure 5.4
Receptive-field models formalize and integrate stimulus dimensions. Suppose we measure tuning curves for the dimensions of orientation, spatial frequency, and contrast. Although these tuning curves provide useful information, it remains unclear how to predict responses to stimuli that differ 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]

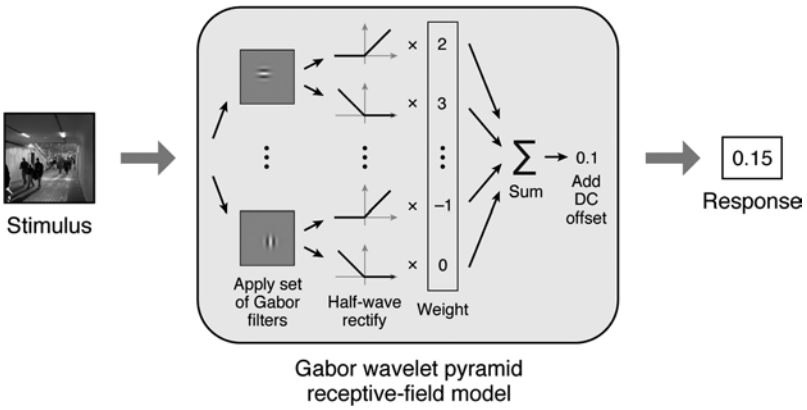


Figure 5.5
Gabor wavelet pyramid receptive-field model. In Kay et al. (2008a), we measured fMRI activity in early visual areas while subjects viewed a large number of grayscale natural images. We then devised a receptive-field model that could potentially characterize the mapping between visual stimuli 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.

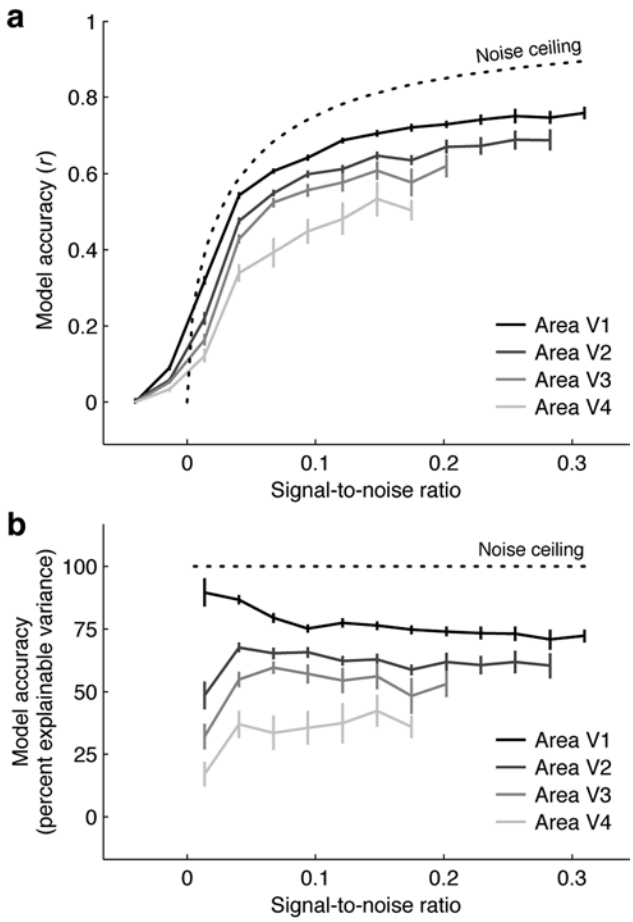


Figure 5.6
Accuracy of the Gabor model. For each voxel, we fit the Gabor model using responses in a training dataset and then assessed how accurately the model predicts responses in a separate validation dataset. (a) Model accuracy as a function of signal-to-noise ratio. In this panel, voxels are binned 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.

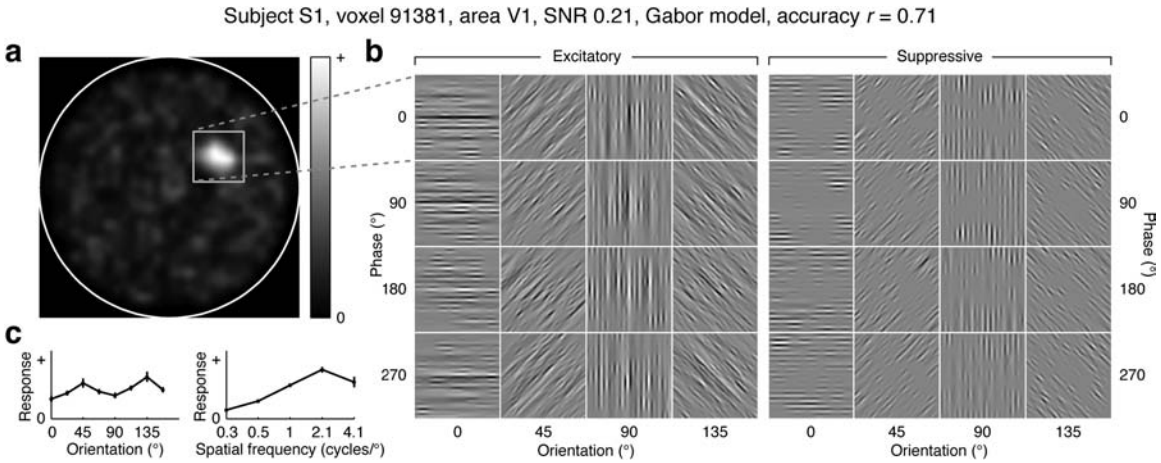


Figure 5.7
Visualization of the receptive field of a representative voxel. (a) Spatial envelope. The receptive field (RF) estimate displayed in this panel was obtained by applying the Gabor model to the full extent of the stimulus (20° × 20°). The intensity of each pixel indicates the sensitivity of 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

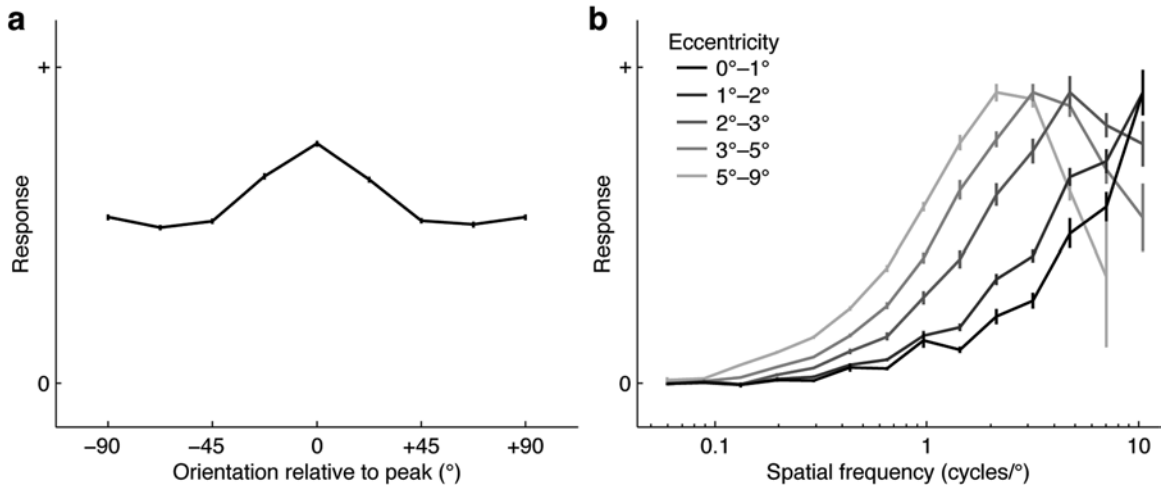


Figure 5.8
Summary of orientation and spatial frequency tuning. We constructed orientation and spatial frequency tuning curves for V1 voxels for which the accuracy (r) of the Gabor model was significantly greater than 0 ($p < 0.01$, one-tailed randomization test). (a) Orientation tuning. To summarize 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).

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.

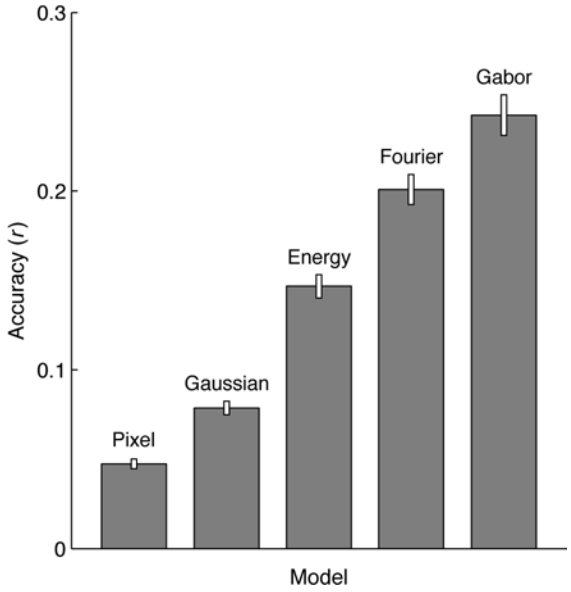


Figure 5.9
Evaluation of alternative models. We formulated several alternative models to compare against the Gabor model. Each model was fit and tested using the same methods used for the Gabor model. In this figure, bar height indicates median accuracy across voxels in V1, and error bars indicate ± 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

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,

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

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.

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

◆ Human Brain Mapping 29:142–156 (2008) ◆

TECHNICAL REPORT

Modeling Low-Frequency Fluctuation and Hemodynamic Response Timecourse in Event-Related fMRI

Kendrick N. Kay,[1] Stephen V. David,[2] Ryan J. Prenger,[3] Kathleen A. Hansen,[1] and Jack L. Gallant[1,4]*

[1]*Department of Psychology, University of California, Berkeley, California*
[2]*Department of Bioengineering, University of California, Berkeley, California*
[3]*Department of Physics, University of California, Berkeley, California*
[4]*Helen Wills Neuroscience Institute, University of California, Berkeley, California*

Abstract: Functional magnetic resonance imaging (fMRI) suffers from many problems that make signal estimation difficult. These include variation in the hemodynamic response across voxels and low signal-to-noise ratio (SNR). We evaluate several analysis techniques that address these problems for event-related fMRI. (1) Many fMRI analyses assume a canonical hemodynamic response function, but this assumption may lead to inaccurate data models. By adopting the finite impulse response model, we show that voxel-specific hemodynamic response functions can be estimated directly from the data. (2) There is a large amount of low-frequency noise fluctuation (LFF) in blood oxygenation level dependent (BOLD) time-series data. To compensate for this problem, we use polynomials as regressors for LFF. We show that this technique substantially improves SNR and is more accurate than high-pass filtering of the data. (3) Model overfitting is a problem for the finite impulse response model because of the low SNR of the BOLD response. To reduce overfitting, we estimate a hemodynamic response timecourse for each voxel and incorporate the constraint of time-event separability, the constraint that hemodynamic responses across event types are identical up to a scale factor. We show that this technique substantially improves the accuracy of hemodynamic response estimates and can be computed efficiently. For the analysis techniques we present, we evaluate improvement in modeling accuracy via 10-fold cross-validation. *Hum Brain Mapp 29:142–156, 2008.* © 2007 Wiley-Liss, Inc.

Key words: hemodynamic response function; low-frequency noise; model evaluation; cross-validation; reverse correlation

Contract grant sponsors: National Institute of Mental Health; The National Eye Institute.

Stephen V. David is currently at Institute for Systems Research, University of Maryland, College Park, MD 20742, USA.

Kathleen A. Hansen is currently at Laboratory of Brain and Cognition, NIMH, Bethesda, MD 20892, USA.

*Correspondence to: Jack L. Gallant, University of California at Berkeley, 3210 Tolman Hall No. 1650, Berkeley, CA 94720, USA. E-mail: gallant@berkeley.edu.

Received for publication 10 April 2006; Revision 13 October 2006; Accepted 2 January 2007

DOI: 10.1002/hbm.20379

Published online 29 March 2007 in Wiley InterScience (www.interscience.wiley.com).

INTRODUCTION

Event-related functional magnetic resonance imaging (fMRI) experimental designs offer several important advantages over block designs: more efficient estimates of the timing and shape of the hemodynamic response (HDR), increased flexibility in experimental design and analysis, and reduction of anticipation and adaptation effects [Josephs and Henson, 1999; Zarahn et al., 1997a]. However, event-related fMRI has reduced statistical power for detecting signal activations [Liu, 2004]. In addition, event-related fMRI increases the complexity of the data and the assumptions underlying the data analysis (e.g. temporal linearity of the BOLD response). It is therefore critical to maximize precision and accuracy in the analysis of event-related fMRI data.

In this study we address three problems in the analysis of event-related fMRI data. Many of the specific techniques we present have been published previously. The goal of the present study is to evaluate rigorously and systematically the value of these techniques, applied in concert, on empirical data. We emphasize cross-validation predictive performance as an objective metric for quantifying model accuracy. (This is in contrast to such metrics as reproducibility and statistical significance, which are important but not directly related to model accuracy.) We also emphasize single voxel modeling, which is likely to become increasingly important as the spatial resolution and signal-to-noise ratio (SNR) of fMRI improve.

One problem in event-related fMRI analysis is variation in the HDR across voxels [Aguirre et al., 1998; Handwerker et al., 2004; Miezin et al., 2000; Neumann et al., 2003; Saad et al., 2001]. Although the assumption of a canonical HDR function (HRF) is common in fMRI analyses, this assumption may lead to incorrect data inferences [Burock and Dale, 2000; Handwerker et al., 2004]. We avoid the assumption of an a priori HRF by adopting the framework of the finite impulse response (FIR) model [Dale, 1999]. Under the FIR model, a HDR is estimated for each voxel to each event type, and there is no constraint on the shape of the responses.

A second problem is the large amount of low-frequency noise fluctuation (LFF) in blood oxygenation level dependent (BOLD) time-series data [Aguirre et al., 1997; Purdon and Weisskoff, 1998; Zarahn et al., 1997b]. LFF has been attributed to scanner and physiological noise [Smith et al., 1999; Zarahn et al., 1997b]. We compensate for LFF by using polynomials [Liu et al., 2001] as regressors for the baseline signal level, i.e. the signal level associated with the absence of the stimulus. We show that this technique improves the SNR and is more accurate than high-pass filtering of the time-series data. Moreover, we show that polynomials can produce more accurate results than Fourier basis functions.

A third problem is model overfitting. Overfitting tends to occur when a model has a large number of parameters relative to the amount of available data. To reduce overfitting by the FIR model, we incorporate the constraint of *time-event separability*. This is the constraint that HDR estimates across event types are identical up to a scale factor, and is reasonable for many experimental paradigms. In a related study, Hinrichs et al. [2000] confirmed increased estimation efficiency under the time-event separable model. We extend their results by demonstrating a simple, fast method for fitting the time-event separable model and by confirming improved cross-validation predictive performance.

We evaluate the proposed analysis techniques on empirical data. These data were obtained from occipital cortex during brief presentations of a checkerboard pattern at different locations in the visual field. Data from this experiment are especially useful for methodological development, because the stimulus is tightly controlled, the SNR is robust, and the data are richly structured. In addition, the sheer number of activated voxels makes it easy to discern population effects. To maximize precision, we analyze the data at the single voxel level, with no spatial smoothing or spatial averaging. We also summarize results from data involving other stimulus designs.

MATERIALS AND METHODS

Stimulus

The stimulus design was similar to that of a previous study from our laboratory [Hansen et al., 2004]. The stimulus consisted of a 7.5-Hz contrast-reversing checkerboard pattern presented within 12 wedges of 30° polar angle width (Fig. 1). The pattern had a radial spatial frequency

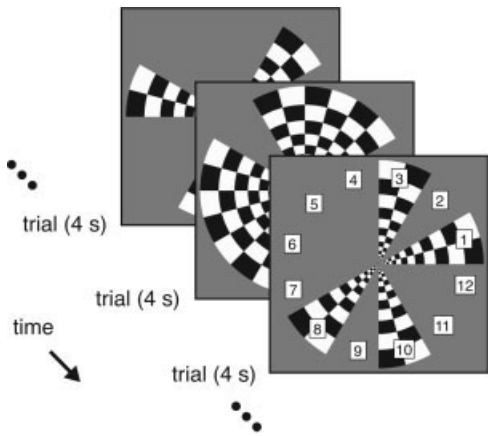


Figure 1.
Schematic of visual stimulus. The stimulus consisted of a 7.5-Hz contrast-reversing checkerboard pattern presented within 12 wedges in the visual field. A cyclically shifted binary m-sequence controlled the presentation timing for each wedge. Each trial lasted 4 s, and there were 255 consecutive trials. For data analysis we define 12 event types, one event type per wedge.

of 12 cycles per revolution and was scaled with eccentricity. At any given time, the pattern was presented within each wedge at 0% (OFF) or 99% (ON) Michelson contrast. The presentation timing was controlled by an m-sequence of level 2, order 8, and length $2^8 - 1 = 255$. The m-sequence was cyclically shifted by 21 elements to produce the ON–OFF pattern for each wedge. The bin duration of the m-sequence was 4 s, and the total stimulus duration was 255 trials $\times$ 4 s = 17 min. For each wedge, there was a total of 128 ON states and 127 OFF states. The minimum, maximum, and mean stimulus onset asynchrony for a given wedge was 4 s, 32 s, and 8 s, respectively.

The use of an m-sequence minimizes correlations between wedges and enables efficient estimation of HDRs [Buracas and Boynton, 2002; Liu, 2004]. m-Sequences have been used in other fMRI studies [de Zwart et al., 2005; Hansen et al., 2004; Kellman et al., 2003]. Code for m-sequence generation was provided by T. Liu (http://fmri-server.ucsd.edu/ttliu/mttfmri_toolbox.html).

The stimulus was displayed by an Epson PowerLite 7700p LCD projector (Epson America, Long Beach, CA) fitted with a custom zoom lens (Buhl Optical, Rochester, NY). The image was focused onto a semitranslucent back-projection screen (Aeroview 100 material, Stewart Filmscreen, Torrance, CA). The subject viewed the screen via a first-surface mirror. The viewing distance was 38 cm, and the stimulus subtended $20^\circ \times 20^\circ$ of visual angle. An occluding device prevented the subject from seeing the unreflected image of the screen. During stimulus presentation, the subject performed a change detection task at a central fixation dot (the mean interval between changes was 2 s). An optical button response box (Current Designs, Philadelphia, PA) recorded subject responses.

The projector operated at a resolution of 1,024 $\times$ 768 at 60 Hz. Luminance output was measured using a Minolta LS-110 photometer (Konica Minolta Photo Imaging, Mahwah, NJ), and the luminance response was linearized via a lookup table. The mean luminance of the stimulus was $\sim$550 cd/m^2. The stimulus was time-locked to the projector refresh rate and synchronized to scanner data acquisition. A Macintosh PowerBook G4 computer (Apple Computer, Cupertino, CA) controlled stimulus presentation and logged button responses, using software written in MATLAB 5.2.1 (The Mathworks, Natick, MA) and Psychophysics Toolbox 2.53 [Brainard, 1997; Pelli, 1997].

Data Collection

The experimental protocol was approved by the UC Berkeley Committee for the Protection of Human Subjects. MRI data were collected at the Brain Imaging Center at UC Berkeley using a 4 T INOVA MR scanner (Varian, Palo Alto, CA) with a whole-body gradient set capable of 35 mT/m with a rise time of 300 μs (Tesla Engineering, Sussex, UK). A curvilinear quadrature transmit/receive surface coil (Midwest RF, LLC, Hartland, WI) was positioned over the occipital pole for enhanced MR SNR. Head motion was minimized with foam padding. Manual shimming of the magnetic field was used to improve image quality and reduce image distortion.

Coronal slices covering occipital cortex were selected: 16 slices, slice thickness 1.8 mm, slice gap 0.2 mm, field-of-view 128 $\times$ 128 mm^2, matrix size 64 $\times$ 64, and nominal resolution 2 $\times$ 2 $\times$ 2 mm^3. For BOLD data, a T2*-weighted, single-shot, slice-interleaved, gradient-echo echo planar imaging (EPI) sequence was used: TR 1 s, TE 0.028 s, flip angle 20°. An initial dummy period was included to allow magnetization to reach steady-state.

During stimulus presentation, the first eight trials were repeated after the end of the 255-trial sequence. BOLD data were collected up through the last trial, and data collected during the initial 8 s $\times$ 4 s = 32 s were ignored [Kellman et al., 2003]. This strategy avoids potential attentional artifacts at the beginning and end of stimulus presentation, compensates for the delay in the HDR, and allows complete sampling of the m-sequence.

Data Preprocessing

A nonlinear phase correction was applied to the image data to reduce Nyquist ghosts and image distortion. Differences in slice acquisition times were corrected via sinc interpolation. To compensate for slow changes in head position, SPM99 motion correction was performed with the following modification: motion parameter estimates were low-pass filtered at 1/20 Hz to remove high-frequency modulations caused by signal activations [Freire and Mangin, 2001]. No additional spatial or temporal filtering was applied.

FIR Model

Our analysis approach is based on the FIR model for event-related fMRI [Dale, 1999]. Our earlier reverse correlation approach [Hansen et al., 2004] is a special case of the FIR model, applicable when stimulus events are uncorrelated.

In the FIR model, the BOLD signal is assumed to be a linear, time-invariant system with respect to the stimulus. A HDR is estimated for each stimulus event type using a set of shifted delta functions as regressors. No assumption on the shape of HDRs is made. Additional regressors are used to model the baseline signal level, i.e. the signal level associated with the absence of the stimulus. The model characterizes two types of effects in the data: *stimulus effects* consist of the transient HDRs to stimulus events, and *nuisance effects* consist of the persistent baseline signal level that may vary over time.

Let e be the number of event types, l be the number of time points in one HDR, m be the number of nuisance terms, and t be the number of time-series data points. The time-series data are modeled as $\mathbf{y}=\mathbf{X}\mathbf{h}+\mathbf{S}\mathbf{b}+\mathbf{n}$, where $\mathbf{y}$ is the data ($t \times 1$), $\mathbf{X}$ is the stimulus matrix ($t \times el$), $\mathbf{h}$ is the concatenation of the HDR associated with each event type

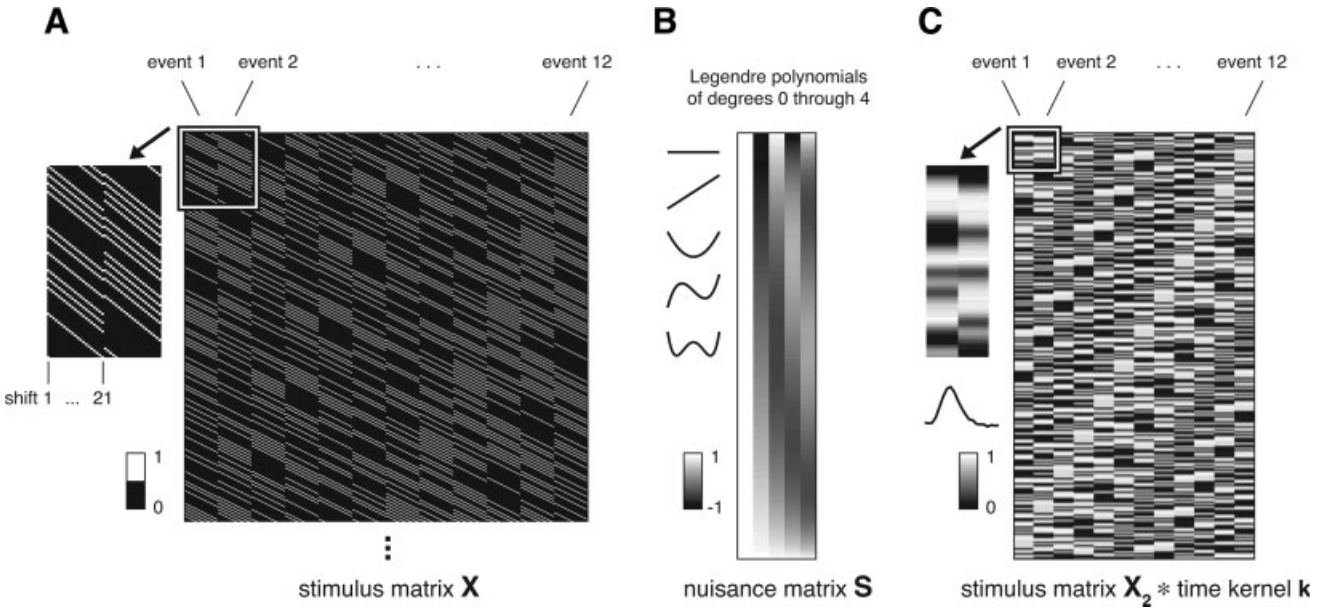


Figure 2.

Schematic of data models. (**A**) Stimulus matrix **X** (FIR model). The matrix dimensions are 1,020 time points × 252 parameters. The matrix is the concatenation of the stimulus convolution matrix for each of the 12 event types. The stimulus convolution matrix for a given event type consists of shifted versions of a binary sequence, where ones indicate event occurrences. There are 21 shifts, one shift for each time point in the HDR estimate. The inset (upper-left) depicts an enlarged view of the parameters for the first two event types. (**B**) Nuisance matrix **S** (polynomial version). The matrix dimensions are 1,020 time points × 5 parameters. The matrix consists of Legendre polynomials of degrees 0 through 4. The inset (upper-left) depicts the polynomials in a line format. (**C**) Convolution of stimulus matrix $\mathbf{X_2}$ and time kernel **k** (time-event separable model). The matrix dimensions are 1,020 time points × 12 parameters. Stimulus matrix $\mathbf{X_2}$ (1,020 × 12) consists of one parameter for each of the 12 event types. The parameter for a given event type is a binary sequence, where ones indicate event occurrences. Time kernel **k** (21 × 1) is a voxel-specific response timecourse estimated from the data. The inset (upper-left) depicts an enlarged view of the parameters for the first two event types. The inset (left) depicts the time kernel in a line format.

($el \times 1$), **S** is the nuisance matrix ($t \times m$), **b** is a set of nuisance parameters ($m \times 1$), and **n** is a noise term ($t \times 1$). The stimulus matrix is the concatenation of the stimulus convolution matrix for each event type. The stimulus convolution matrix for a given event type consists of shifted versions of a binary sequence, where ones indicate event occurrences (Fig. 2). Stimulus effects are given by **Xh**, and nuisance effects are given by **Sb**.

For our data, there are a total of 1,020 time-series data points ($t = 1{,}020$). We define 12 event types, one event type per wedge ($e = 12$). We treat the ON state (99% contrast) of a wedge at the beginning of a trial as an event occurrence. We estimate a HDR of duration 20 s for each event type ($l = 21$). The baseline signal level is the signal level associated with viewing the fixation dot against the gray background.

Modeling LFF

We evaluate several versions of the FIR model. These versions differ in how they compensate for LFF.

In the simple version of the FIR model, LFF is ignored and nuisance matrix **S** consists of only a constant term. This constant term characterizes the baseline signal level as a DC offset in the time-series data. HDR estimates obtained under this version of the FIR model will be poor if the magnitude of LFF is large. This is because LFF adds noise to the time-series data.

One strategy for compensating for LFF is to include in nuisance matrix **S** regressors that model the timecourse of LFF. This strategy enables the modeled baseline signal level to vary over time. Fourier basis functions are commonly used as regressors; in this case, the nuisance matrix consists of a constant term and a set of sine and cosine functions. A different choice of regressors is a set of polynomials of increasing degree (Fig. 2). We use Legendre polynomials [Liu et al., 2001] which are pairwise orthogonal. Equivalent model fits can be obtained with other sets of polynomials (e.g., 1, t, t^2, etc.) that span the same subspace as Legendre polynomials.

Another strategy for compensating for LFF is to detrend the time-series data as a preprocessing step [Kruggel et al., 1999; Marchini and Ripley, 2000; Skudlarski et al., 1999; Tanabe et al., 2002]. We use a high-pass filtering technique: we first remove a linear trend to avoid wrap-around

effects and then high-pass filter the data. On the filtered data, we fit the simple version of the FIR model in which nuisance matrix **S** consists of a constant term.

Time-Event Separable Model

The FIR model uses a large number of parameters to characterize stimulus effects. In our case, there are $(e = 12) \times (l = 21) = 252$ parameters in stimulus matrix **X** and only 1,020 data points. Given the limited amount of data available in a typical fMRI experiment, the FIR model risks overfitting the data.

To reduce the number of model parameters, we incorporate the constraint of time-event separability. This is the condition that HDR estimates across event types are identical up to a scale factor. (More loosely, time-event separability is the condition that the shape of the HDR is the same for any event type.) Under the time-event separable model, stimulus effects are characterized by a single response timecourse—the time kernel—and an amplitude value for each event type. The HDR to an event type is the product of the time kernel and the amplitude value associated with the event type.

The time-series data are modeled as $\mathbf{y}=(\mathbf{X_2} * \mathbf{k})\mathbf{h_2}+\mathbf{Sb}+\mathbf{n}$, where $\mathbf{X_2}$ is the stimulus matrix ($t \times e$), **k** is the time kernel ($l \times 1$), $*$ represents convolution, $\mathbf{h_2}$ is a set of event amplitudes ($e \times 1$), and **S**, **h**, and **n** are as in the FIR model. The stimulus matrix consists of one parameter for each event type. The parameter for a given event type is a binary sequence, where ones indicate event occurrences (Fig. 2). Stimulus effects are given by $(\mathbf{X_2} * \mathbf{k})\mathbf{h_2}$, and nuisance effects are given by **Sb**.

For our data, the time-event separable model uses $(l = 21) + (e = 12) = 33$ parameters to characterize stimulus effects. This is much fewer than the 252 parameters used in the FIR model.

Model Fitting

We fit the FIR model by obtaining the ordinary least-squares estimate $\begin{bmatrix}\hat{\mathbf{h}}\\ \hat{\mathbf{b}}\end{bmatrix} = (\mathbf{W}^\mathbf{T}\mathbf{W})^{-1}\mathbf{W}^\mathbf{T}\mathbf{y}$ where $\mathbf{W}=[\mathbf{X}\ \mathbf{S}]$. This produces $\hat{\mathbf{h}}$, a set of HDR estimates, and $\hat{\mathbf{b}}$, a set of nuisance parameter estimates.

We fit the time-event separable model using two different methods. In the first method (SEPNL), we use an iterative fitting approach [Hinrichs et al., 2000]. We estimate the time kernel, event amplitudes, and nuisance parameters using nonlinear least-squares optimization (MATLAB Optimization Toolbox, Levenberg-Marquardt method). This method determines all model parameters simultaneously, minimizing the squared error between the model fit and the data. A disadvantage of the iterative fitting method is that it is computationally intensive—the method may be impractical given that thousands of voxels are analyzed in a typical fMRI experiment. Also, the fitting method may converge to a local minimum of the error function.

In the second method for fitting the time-event separable model (SEPSVD), we estimate the time kernel before the other model parameters. This approach avoids iterative computation but may not produce an optimal model fit (in the least-squares sense). The method proceeds as follows. We obtain HDR estimates $\hat{\mathbf{h}}$ from the FIR model. We reshape $\hat{\mathbf{h}}$ into a matrix with rows corresponding to event types and columns corresponding to time points ($e \times l$). We perform singular value decomposition on this matrix to obtain the singular vector associated with the largest singular value. This vector is the l-dimensional vector along which variance in $\hat{\mathbf{h}}$ is maximized; this is the time kernel estimate $\hat{\mathbf{k}}$. (Another way to conceptualize $\hat{\mathbf{k}}$ is as the l-dimensional vector that best reconstructs $\hat{\mathbf{h}}$ in the least-squares sense.) Using $\hat{\mathbf{k}}$, we obtain the ordinary least-squares estimate $\begin{bmatrix}\hat{\mathbf{h}}_2\\ \hat{\mathbf{b}}\end{bmatrix} = (\mathbf{W}^\mathbf{T}\mathbf{W})^{-1}\mathbf{W}^\mathbf{T}\mathbf{y}$ where $\mathbf{W} = \left[(\mathbf{X_2} * \hat{\mathbf{k}})\mathbf{S}\right]$. This produces $\hat{\mathbf{h}}_2$, a set of event amplitude estimates, and $\hat{\mathbf{b}}$, a set of nuisance parameter estimates. Note that the time kernel estimate is based on the FIR model fit. Thus, overfitting by the FIR model has some effect on the time kernel estimate. In practice, however, the SEPSVD method performs quite well (see Results).

To obtain standard errors on the parameter estimates of a model, we use a nonparametric jackknife procedure [Efron and Tibshirani, 1993]. We randomly divide the time-series data points into 10 subsets and fit the model 10 times, each time with a different subset excluded. (To exclude data points, we delete rows of **y** and the corresponding rows of **X**, **S**, and $\mathbf{X_2} * \mathbf{k}$.) Standard errors are calculated from the distributions of parameter estimates across the 10 model fits.

To quantify the amplitude of a HDR, we sum over a time window corresponding to the peak of the positive BOLD response [de Zwart et al., 2005]. (For our data, we use the time window of 3–7 s based on inspection of HDR estimates across voxels and event types (Fig. 3).) We quantify the SNR of an event type as the absolute value of the HDR amplitude divided by the standard error of the HDR amplitude. (The standard error is calculated via a jackknife procedure; see earlier.) We quantify the SNR of a voxel as the maximum SNR achieved over all event types. We calculate percent BOLD change relative to the DC parameter estimate (i.e. the parameter estimate for the constant term included in the nuisance matrix).

In one instance we use an alternative SNR metric, which we denote by $\mathrm{SNR_{alt}}$. This metric is useful for comparing the SNR of different models. For a given voxel, we calculate the maximum absolute HDR amplitude (MAX) obtained under any of the models. We then quantify the $\mathrm{SNR_{alt}}$ for each model as MAX divided by the median standard error on HDR amplitudes across events. This metric prevents variability in HDR amplitude estimates from influencing SNR values.

Note that the SNR metrics described earlier are similar to the conventional t-statistic. Thus, one can interpret changes in SNR in terms of statistical significance and

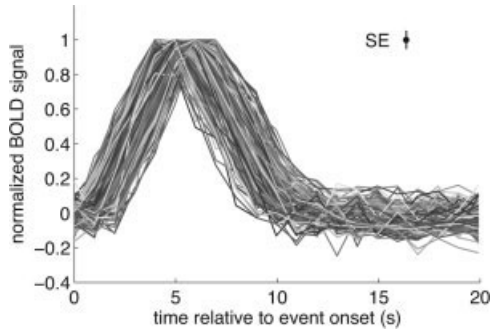


Figure 3.

Inspection of HDR estimates across voxels and event types. Using the POLY model (FIR model combined with polynomials), we obtained for each voxel an estimate of the HDR to each of the 12 event types. The figure depicts positive HDR estimates with a SNR of at least 30 ($n = 216$ from 212 unique voxels). The *x*-axis indicates time relative to event onset; the *y*-axis indicates the BOLD signal. For display purposes, each HDR estimate is normalized by dividing by its maximum value. The inset indicates the median standard error for the depicted data points. The robustness of the shapes of the timecourses indicates the high SNR in the data, despite the small voxel size (2 mm) and the moderate amount of data (17 min).

effect size. For example, suppose we wish to detect a signal change whose magnitude is four times the magnitude of the noise, given a fixed amount of data. At an α value of 0.001 and 9 degrees of freedom (10 jackknifes were taken), the power to detect such a change is 0.23. With a 50% increase in SNR, the power to detect such a change increases to 0.87.

To quantify the magnitude of LFF, we calculate the median absolute deviation (relative to the mean) of the time points of the estimated nuisance effects. We convert the raw BOLD units to standard deviation units, where one standard deviation unit equals the standard deviation of the time-series data with the nuisance effects subtracted. We define the resulting quantity as the *LFF magnitude index*. Intuitively, this index quantifies the typical deviation of the baseline signal level over the course of the time-series. For example, a value of 0.4 indicates that, on average, the baseline signal level is 0.4 standard deviation units away from the mean baseline signal level.

Model Evaluation

We quantify the *fit accuracy* of a model as the coefficient of multiple determination (R^2) between the data and the model fit to the data. This value is the amount of variance in the data explained by the model fit.

When comparing models, an improvement in fit accuracy could reflect improvement in model accuracy, but could also reflect model overfitting. To measure the accuracy of a model while controlling for overfitting, we use a nonparametric *n*-fold cross-validation procedure where $n = 10$. We randomly divide the time-series data points into 10 subsets. We exclude one subset and fit the model on the remaining data points. (To exclude data points, we delete rows of $\mathbf{y}$ and the corresponding rows of $\mathbf{X}$, $\mathbf{S}$, and $\mathbf{X_2} * \mathbf{k}$.) We use the obtained model parameter estimates to predict the data in the excluded subset. The process is repeated 10 times, such that each subset is excluded once. We thereby obtain a prediction for each data point. We quantify the *prediction accuracy* of a model as the coefficient of multiple determination (R^2) between the data and the model prediction of the data. This value is the amount of variance in the data explained by the model prediction. The prediction accuracy of a model is how well the model generalizes to new data, i.e. data not used in the fitting of the model.

LFF often dominate the variance in the time-series data. In these cases, the coefficient of multiple determination is artificially high (e.g., $\gg 0.9$) and reflects primarily how well LFF is modeled. To obtain a prediction accuracy metric that reflects strictly how well stimulus effects are modeled, we perform the following procedure. We subtract the predicted nuisance effects from both the original data and the model prediction. We then calculate the coefficient of multiple determination between the adjusted data and adjusted prediction. We define the resulting value as the *LFF-adjusted* prediction accuracy. (Because the predicted nuisance effects are only estimates and not the true nuisance effects, the metric is potentially biased. However, we observe the same trends in model performance with either prediction accuracy metric.)

In the present context, the coefficient of multiple determination (R^2) directly quantifies how well a given model explains the observed data. Reporting R^2 values is not common in the literature [one exception is Razavi et al., 2003]. When comparing models with respect to R^2 values, a difference of 1–2% can be considered a small effect, a difference of 5% can be considered a moderate effect, and a difference of 10% can be considered a large effect.

Additional Data Sets

We also collected data sets using different subjects, imaging parameters, and stimulus designs. From the perspective of the present study, there is no specific motivation for the particular characteristics of these other data sets. The purpose of these additional data sets is to show that results are not specific to a particular experiment.

Data set 1 is the primary data set described earlier, and involved subject KH (an author). Data set 2 involved subject KK (an author), a volume coil, a two-shot EPI sequence (TR 1 s per shot), and a $3 \times 3 \times 3$ mm^3 voxel size. The stimulus was the same as in data set 1.

Data set 3 involved subject TN and a $2 \times 2 \times 2.5$ mm^3 voxel size. The stimulus consisted of achromatic sinusoidal gratings of eight different orientations. One trial consisted

TABLE I. Summary of data models

Model	Stimulus effects	Nuisance effects
DC	Finite impulse response	Constant term
FOURIER	Finite impulse response	Constant term, sine and cosine functions with 1, 2, and 3 cycles
POLY	Finite impulse response	Polynomials of degrees 0 through 4
FILTER	Finite impulse response	Constant term, after removing a linear trend and high-pass filtering at 1/60 Hz
SEPNL	Time-event separable, iterative fitting method	Polynomials of degrees 0 through 4
SEPSVD	Time-event separable, singular value decomposition fitting method	Polynomials of degrees 0 through 4

This table lists how each model characterizes stimulus effects (i.e. hemodynamic responses to stimulus events) and how each model compensates for nuisance effects (i.e. the baseline signal level).

of the presentation of a grating for 1 s followed by 3 s of a gray background. The eight orientations were repeated 15 times each, and the presentation order was randomly chosen. The stimulus alternated between 16-s periods during which a gray background was presented and 80 s periods during which trials were presented. The stimulus duration was 9.9 min. For data analysis we used eight event types, one event type for each grating orientation.

Data set 4 involved subject TN and a $2 \times 2 \times 2.5$ mm^3 voxel size. The stimulus consisted of 12 grayscale natural photos. One trial consisted of the presentation of a photo for 1 s followed by 3 s of a gray background. Each photo was repeated 13 times; the presentation order was controlled by an m-sequence of level 13, order 2, and length $13^2 - 1 = 168$. The stimulus duration was 11.2 min. For data analysis we used 12 event types, one event type for each distinct photo.

RESULTS

We collected data using multiple subjects, imaging parameters, and stimulus designs. Our analysis results were largely consistent across data sets. In this section we present in-depth results for a single data set (Figs. 1–8), indicate which results were variable in other data sets, and summarize results for all data sets (Fig. 9).

Basic Data Inspection

We conducted an event-related fMRI experiment involving brief (4 s) presentations of a checkerboard pattern within 12 wedges in the visual field (Fig. 1). Our analysis approach is based on the FIR model for event-related fMRI [Dale, 1999]. We define 12 event types, one event type per wedge. For each voxel, a HDR to each of the 12 event types is estimated using a set of shifted delta functions. No assumption on the shape of HDRs is made. Additional regressors are used to model the time-varying baseline signal level (Fig. 2).

We obtained strong BOLD activations in occipital cortex. Figure 3 depicts positive HDR estimates obtained under the POLY model (Table I) with a SNR of at least 30. (This strict criterion selects only those estimates that are nearly noise-free.) The robustness of the shapes of the time-courses confirms the high SNR in the data, despite the small voxel size (2 mm) and the moderate amount of data (17 min). The high SNR is due to the high magnetic field (4 T), the use of a surface coil, the use of an experienced fMRI subject, the m-sequence experimental design, and the high-contrast visual stimulus.

Compensation for LFF

We evaluated several strategies for compensating for LFF in the time-series data. (1) The DC model ignores LFF and uses only a constant term to model DC offset. (2) The FOURIER model uses a constant term and Fourier basis functions with 1, 2, and 3 cycles to model LFF. (3) The POLY model uses Legendre polynomials of degrees 0 through 4 to model LFF. (The spectral content of these polynomials approximately match those of the Fourier basis functions.) (4) The FILTER model removes a linear trend and high-pass filters the time-series data at 1/60 Hz as a preprocessing step.

Panel A of Figure 4 shows that the POLY model greatly increased prediction accuracy compared to the DC model (median increase 14.8%; $P < 0.001$). This indicates ignoring LFF resulted in model fits with poor generalizability. This also indicates that a substantial amount of LFF exists in the time-series data.

Panel B of Figure 4 shows that the POLY model somewhat increased prediction accuracy compared to the FOURIER model (median increase 2.3%; $P < 0.001$). This indicates that polynomials more accurately characterized LFF compared to Fourier basis functions. However, in other data sets, the POLY and FOURIER models had comparable performance (Fig. 9).

Panel C of Figure 4 shows the POLY model substantially increased LFF-adjusted prediction accuracy compared to

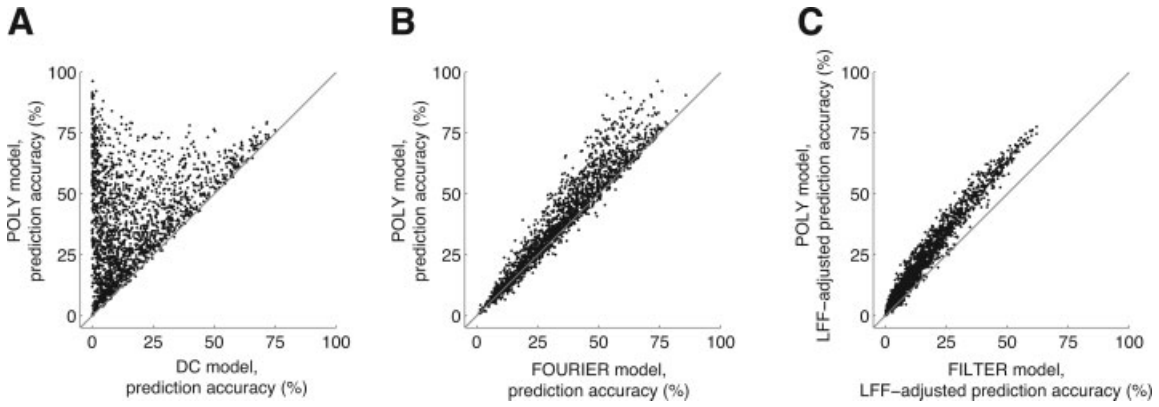


Figure 4.

Modeling LFF with polynomials maximizes prediction accuracy. In these graphs we compare different strategies for LFF compensation (Table I). For each graph, we selected voxels with a minimum SNR of 10 under either of the models being compared. Each point in a graph represents prediction accuracy for a single voxel. (**A**) DC vs. POLY. The *x*- and *y*-axes indicate prediction accuracy under the DC and POLY models, respectively. There was a large increase in accuracy under the POLY model compared to the DC model ($n = 1904$; median increase 14.8%; $P < 0.001$). This indicates that ignoring LFF resulted in model fits with poor generalizability, and that a substantial amount of LFF exists in the time-series data. Some voxels exhibited very large increases in prediction accuracy; in these cases, the contribution of LFF to variance in the time-series data is much larger than the contribution of stimulus effects. (**B**) FOURIER vs. POLY. The *x*- and *y*-axes indicate the prediction accuracy under the FOURIER and POLY models, respectively. There was a small increase in accuracy under the POLY model compared to the FOURIER model ($n = 1,971$; median increase 2.3%; $P < 0.001$). This indicates polynomials more accurately characterized LFF compared to Fourier basis functions in this data set. (**C**) FILTER vs. POLY. The *x*- and *y*-axes indicate the LFF-adjusted prediction accuracy under the FILTER and POLY models, respectively. There was a large increase in accuracy under the POLY model compared to the FILTER model ($n = 1,880$; median increase 6.5%; $P < 0.001$). This indicates that stimulus effects were better characterized when polynomials were used to model LFF compared to when the time-series data were high-pass filtered to remove LFF.

the FILTER model (median increase 6.5%; $P < 0.001$). The use of the LFF-adjusted prediction accuracy metric (see Methods) ensures that increased accuracy under the POLY model is not simply due to the modeling of LFF. The result indicates that stimulus effects were better characterized when polynomials were used to model LFF compared to when the time-series data were high-pass filtered to remove LFF. (We also evaluated the FILTER model using a frequency cutoff of 1/500 Hz; compared to this model, the POLY model still provided a median increase of 2.8% LFF-adjusted prediction accuracy.)

Characteristics of LFF

We investigated in more detail the timecourses of LFF. Panel A of Figure 5 illustrates the effect of manipulating the maximum degree of the polynomials included in the POLY model. Dramatic increases in LFF-adjusted prediction accuracy were obtained by increasing the maximum degree from 0 (median accuracy 5.8%) to 4 (median accuracy 10.8%). Polynomials with degree greater than 4 only marginally increased accuracy; moreover, these increases were inconsistent across voxels (data not shown). Panel A also illustrates the effect of maximum polynomial degree on the SNR. Substantial increases in SNR were obtained by increasing the maximum degree from 0 (median SNR 9.4) to 3 (median SNR 11.6), beyond which SNR did not increase appreciably.

Panel B of Figure 5 depicts the spectral content of Legendre polynomials of degrees 0 through 4. These polynomials consist predominantly of very low frequencies (0–0.004 Hz). With each additional polynomial degree, higher frequencies in the time-series data can be modeled. Panel C of Figure 5 illustrates several example LFF timecourses. Note that the shape and magnitude of LFF vary across voxels.

We quantified the magnitude of LFF with the LFF magnitude index. The index quantifies the typical deviation of the baseline signal level over the time-series data, and is in standard deviation units (see Methods). The 25th and 75th percentiles of the index are 0.18 and 0.57, respectively. (These percentiles were calculated for voxels with a minimum SNR of 10 under the POLY model.) This indicates that noise due to LFF accounts for a substantial fraction of the variation in the time-series data.

Overfitting by the FIR Model

Overfitting tends to occur when a model has a large number of parameters relative to the amount of available data. Two lines of evidence show that the FIR model suffers from overfitting.

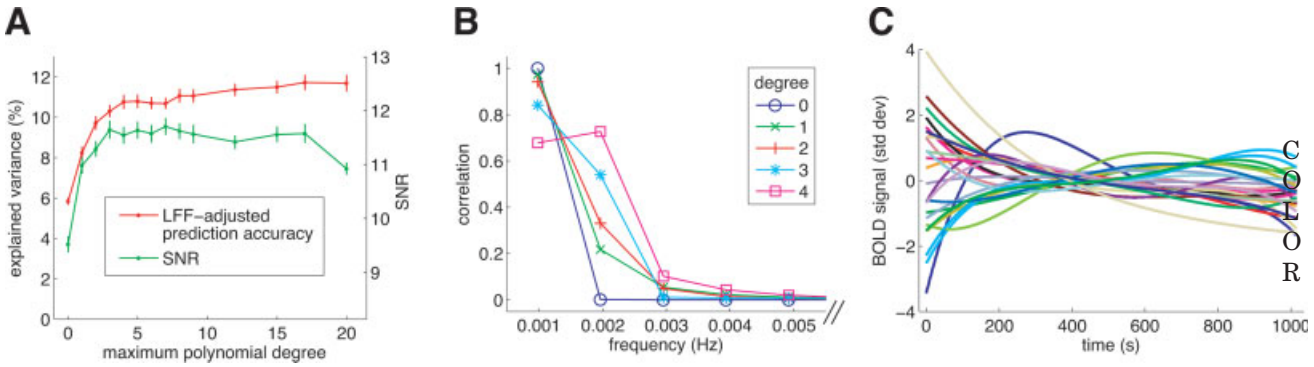


Figure 5.

Characteristics of LFF. (**A**) The effect of the maximum polynomial degree on model performance. We manipulated the maximum degree of the polynomials included in the POLY model (*x*-axis) and evaluated the effect on LFF-adjusted prediction accuracy (*y*-axis; red line) and SNR (*y*-axis; green line). For this graph we selected voxels with a minimum SNR of 10 under any of the model variants ($n = 2{,}890$). Dots indicate the median across voxels, and error bars indicate ± 1 SE (bootstrap procedure). With increasing polynomial degree, both LFF-adjusted prediction accuracy and SNR dramatically increased. (**B**) Spectral content of Legendre polynomials of degrees 0 through 4. The polynomials extend over the course of the time-series data (17 min). We calculated the discrete Fourier transform of each polynomial after applying a Hanning window to avoid edge artifacts and subtracting the mean value. The correlation (*y*-axis) between the time-series data and the Fourier component at each frequency (*x*-axis) is plotted. For display purposes the zero-frequency point is omitted. Note that the polynomials consist predominantly of very low frequencies (0–0.004 Hz). (**C**) Example timecourses of LFF. For 25 voxels we plot nuisance effects as determined under the POLY model. These voxels were randomly selected from voxels with a minimum SNR of 10 ($n = 1{,}730$). The *x*-axis indicates time; the *y*-axis indicates standard deviation units (see Methods). For display purposes, the mean of each timecourse is removed.

The specific HDR window used in the FIR model substantially affected the quality of model fits. Panel A of Figure 6 shows that fit accuracy monotonically increased with window duration. This reflects the fact that, with a longer window duration, additional model parameters are available to fit the data. However, prediction accuracy did not monotonically increase, but was maximized at a duration of 9 s. This indicates that on average, estimating HDRs beyond 9 s

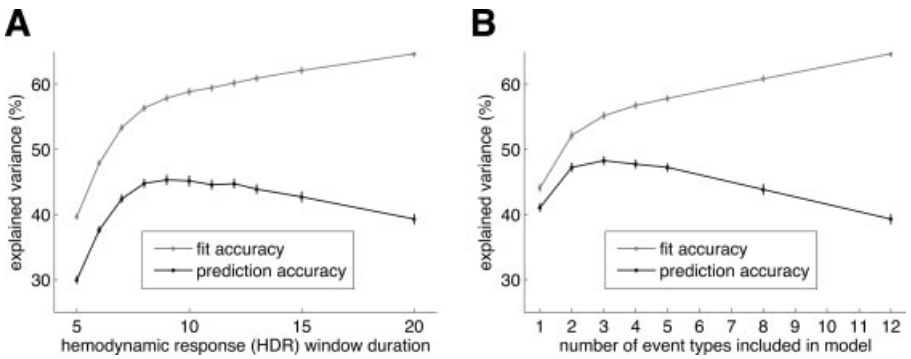


Figure 6.

Overfitting by the FIR model. We manipulated two characteristics of the POLY model (Table 1) and evaluated the effect on fit accuracy (gray line) and prediction accuracy (black line). For these graphs we selected voxels with a minimum SNR of 10 under the POLY model ($n = 1{,}730$). Dots indicate the median across voxels, and error bars indicate ± 1 SE (bootstrap procedure). (**A**) The effect of HDR window duration on fit accuracy and prediction accuracy. The *x*-axis indicates the HDR window duration used in the model; the *y*-axis indicates explained variance. Prediction accuracy was maximized at a duration of 9 s. This indicates that, on average, estimating HDRs beyond 9 s resulted in overfitting and reduced model generalizability. (**B**) The effect of the number of event types on fit accuracy and prediction accuracy. Based on SNR estimates obtained under the POLY model, we refit the model including only the top event types with respect to SNR. (Because different voxels respond to different event types, the included event types varied on a voxel-by-voxel basis.) The *x*-axis indicates the number of event types; the *y*-axis indicates explained variance. Prediction accuracy was maximized at three event types. This indicates that, on average, estimating more than three event types resulted in overfitting and reduced model generalizability. This result is explained by the fact that voxels in visual cortex are often highly selective for spatial position, in such a way that stimuli positioned at nonpreferred locations produce no discernable activation.

resulted in overfitting and reduced model generalizability. This result is consistent with the observation that HDRs have mostly died off by 9 s after event onset (see Fig. 3).

The number of event types included in the FIR model also substantially affected the quality of model fits. We evaluated variants of the model in which only the top event types with respect to SNR are included. (The top event types were determined on a voxel-by-voxel basis.) Panel B of Figure 6 indicates that fit accuracy monotonically increased with number of event types. This reflects the fact that, with more event types, additional model parameters are available to fit the data. However, prediction accuracy did not monotonically increase, but was maximized at three event types. This indicates that on average estimating more than three event types resulted in overfitting and reduced model generalizability. This result is explained by the fact that voxels in visual cortex are often highly selective for spatial position, in such a way that stimuli positioned at nonpreferred locations produce no discernable activation.

Time-Event Separability

To reduce overfitting by the FIR model, we incorporated the constraint of time-event separability. Under the time-event separable model, stimulus effects are characterized by a single response timecourse—the time kernel—and an amplitude value for each event type (Fig. 2). This reduces the number of model parameters that need to be estimated. We evaluated two methods for fitting the time-event separable model, SEPNL and SEPSVD (see Methods).

Panel A of Figure 7 shows that the SEPSVD model greatly increased LFF-adjusted prediction accuracy compared to the POLY model (median increase 9.9%; $P < 0.001$). This indicates that voxel responses were largely time-event separable, and that time-event separability improved the accuracy of HDR estimates. Panel B of Figure 7 shows that the SEPNL model slightly increased LFF-adjusted prediction accuracy compared to the SEPSVD model (median increase 0.5%; $P < 0.001$). This indicates that the two fitting methods produced very similar results. However, in one of the other data sets, the SEPNL model performed substantially better than the SEPSVD model (Fig. 9).

The incorporation of time-event separability also increased the SNR. We selected voxels with a minimum SNR of 10 under either the POLY or SEPSVD model. Of these voxels, the median SNR_{alt} for the POLY model was 14.3, while the median SNR_{alt} for the SEPSVD model was 15.5. This increase was statistically significant ($P < 0.001$).

Example Voxels

We have presented population results thus far, but it is also useful to inspect results for individual voxels. Panels A–E of Figure 8 show model parameter estimates for a typical voxel. Notice the DC model produced very noisy HDR estimates; the FILTER model produced HDR estimates considerably different from those produced by other models; and the SEPSVD model produced the most accurate HDR estimates. Panel F of Figure 8 depicts the spectral content of stimulus effects for the voxel. Notice power is distributed over a wide range of frequencies. Panel G of Figure 8 shows model parameter estimates for another typical voxel. Again, the SEPSVD model produced the most accurate HDR estimates.

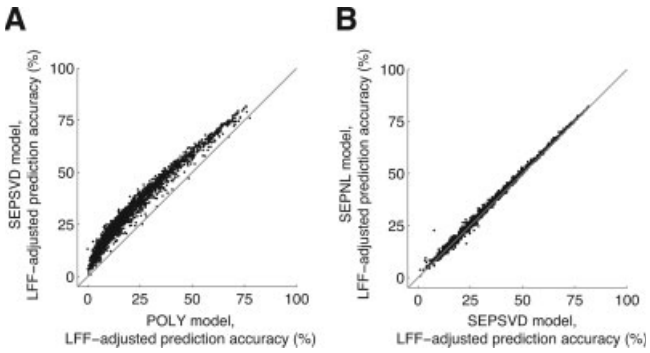


Figure 7.

Time-event separability reduces overfitting and increases prediction accuracy. In these graphs we compare the FIR model to the time-event separable model (Table I). Each point in a graph represents prediction accuracy for a single voxel. (**A**) POLY vs. SEPSVD. The *x*- and *y*-axes indicate the LFF-adjusted prediction accuracy under the POLY and SEPSVD models, respectively. The graph depicts voxels with a minimum SNR of 10 under either data model ($n = 1{,}884$). There was a large increase in accuracy under the SEPSVD model compared to the POLY model (median increase 9.9%; $P < 0.001$). This indicates that voxel responses were largely time-event separable, and that time-event separability improved the accuracy of HDR estimates. (**B**) SEPSVD vs. SEPNL. The *x*- and *y*-axes indicate the LFF-adjusted prediction accuracy under the SEPSVD and SEPNL models, respectively. The graph depicts voxels with a minimum SNR of 10 under the POLY model ($n = 1{,}730$). There was a tiny increase in accuracy under the SEPNL model compared to the SEPSVD model (median increase 0.5%; $P < 0.001$). This indicates that the singular value decomposition fitting method compared favorably against the iterative fitting method in this data set.

Model Performance Summary

Figure 9 summarizes the LFF-adjusted prediction accuracy of the data models we evaluated, and includes results from additional data sets. Across four data sets, the same basic trend in accuracy was observed: the SEPNL and SEPSVD models were the most accurate, the FOURIER and POLY models were moderately accurate, and the DC and FILTER models were the least accurate.

There were two interesting anomalies. First, whereas the POLY model outperformed the FOURIER model for data

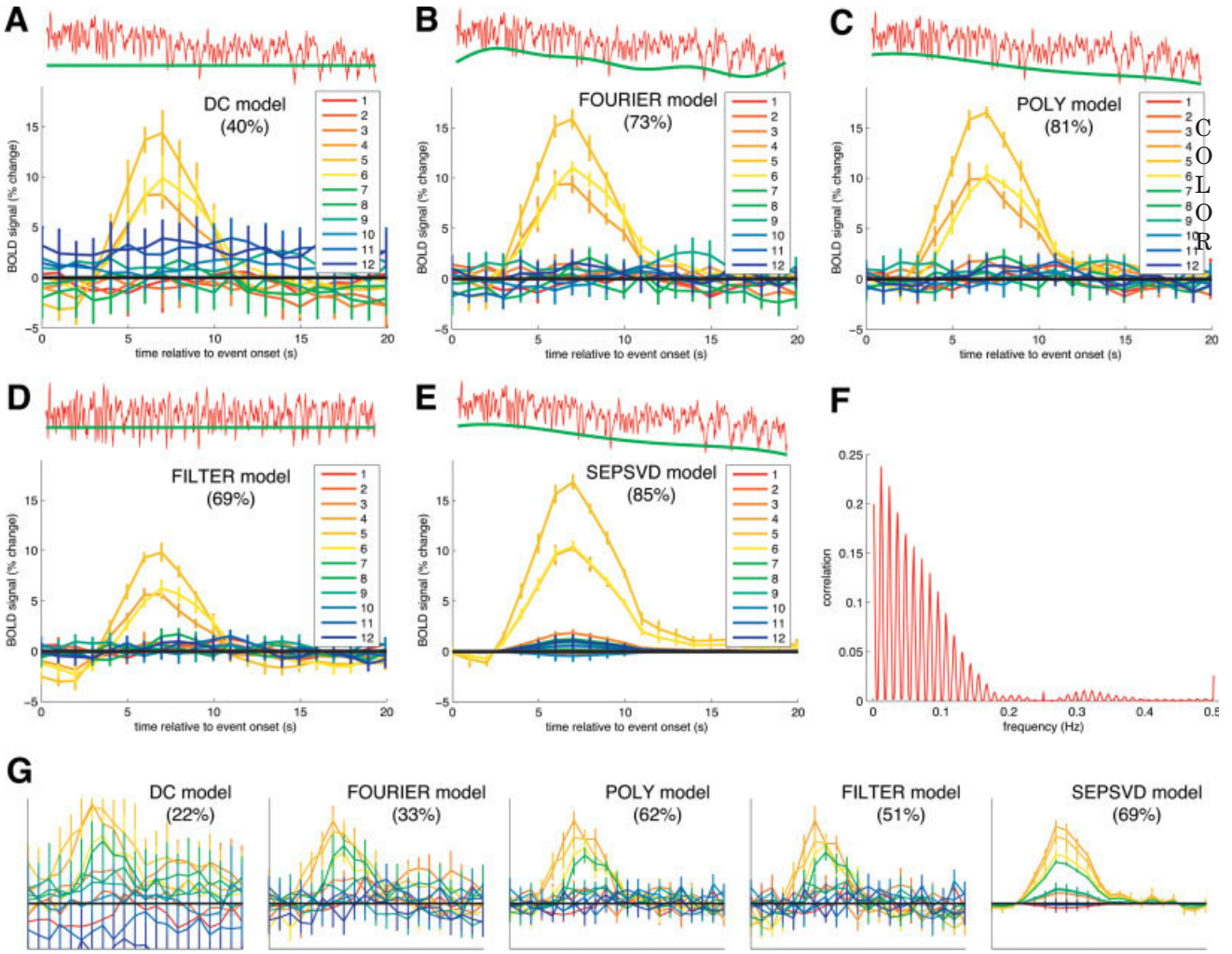


Figure 8.

Comparison of data models for two typical voxels in occipital cortex. Panels A–F depict one voxel, and panel G depicts a second voxel. In panels A–E and G, the main axes show HDR estimates for the 12 event types. The *x*-axis indicates time relative to event onset; the *y*-axis indicates percent BOLD change. A thick black horizontal line indicates zero percent BOLD change. Error bars indicate ±1 SE (jackknife procedure). Indicated in parentheses is the LFF-adjusted prediction accuracy, which is calculated via 10-fold cross-validation. The inset axes above the main axes depict the time-series data (red line) and nuisance effects (green line). (**A**) DC model. This model ignores LFF and uses only a constant term to characterize the baseline signal level. Under this model, HDR estimates were very noisy and prediction accuracy was poor. (**B**) FOURIER model. This model uses a constant term and Fourier basis functions with 1, 2, and 3 cycles to model LFF. Compared to the DC model, HDR estimates were less noisy and prediction accuracy was better. Note that the nuisance effects poorly track the time-series data at the beginning and end of the time-series. (**C**) POLY model. This model uses polynomials of degrees 0 through 4 to model LFF. Compared to the FOURIER model, HDR estimates were slightly less noisy and prediction accuracy was better. Notice the nuisance effects track the time-series data well. The LFF magnitude index is 0.87. (**D**) FILTER model. This model high-pass filters the time-series data at 1/60 Hz to remove LFF as a preprocessing step. The filtered data are shown in red in the inset (above). HDR estimates were considerably different from those obtained under other data models. (**E**) SEPSVD model. This model incorporates the constraint of time-event separability and uses polynomials of degrees 0 through 4 to model LFF. Time-event separability is the condition that HDR estimates across event types are identical up to a scale factor. Prediction accuracy was highest under the SEPSVD model. The LFF magnitude index is 0.86. (**F**) Spectral content of stimulus effects. We obtained the estimated timecourse of stimulus effects under the SEPSVD model. We calculated the discrete Fourier transform of this timecourse after subtracting the mean value. The correlation (*y*-axis) between the time-series data and the Fourier component at each frequency (*x*-axis) is plotted. For display purposes the zero-frequency point is omitted. Note that the power is distributed over a wide range of frequencies. (**G**) Comparison of data models for a second voxel. The format is identical to that of panels A–E, except that the *y*-axis ranges from −3 to 7. Again, the SEPSVD model had the highest prediction accuracy.

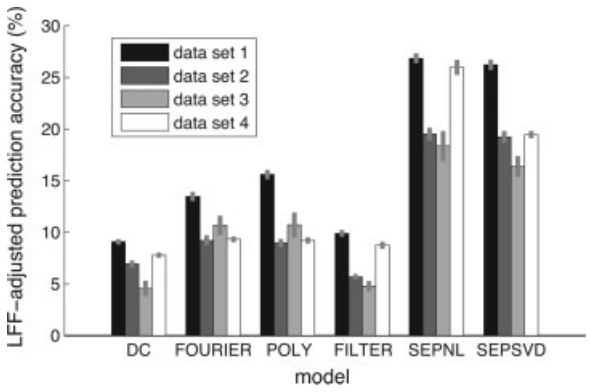


Figure 9.

Summary of data model performance. This graph summarizes results from four data sets involving different subjects, imaging parameters, and stimulus designs: the primary data set (illustrated in Figs. 1–8) and three additional data sets. For each data set, we selected voxels (n = 2,223, 699, 236, 825, respectively) passing a minimum SNR threshold (= 10, 10, 7, 10, respectively) under any of the models that do not involve iterative fitting (this excludes the SEPNL model). (The SNR threshold is lowered for data set 3 due to low signal in that data set.) The x-axis indicates the data models we evaluated; and the y-axis indicates the LFF-adjusted prediction accuracy. The height of each bar indicates the median across voxels, and the bar shading indicates the data set. Error bars indicate ±1 SE (bootstrap procedure). The same basic trend in performance was observed across the four data sets: the SEPNL and SEPSVD models were the most accurate, the FOURIER and POLY models were moderately accurate, and the DC and FILTER models were the least accurate.

set 1 (see also Fig. 4), the two models had similar performance in other data sets. The only difference between the two models is the choice of regressors for LFF. The variable results across data sets suggest that LFF characteristics are dependent on the subject, imaging parameters, and/or stimulus design.

Second, whereas in data sets 1–3, the SEPNL and SEPSVD models had similar performance, in data set 4, the SEPNL model substantially increased accuracy compared to the SEPSVD model (median increase across voxels 5.4%; $P < 0.001$). The reason for the large but inconsistent increase in accuracy under the SEPNL model is an issue for further investigation. We speculate that the variable results may be due to strong temporal nonlinearities in data set 4.

DISCUSSION

Linearity Assumptions

The FIR and time-event separable models assume that the BOLD response is a linear, time-invariant system with respect to the stimulus. However, violations of time-invariance have been widely documented [Boynton et al., 1996; Buxton et al., 2004; Dale and Buckner, 1997; Friston et al., 2000b; Glover, 1999; Huettel and McCarthy, 2001; Logothetis, 2003; Miezin et al., 2000; Wager et al., 2005]. In general, the response to an event closely preceded by another event has a greater delay and lower amplitude than expected. This temporal nonlinearity may be neural in origin (e.g. adaptation) and/or related to the coupling between neural activity and the BOLD response [Bandettini et al., 2002; Birn et al., 2001; Boynton and Finney, 2003; Huettel et al., 2004; Janz et al., 2001; Ogawa et al., 2000].

We dampened the impact of temporal nonlinearities in our experimental design by the use of a 4-s bin duration. This is because deviations from linearity are large only at short-stimulus durations [Birn et al., 2001; Boynton et al., 1996; Pfeuffer et al., 2003; Vazquez and Noll, 1998]. Whereas the response to a moderate-length stimulus (4 s) well predicts the response to a longer stimulus (8 s), the response to a short stimulus (1 s) poorly predicts the response to a longer stimulus (2 s). It may be possible to devise models to account for temporal nonlinearities when they exist [Friston et al., 2000b; Wager et al., 2005].

Our experimental design involves simultaneous presentation of different event types (i.e. multiple wedges in the visual field at any given time). The primary purpose of simultaneous presentation is to increase the number of event repetitions and thereby increase the SNR. Note that both the FIR and time-event separable models assume that the BOLD response is additive across events: that is, the response to events presented simultaneously is equal to the sum of the responses to the events presented in isolation. The validity of this assumption depends on the experimental paradigm [Hansen et al., 2004]. However, the analysis techniques we present are not specific to experimental designs using simultaneous event presentation.

Low-Frequency Fluctuation

We found that using polynomials to model LFF resulted in more accurate HDR estimates than those obtained with other strategies. We used an event-related experimental design, and our study complements studies that investigated LFF for block designs [LaConte et al., 2003; Razavi et al., 2003].

The poor performance of high-pass filtering is explained by the fact that stimulus effects in our data exist at low frequencies. High-pass filtering removes LFF but also removes a portion of the stimulus effects [Kruggel et al., 1999; Ollinger et al., 2001; Skudlarski et al., 1999; Smith et al., 1999]. Moreover, the removal of stimulus effects induces bias in HDR estimates (Fig. 8). Detrending techniques (of which high-pass filtering is one instance) are appropriate only when stimulus effects can be assumed to be absent at low frequencies (e.g. a periodic ON–OFF block experimental design).

Using regressors to model LFF is not equivalent to removing these regressors from the time-series data before fitting the data model [Liu et al., 2001]. The latter is effec-

tively a detrending technique. As such, it neglects potential correlation between stimulus effects and the regressors that are removed. Detrending may also increase autocorrelation in the noise component of the time-series data, and thereby decrease the validity of a model that assumes uncorrelated noise [Razavi et al., 2003]. It is necessary to fit both stimulus effects and nuisance effects simultaneously in order to estimate the individual contributions of these two effects at low frequencies.

We found that a set of polynomials of degrees 0 through 4 modeled LFF well. Most of the spectral power in these polynomials is between 0 and 0.004 Hz (for a 17-min data set). Because the 1-Hz data sampling rate is sufficient for characterizing the respiratory cycle (~0.25 Hz), it is unlikely that LFF reflects respiration-related noise. However, the 1-Hz data sampling rate is insufficient for characterizing the cardiac cycle (~1 Hz), and so aliasing of cardiac-related signals could be contributing to LFF. Measurement of the cardiac cycle during data acquisition could perhaps be used to improve modeling of the time-series data. However, there is some evidence that LFF is dominated by nonphysiological factors such as scanner instability [Smith et al., 1999].

Including polynomials of higher degree increased prediction accuracy, but only marginally and inconsistently. This indicates that the magnitude of LFF at higher frequencies was relatively small, and that including additional polynomials risked overfitting. Tailoring the number of polynomials on a voxel-by-voxel basis is a possible strategy.

In our data, the baseline signal level is generally not the same at the beginning and at the end of the time-series data. This is one reason that Fourier basis functions, which are periodic, did not model LFF as well as polynomials in data set 1. However, the characteristics of LFF may be specific to the experimental setup [Aguirre et al., 1997; Purdon and Weisskoff, 1998; Zarahn et al., 1997b]. It is therefore necessary to evaluate different models for LFF on a case-by-case basis (Fig. 9).

A different way to approach the problem of LFF is to focus on the autocorrelation in the noise in BOLD time-series data [for reviews, see Bullmore et al., 2001; Friston et al., 2000a]. Prewhitening strategies have been proposed for obtaining estimates under the general linear model that have less variance than ordinary least-squares estimates [Bullmore et al., 1996; Burock and Dale, 2000; Friman et al., 2004; Locascio et al., 1997; Marchini and Ripley, 2000; Purdon and Weisskoff, 1998; Woolrich et al., 2001; Worsley et al., 2002]. The addition of prewhitening to the use of regressors for LFF may result in further improvements in SNR and prediction accuracy.

Time Kernel Estimation

The time kernel for a voxel can be viewed as a voxel-specific HRF. The technique of time kernel estimation occupies a middle ground between assuming a canonical HRF and making no assumption about the shape of HDRs (FIR model). In the first case, the shape of the HDR is assumed to be known, and the only free model parameters are the amplitude for each event type. Overfitting is unlikely because there are few model parameters, but model accuracy is suboptimal because of variation in HDR shape across voxels. In the second case, a separate HDR is estimated for each event type, resulting in many free model parameters. Variation in HDR shape can be accounted for, but model accuracy is suboptimal because of overfitting. By estimating a time kernel, we greatly reduce the number of free model parameters, but still make no assumption about the shape of the HDR across voxels.

Increased prediction accuracy under the time-event separable model is contingent on the degree to which voxel responses are in fact time-event separable. The large increase in prediction accuracy in our data indicates that voxel responses were largely time-event separable, but does not necessarily imply complete separability. Time-event separability likely holds in many experimental paradigms. Because the BOLD response temporally blurs underlying neural activity, we expect time-event separability to hold whenever the timescale of neural activity is roughly the same across event types.

In some experimental paradigms, we may expect aspects of the HDR timecourse (e.g. onset, width) to vary across event types. For example, we might expect the delay of the neural activity in a voxel to be dependent on the level of difficulty of a cognitive task. In such a case, we would expect the onset of the HDR timecourse to vary across easy and hard instances of the task. The assumption of time-event separability would be inappropriate for such experimental paradigms.

Overfitting and Regularization

The FIR model substantially overfitted our data, producing HDR estimates that had suboptimal prediction accuracy. Overfitting is a substantial problem for event-related fMRI because of the low SNR of the BOLD response and the large number of parameters necessary to accommodate variations in the shape of the HDR.

Overfitting by the FIR model can be reduced by tailoring the HDR window or by modeling only a subset of the event types. However, the optimal HDR window and subset of event types for a given voxel may not generalize to other voxels. Searching for the optimal parameters on a voxel-by-voxel basis is computationally impractical.

A practical solution to overfitting by the FIR model is to incorporate the constraint of time-event separability. This constraint greatly reduces the number of model parameters—in our case, the number of model parameters is reduced from 252 to 33. Note that the time-event separable model does not have any additional descriptive power compared to the FIR model: any set of HDRs that can be characterized by the time-event separable model can also

be characterized by the FIR model. Thus, we can view time-event separability as a means of regularizing the FIR model. (Regularization refers to techniques that attempt to improve prediction accuracy by introducing a specific bias to model parameter estimates.)

Other regularization techniques are possible and are not mutually exclusive to time-event separability. They include fitting a parametric function, such as a γ function, to HDR estimates [Boynton et al., 1996; Cohen, 1997; Glover, 1999]; incorporating temporal basis set restrictions or other priors into the FIR model [Burock and Dale, 2000; Dale, 1999; Goutte et al., 2000]; and incorporating constraints on the spatial pattern of signal activations [Katanoda et al., 2002; Kiebel et al., 2000; Kruggel et al., 1999; Purdon et al., 2001]. If these techniques are used, it is important to verify that they improve prediction accuracy.

ACKNOWLEDGMENTS

We thank B. Inglis for extensive MRI assistance; R. Redfern for equipment design and construction; M. Banks, K. Schreiber, and S. Watt for stimulus presentation advice; B. Pasley, M. Silver, and T. Liu for helpful discussions; J. Ollinger for image reconstruction software; and K. Gustavsen, B. Inglis, A. Rokem, and anonymous reviewers for comments on the manuscript. K. Kay was supported by the National Defense Science and Engineering Graduate Fellowship.

REFERENCES

Aguirre GK, Zarahn E, D'Esposito M (1997): Empirical analyses of BOLD fMRI statistics. II. Spatially smoothed data collected under null-hypothesis and experimental conditions. Neuroimage 5:199–212.

Aguirre GK, Zarahn E, D'Esposito M (1998): The variability of human, BOLD hemodynamic responses. Neuroimage 8:360–369.

Bandettini PA, Birn RM, Kelley D, Saad Z (2002): Dynamic nonlinearities in BOLD contrast: Neuronal or hemodynamic? Int Congr Ser 1235:73–95.

Birn RM, Saad ZS, Bandettini PA (2001): Spatial heterogeneity of the nonlinear dynamics in the FMRI BOLD response. Neuroimage 14:817–826.

Boynton GM, Finney EM (2003): Orientation-specific adaptation in human visual cortex. J Neurosci 23:8781–8787.

Boynton GM, Engel SA, Glover GH, Heeger DJ (1996): Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci 16:4207–4221.

Brainard DH (1997): The psychophysics toolbox. Spat Vis 10:433–436.

Bullmore E, Brammer M, Williams SC, Rabe-Hesketh S, Janot N, David A, Mellers J, Howard R, Sham P (1996): Statistical methods of estimation and inference for functional MR image analysis. Magn Reson Med 35:261–277.

Bullmore E, Long C, Suckling J, Fadili J, Calvert G, Zelaya F, Carpenter TA, Brammer M (2001): Colored noise and computational inference in neurophysiological (fMRI) time series analysis: Resampling methods in time and wavelet domains. Hum Brain Mapp 12:61–78.

Buracas GT, Boynton GM (2002): Efficient design of event-related fMRI experiments using M-sequences. Neuroimage 16(3, Part 1):801–813.

Burock MA, Dale AM (2000): Estimation and detection of event-related fMRI signals with temporally correlated noise: A statistically efficient and unbiased approach. Hum Brain Mapp 11:249–260.

Buxton RB, Uludag K, Dubowitz DJ, Liu TT (2004): Modeling the hemodynamic response to brain activation. Neuroimage 23(Suppl 1):S220–S233.

Cohen MS (1997): Parametric analysis of fMRI data using linear systems methods. Neuroimage 6:93–103.

Dale AM (1999): Optimal experimental design for event-related fMRI. Hum Brain Mapp 8:109–114.

Dale AM, Buckner RL (1997): Selective averaging of rapidly presented individual trials using fMRI. Hum Brain Mapp 5:329–340.

de Zwart JA, Silva AC, van Gelderen P, Kellman P, Fukunaga M, Chu R, Koretsky AP, Frank JA, Duyn JH (2005): Temporal dynamics of the BOLD fMRI impulse response. Neuroimage 24:667–677.

Efron B, Tibshirani R (1993): An Introduction to the Bootstrap. Vol. 16. New York: Chapman & Hall. 436pp.

Freire L, Mangin JF (2001): Motion correction algorithms may create spurious brain activations in the absence of subject motion. Neuroimage 14:709–722.

Friman O, Borga M, Lundberg P, Knutsson H (2004): Detection and detrending in fMRI data analysis. Neuroimage 22:645–655.

Friston KJ, Josephs O, Zarahn E, Holmes AP, Rouquette S, Poline J (2000a): To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. Neuroimage 12:196–208.

Friston KJ, Mechelli A, Turner R, Price CJ (2000b): Nonlinear responses in fMRI: The balloon model, volterra kernels, and other hemodynamics. Neuroimage 12:466–477.

Glover GH (1999): Deconvolution of impulse response in event-related BOLD fMRI. Neuroimage 9:416–429.

Goutte C, Nielsen FA, Hansen LK (2000): Modeling the haemodynamic response in fMRI using smooth FIR filters. IEEE Trans Med Imaging 19:1188–1201.

Handwerker DA, Ollinger JM, D'Esposito M (2004): Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. Neuroimage 21:1639–1651.

Hansen KA, David SV, Gallant JL (2004): Parametric reverse correlation reveals spatial linearity of retinotopic human V1 BOLD response. Neuroimage 23:233–241.

Hinrichs H, Scholz M, Tempelmann C, Woldorff MG, Dale AM, Heinze HJ (2000): Deconvolution of event-related fMRI responses in fast-rate experimental designs: Tracking amplitude variations. J Cogn Neurosci 12(Suppl 2):76–89.

Huettel SA, McCarthy G (2001): Regional differences in the refractory period of the hemodynamic response: An event-related fMRI study. Neuroimage 14:967–976.

Huettel SA, Obembe OO, Song AW, Woldorff MG (2004): The BOLD fMRI refractory effect is specific to stimulus attributes: Evidence from a visual motion paradigm. Neuroimage 23:402–408.

Janz C, Heinrich SP, Kornmayer J, Bach M, Hennig J (2001): Coupling of neural activity and BOLD fMRI response: New insights by combination of fMRI and VEP experiments in transition from single events to continuous stimulation. Magn Reson Med 46:482–486.

Josephs O, Henson RN (1999): Event-related functional magnetic resonance imaging: Modelling, inference and optimization. Philos Trans R Soc Lond B Biol Sci 354:1215–1228.

Katanoda K, Matsuda Y, Sugishita M (2002): A spatio-temporal regression model for the analysis of functional MRI data. Neuroimage 17:1415–1428.

Kellman P, Gelderen P, de Zwart JA, Duyn JH (2003): Method for functional MRI mapping of nonlinear response. Neuroimage 19:190–199.

Kiebel SJ, Goebel R, Friston KJ (2000): Anatomically informed basis functions. Neuroimage 11(6, Part 1):656–667.

Kruggel F, von Cramon DY, Descombes X (1999): Comparison of filtering methods for fMRI datasets. Neuroimage 10:530–543.

LaConte S, Anderson J, Muley S, Ashe J, Frutiger S, Rehm K, Hansen LK, Yacoub E, Hu X, Rottenberg D, Strother S (2003): The evaluation of preprocessing choices in single-subject BOLD fMRI using NPAIRS performance metrics. Neuroimage 18:10–27.

Liu TT (2004): Efficiency, power, and entropy in event-related fMRI with multiple trial types. II. Design of experiments. Neuroimage 21:401–413.

Liu TT, Frank LR, Wong EC, Buxton RB (2001): Detection power, estimation efficiency, and predictability in event-related fMRI. Neuroimage 13:759–773.

Locascio JL, Jennings PJ, Moore CI, Corkin S (1997): Time series analysis in the time domain and resampling methods for studies of functional magnetic resonance brain imaging. Hum Brain Mapp 5:168–193.

Logothetis NK (2003): The underpinnings of the BOLD functional magnetic resonance imaging signal. J Neurosci 23:3963–3971.

Marchini JL, Ripley BD (2000): A new statistical approach to detecting significant activation in functional MRI. Neuroimage 12:366–380.

Miezin FM, Maccotta L, Ollinger JM, Petersen SE, Buckner RL (2000): Characterizing the hemodynamic response: Effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. Neuroimage 11(6, Part 1):735–759.

Neumann J, Lohmann G, Zysset S, von Cramon DY (2003): Within-subject variability of BOLD response dynamics. Neuroimage 19:784–796.

Ogawa S, Lee TM, Stepnoski R, Chen W, Zhu XH, Ugurbil K (2000): An approach to probe some neural systems interaction by functional MRI at neural time scale down to milliseconds. Proc Natl Acad Sci USA 97:11026–11031.

Ollinger JM, Corbetta M, Shulman GL (2001): Separating processes within a trial in event-related functional MRI. Neuroimage 13:218–229.

Pelli DG (1997): The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spat Vis 10:437–442.

Pfeuffer J, McCullough JC, Van de Moortele PF, Ugurbil K, Hu X (2003): Spatial dependence of the nonlinear BOLD response at short stimulus duration. Neuroimage 18:990–1000.

Purdon PL, Weisskoff RM (1998): Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. Hum Brain Mapp 6:239–249.

Purdon PL, Solo V, Weisskoff RM, Brown EN (2001): Locally regularized spatiotemporal modeling and model comparison for functional MRI. Neuroimage 14:912–923.

Razavi M, Grabowski TJ, Vispoel WP, Monahan P, Mehta S, Eaton B, Bolinger L (2003): Model assessment and model building in fMRI. Hum Brain Mapp 20:227–238.

Saad ZS, Ropella KM, Cox RW, DeYoe EA (2001): Analysis and use of FMRI response delays. Hum Brain Mapp 13:74–93.

Skudlarski P, Constable RT, Gore JC (1999): ROC analysis of statistical methods used in functional MRI: Individual subjects. Neuroimage 9:311–329.

Smith AM, Lewis BK, Ruttimann UE, Ye FQ, Sinnwell TM, Yang Y, Duyn JH, Frank JA (1999): Investigation of low frequency drift in fMRI signal. Neuroimage 9:526–533.

Tanabe J, Miller D, Tregellas J, Freedman R, Meyer FG (2002): Comparison of detrending methods for optimal fMRI preprocessing. Neuroimage 15:902–907.

Vazquez AL, Noll DC (1998): Nonlinear aspects of the BOLD response in functional MRI. Neuroimage 7:108–118.

Wager TD, Vazquez A, Hernandez L, Noll DC (2005): Accounting for nonlinear BOLD effects in fMRI: Parameter estimates and a model for prediction in rapid event-related studies. Neuroimage 25:206–218.

Woolrich MW, Ripley BD, Brady M, Smith SM (2001): Temporal autocorrelation in univariate linear modeling of FMRI data. Neuroimage 14:1370–1386.

Worsley KJ, Liao CH, Aston J, Petre V, Duncan GH, Morales F, Evans AC (2002): A general statistical analysis for fMRI data. Neuroimage 15:1–15.

Zarahn E, Aguirre G, D'Esposito M (1997a): A trial-based experimental design for fMRI. Neuroimage 6:122–138.

Zarahn E, Aguirre GK, D'Esposito M (1997b): Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. Neuroimage 5:179–197.

doi:10.1038/nature06713

LETTERS

Identifying natural images from human brain activity

Kendrick N. Kay[1], Thomas Naselaris[2], Ryan J. Prenger[3] & Jack L. Gallant[1,2]

A challenging goal in neuroscience is to be able to read out, or decode, mental content from brain activity. Recent functional magnetic resonance imaging (fMRI) studies have decoded orientation[1,2], position[3] and object category[4,5] from activity in visual cortex. However, these studies typically used relatively simple stimuli (for example, gratings) or images drawn from fixed categories (for example, faces, houses), and decoding was based on previous measurements of brain activity evoked by those same stimuli or categories. To overcome these limitations, here we develop a decoding method based on quantitative receptive field models that characterize the relationship between visual stimuli and fMRI activity in early visual areas. These models describe the tuning of individual voxels for space, orientation and spatial frequency, and are estimated directly from responses evoked by natural images. We show that these receptive field models make it possible to identify, from a large set of completely novel natural images, which specific image was seen by an observer. Identification is not a mere consequence of the retinotopic organization of visual areas; simpler receptive field models that describe only spatial tuning yield much poorer identification performance. Our results suggest that it may soon be possible to reconstruct a picture of a person's visual experience from measurements of brain activity alone.

Imagine a general brain reading device that could reconstruct a picture of a person's visual experience at any moment in time[6]. This general visual decoder would have great scientific and practical use. For example, we could use the decoder to investigate differences in perception across people, to study covert mental processes such as attention, and perhaps even to access the visual content of purely mental phenomena such as dreams and imagery. The decoder would also serve as a useful benchmark of our understanding of how the brain represents sensory information.

How do we build a general visual decoder? We consider as a first step the problem of image identification[3,7,8]. This problem is analogous to the classic 'pick a card, any card' magic trick. We begin with a large, arbitrary set of images. The observer picks an image from the set and views it while brain activity is measured. Is it possible to use the measured brain activity to identify which specific image was seen?

To ensure that a solution to the image identification problem will be applicable to general visual decoding, we introduce two challenging requirements[6]. First, it must be possible to identify novel images. Conventional classification based decoding methods can be used to identify images if brain activity evoked by those images has been measured previously, but they cannot be used to identify novel images (see Supplementary Discussion). Second, it must be possible

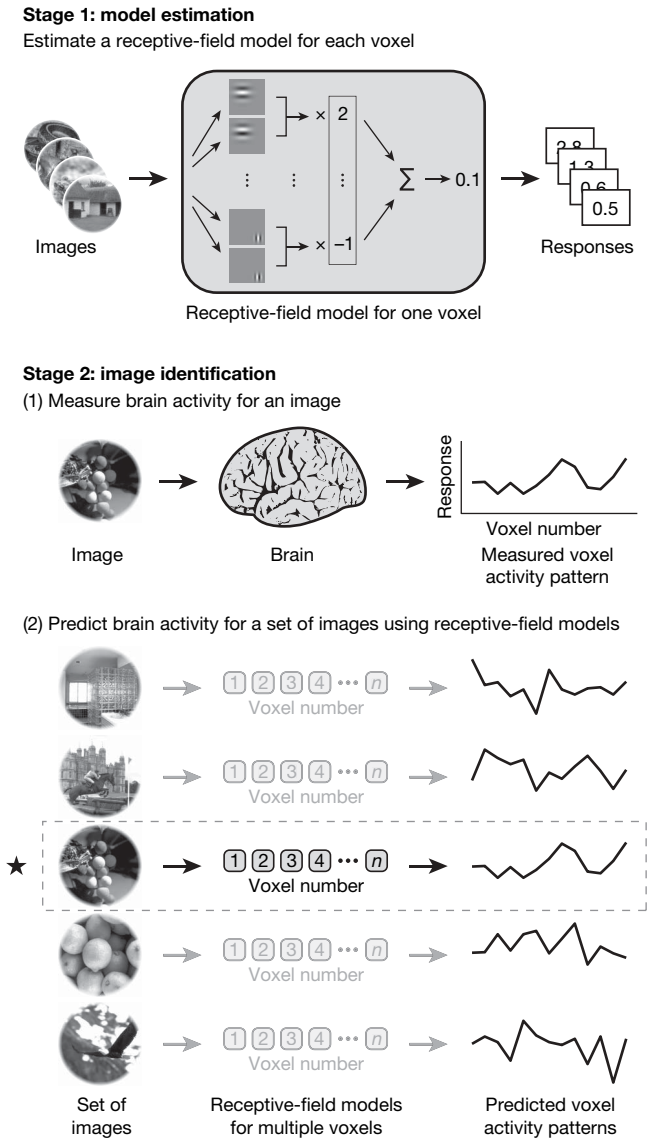


Figure 1 | Schematic of experiment. The experiment consisted of two stages. In the first stage, model estimation, fMRI data were recorded while each subject viewed a large collection of natural images. These data were used to estimate a quantitative receptive field model[10] for each voxel. The model was based on a Gabor wavelet pyramid[11–13] and described tuning along the dimensions of space[3,14–19], orientation[1,2,20] and spatial frequency[21,22]. In the second stage, image identification, fMRI data were recorded while each subject viewed a collection of novel natural images. For each measurement of brain activity, we attempted to identify which specific image had been seen. This was accomplished by using the estimated receptive field models to predict brain activity for a set of potential images and then selecting the image whose predicted activity most closely matches the measured activity.

[1]Department of Psychology, University of California, Berkeley, California 94720, USA. [2]Helen Wills Neuroscience Institute, University of California, Berkeley, California 94720, USA. [3]Department of Physics, University of California, Berkeley, California 94720, USA.

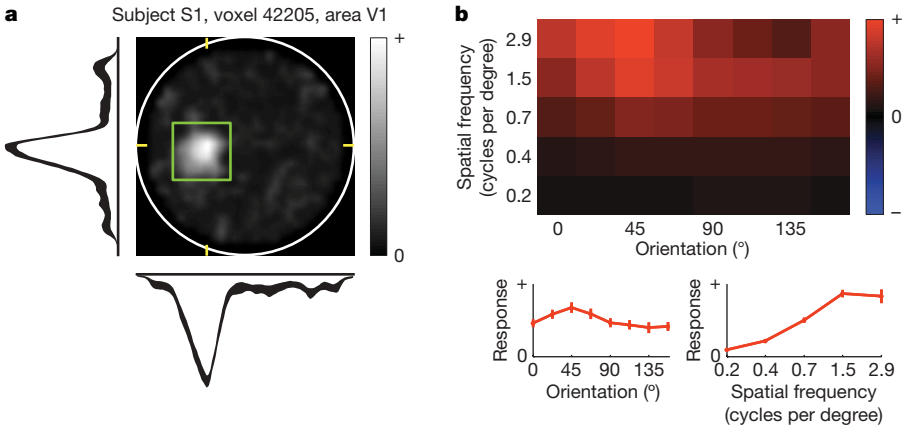


Figure 2 | Receptive-field model for a representative voxel. a, Spatial envelope. The intensity of each pixel indicates the sensitivity of the receptive field to that location. The white circle delineates the bounds of the stimulus ($20^\circ \times 20^\circ$) and the green square delineates the estimated receptive field location. Horizontal and vertical slices through the spatial envelope are shown below and to the left. These intersect the peak of the spatial envelope, as indicated by yellow tick marks. The thickness of each slice profile indicates ± 1 s.e.m. This receptive field is located in the left hemifield, just below the horizontal meridian. **b**, Orientation and spatial frequency tuning curves. The top matrix depicts the joint orientation and spatial frequency tuning of the receptive field, and the bottom two plots give the marginal orientation and spatial frequency tuning curves. Error bars indicate ± 1 s.e.m. This receptive field has broadband orientation tuning and high pass spatial frequency tuning. For additional receptive field examples and population summaries of receptive field properties, see Supplementary Figs 9 11.

to identify natural images. Natural images have complex statistical structure[9] and are much more difficult to parameterize than simple artificial stimuli such as gratings or pre segmented objects. Because neural processing of visual stimuli is nonlinear, a decoder that can identify simple stimuli may fail when confronted with complex natural images.

Our experiment consisted of two stages (Fig. 1). In the first stage, model estimation, fMRI data were recorded from visual areas V1, V2 and V3 while each subject viewed 1,750 natural images. We used these data to estimate a quantitative receptive field model[10] for each voxel (Fig. 2). The model was based on a Gabor wavelet pyramid[11 13] and described tuning along the dimensions of space[3,14 19], orienta tion[1,2,20] and spatial frequency[21,22]. (See Supplementary Discussion for a comparison of our receptive field analysis with those of pre vious studies.)

In the second stage, image identification, fMRI data were recorded while each subject viewed 120 novel natural images. This yielded 120 distinct voxel activity patterns for each subject. For each voxel activity pattern we attempted to identify which image had been seen. To do this, the receptive field models estimated in the first stage of the experiment were used to predict the voxel activity pattern that would be evoked by each of the 120 images. The image whose predicted voxel activity pattern was most correlated (Pearson's *r*) with the measured voxel activity pattern was selected.

Identification performance for one subject is illustrated in Fig. 3. For this subject, 92% (110/120) of the images were identified correctly (subject S1), whereas chance performance is just 0.8% (1/120). For a second subject, 72% (86/120) of the images were identified correctly (subject S2). These high performance levels demonstrate the validity of our decoding approach, and indicate that our receptive field models accurately characterize the selectivity of individual voxels to natural images.

A general visual decoder would be especially useful if it could operate on brain activity evoked by a single perceptual event. However, because fMRI data are noisy, the results reported above were obtained using voxel activity patterns averaged across 13 repeated trials. We therefore attempted identification using voxel activity patterns from single trials. Single trial performance was 51% (834/1620) and 32% (516/1620) for subjects S1 and S2, respec tively (Fig. 4a); once again, chance performance is just 0.8% (13.5/1620). These results suggest that it may be feasible to decode the content of perceptual experiences in real time[7,23].

We have so far demonstrated identification of a single image drawn from a set of 120 images, but a general visual decoder should be able to handle much larger sets of images. To investigate this issue, we measured identification performance for various set sizes up to 1,000 images (Fig. 4b). As set size increased tenfold from 100 to 1,000, performance only declined slightly, from 92% to 82% (subject S1,

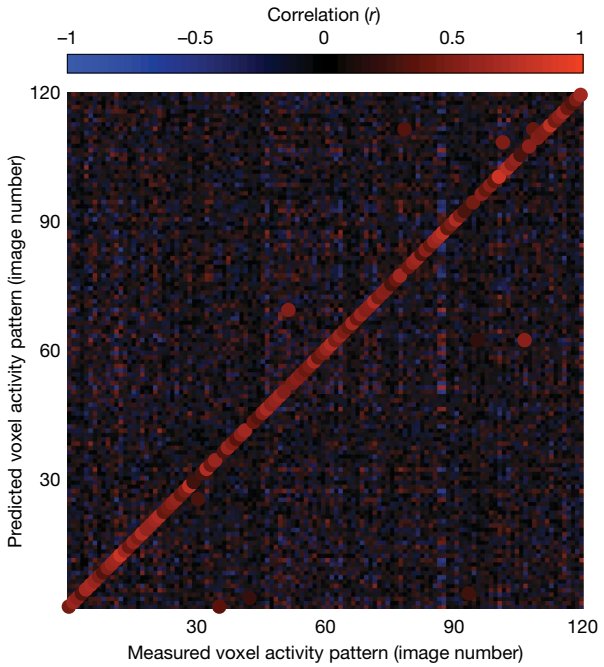


Figure 3 | Identification performance. In the image identification stage of the experiment, fMRI data were recorded while each subject viewed 120 novel natural images that had not been used to estimate the receptive field models. For each of the 120 measured voxel activity patterns, we attempted to identify which image had been seen. This figure illustrates identification performance for one subject (S1). The colour at the *m*th column and *n*th row represents the correlation between the measured voxel activity pattern for the *m*th image and the predicted voxel activity pattern for the *n*th image. The highest correlation in each column is designated by an enlarged dot of the appropriate colour, and indicates the image selected by the identification algorithm. For this subject 92% (110/120) of the images were identified correctly.

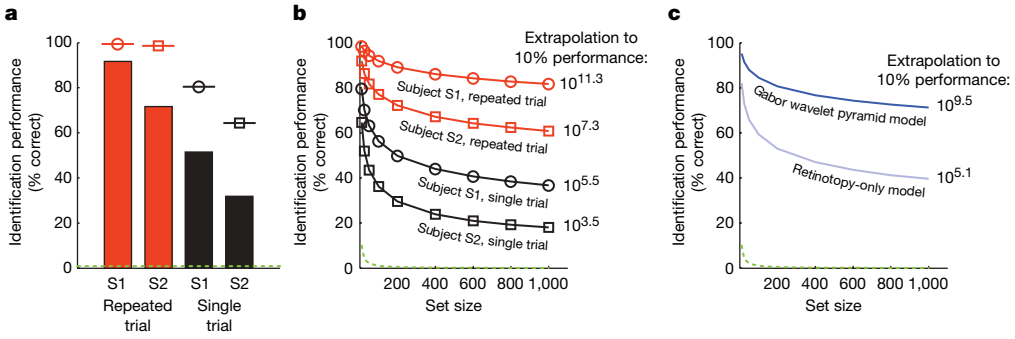


Figure 4 | Factors that impact identification performance. **a**, Summary of identification performance. The bars indicate empirical performance for a set size of 120 images, the marker above each bar indicates the estimated noise ceiling (that is, the theoretical maximum performance given the level of noise in the data), and the dashed green line indicates chance performance. The noise ceiling estimates suggest that the difference in performance across subjects is due to intrinsic differences in the level of noise. **b**, Scaling of identification performance with set size. The *x* axis indicates set size, the *y* axis indicates identification performance, and the number to the right of each line gives the estimated set size at which performance declines to 10% correct. In all cases performance scaled very well with set size. **c**, Retinotopy only model versus Gabor wavelet pyramid model. Identification was attempted using an alternative retinotopy only model that captures only the location and size of each voxel's receptive field. This model performed substantially worse than the Gabor wavelet pyramid model, indicating that spatial tuning alone is insufficient to achieve optimal identification performance. (Results reflect repeated trial performance averaged across subjects; see Supplementary Fig. 5 for detailed results.)

repeated trial). Extrapolation of these measurements (see Supplementary Methods) suggests that performance for this subject would remain above 10% even up to a set size of $10^{11.3}$ images. This is more than 100 times larger than the number of images currently indexed by Google ($10^{8.9}$ images; source: http://www.google.com/whatsnew/, 4 June 2007).

Early visual areas are organized retinotopically, and voxels are known to reflect this organization[14,16,18]. Could our results be a mere consequence of retinotopy? To answer this question, we attempted identification using an alternative model that captures the location and size of each voxel's receptive field but discards orientation and spatial frequency information (Fig. 4c). Performance for this retinotopy only model declined to 10% correct at a set size of just $10^{5.1}$ images, whereas performance for the Gabor wavelet pyramid model did not decline to 10% correct until $10^{9.5}$ images were included in the set (repeated trial performance extrapolated and averaged across subjects). This result indicates that spatial tuning alone does not yield optimal identification performance; identification improves substantially when orientation and spatial frequency tuning are included in the model.

To further investigate the impact of orientation and spatial frequency tuning, we measured identification performance after imposing constraints on the orientation and spatial frequency tuning of the Gabor wavelet pyramid model (Supplementary Fig. 8). The results indicate that both orientation and spatial frequency tuning contribute to identification performance, but that the latter makes the larger contribution. This is consistent with recent studies demonstrating that voxels have only slight orientation bias[1,2]. We also find that voxel to voxel variation in orientation and spatial frequency tuning contributes to identification performance. This reinforces the growing realization in the fMRI community that information may be present in fine grained patterns of voxel activity[6].

To be practical our identification algorithm must perform well even when brain activity is measured long after estimation of the receptive field models. To assess performance over time[2,4,6,23] we attempted identification for a set of 120 novel natural images that were seen approximately two months after the initial experiment. In this case 82% (99/120) of the images were identified correctly (chance performance 0.8%; subject S1, repeated trial). We also evaluated identification performance for a set of 12 novel natural images that were seen more than a year after the initial experiment. In this case 100% (12/12) of the images were identified correctly (chance performance 8%; subject S1, repeated trial). These results demonstrate that the stimulus related information that can be decoded from voxel activity remains largely stable over time.

Why does identification sometimes fail? Inspection revealed that identification errors tended to occur when the selected image was visually similar to the correct image. This suggests that noise in measured voxel activity patterns causes the identification algorithm to confuse images that have similar features.

Functional MRI signals have modest spatial resolution and reflect haemodynamic activity that is only indirectly coupled to neural activity[24,25]. Despite these limitations, we have shown that fMRI signals can be used to achieve remarkable levels of identification performance. This indicates that fMRI signals contain a considerable amount of stimulus related information[4] and that this information can be successfully decoded in practice.

Identification of novel natural images brings us close to achieving a general visual decoder. The final step will require devising a way to reconstruct the image seen by the observer, instead of selecting the image from a known set. Stanley and co workers[26] reconstructed natural movies by modelling the luminance of individual image pixels as a linear function of single unit activity in cat lateral geniculate nucleus. This approach assumes a linear relation between luminance and the activity of the recorded units, but this condition does not hold in fMRI[27,28].

An alternative approach to reconstruction is to incorporate receptive field models into a statistical inference framework. In such a framework, receptive field models are used to infer the most likely image given a measured activity pattern. This model based approach has a long history in both theoretical and experimental neuroscience[29,30]. Recently, Thirion and co workers[3] used it to reconstruct spatial maps of contrast from fMRI activity in human visual cortex. The success of the approach depends critically on how well the receptive field models predict brain activity. The present study demonstrates that our receptive field models have sufficient predictive power to enable identification of novel natural images, even for the case of extremely large sets of images. We are therefore optimistic that the model based approach will make possible the reconstruction of natural images from human brain activity.

METHODS SUMMARY

The stimuli consisted of sequences of 20° × 20° greyscale natural photographs (Supplementary Fig. 1a). Photographs were presented for 1 s with a delay of 3 s between successive photographs (Supplementary Fig. 1b). Subjects (S1: author T.N.; S2: author K.N.K.) viewed the photographs while fixating a central white square. MRI data were collected at the Brain Imaging Center at University of California, Berkeley using a 4 T INOVA MR scanner (Varian, Inc.) and a quadrature transmit/receive surface coil (Midwest RF, LLC). Functional BOLD data were recorded from occipital cortex at a spatial resolution of 2 mm × 2 mm × 2.5 mm and a temporal resolution of 1 Hz. Brain volumes were

reconstructed and then co registered to correct differences in head positioning within and across scan sessions. The time series data were pre processed such that voxel specific response time courses were deconvolved from the data. Voxels were assigned to visual areas based on retinotopic mapping data[17] col lected in separate scan sessions.

In the model estimation stage of the experiment, a receptive field model was estimated for each voxel. The model was based on a Gabor wavelet pyramid[11 13] (Supplementary Figs 2 and 3), and was able to characterize responses of voxels in early visual areas V1, V2 and V3 (Supplementary Table 1). Alternative receptive field models were also used, including the retinotopy only model and several constrained versions of the Gabor wavelet pyramid model. Details of these models and model estimation procedures are given in Supplementary Methods.

In the image identification stage of the experiment, the estimated receptive field models were used to identify images viewed by the subjects, based on measured voxel activity. The identification algorithm is described in the main text. For details of voxel selection, performance for different set sizes, and noise ceiling estimation, see Supplementary Fig. 4 and Supplementary Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 June 2007; accepted 17 January 2008.
Published online 5 March 2008.

1. Haynes, J. D. & Rees, G. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neurosci.* **8,** 686 691 (2005).
2. Kamitani, Y. & Tong, F. Decoding the visual and subjective contents of the human brain. *Nature Neurosci.* **8,** 679 685 (2005).
3. Thirion, B. *et al.* Inverse retinotopy: inferring the visual content of images from brain activation patterns. *Neuroimage* **33,** 1104 1116 (2006).
4. Cox, D. D. & Savoy, R. L. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage* **19,** 261 270 (2003).
5. Haxby, J. V. *et al.* Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293,** 2425 2430 (2001).
6. Haynes, J. D. & Rees, G. Decoding mental states from brain activity in humans. *Nature Rev. Neurosci.* **7,** 523 534 (2006).
7. Hung, C. P., Kreiman, G., Poggio, T. & DiCarlo, J. J. Fast readout of object identity from macaque inferior temporal cortex. *Science* **310,** 863 866 (2005).
8. Tsao, D. Y., Freiwald, W. A., Tootell, R. B. & Livingstone, M. S. A cortical region consisting entirely of face selective cells. *Science* **311,** 670 674 (2006).
9. Simoncelli, E. P. & Olshausen, B. A. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24,** 1193 1216 (2001).
10. Wu, M. C., David, S. V. & Gallant, J. L. Complete functional characterization of sensory neurons by system identification. *Annu. Rev. Neurosci.* **29,** 477 505 (2006).
11. Daugman, J. G. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two dimensional visual cortical filters. *J. Opt. Soc. Am. A* **2,** 1160 1169 (1985).
12. Jones, J. P. & Palmer, L. A. An evaluation of the two dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiol.* **58,** 1233 1258 (1987).
13. Lee, T. S. Image representation using 2D Gabor wavelets. *IEEE Trans. Pattern Anal.* **18,** 959 971 (1996).
14. DeYoe, E. A. *et al.* Mapping striate and extrastriate visual areas in human cerebral cortex. *Proc. Natl Acad. Sci. USA* **93,** 2382 2386 (1996).
15. Dumoulin, S. O. & Wandell, B. A. Population receptive field estimates in human visual cortex. *Neuroimage* **39,** 647 660 (2008).
16. Engel, S. A. *et al.* fMRI of human visual cortex. *Nature* **369,** 525 (1994).
17. Hansen, K. A., David, S. V. & Gallant, J. L. Parametric reverse correlation reveals spatial linearity of retinotopic human V1 BOLD response. *Neuroimage* **23,** 233 241 (2004).
18. Sereno, M. I. *et al.* Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* **268,** 889 893 (1995).
19. Smith, A. T., Singh, K. D., Williams, A. L. & Greenlee, M. W. Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cereb. Cortex* **11,** 1182 1190 (2001).
20. Sasaki, Y. *et al.* The radial bias: a different slant on visual orientation sensitivity in human and nonhuman primates. *Neuron* **51,** 661 670 (2006).
21. Olman, C. A., Ugurbil, K., Schrater, P. & Kersten, D. BOLD fMRI and psychophysical measurements of contrast response to broadband images. *Vision Res.* **44,** 669 683 (2004).
22. Singh, K. D., Smith, A. T. & Greenlee, M. W. Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. *Neuroimage* **12,** 550 564 (2000).
23. Haynes, J. D. & Rees, G. Predicting the stream of consciousness from activity in human visual cortex. *Curr. Biol.* **15,** 1301 1307 (2005).
24. Heeger, D. J. & Ress, D. What does fMRI tell us about neuronal activity? *Nature Rev. Neurosci.* **3,** 142 151 (2002).
25. Logothetis, N. K. & Wandell, B. A. Interpreting the BOLD signal. *Annu. Rev. Physiol.* **66,** 735 769 (2004).
26. Stanley, G. B., Li, F. F. & Dan, Y. Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *J. Neurosci.* **19,** 8036 8042 (1999).
27. Haynes, J. D., Lotto, R. B. & Rees, G. Responses of human visual cortex to uniform surfaces. *Proc. Natl Acad. Sci. USA* **101,** 4286 4291 (2004).
28. Rainer, G., Augath, M., Trinath, T. & Logothetis, N. K. Nonmonotonic noise tuning of BOLD fMRI signal to natural images in the visual cortex of the anesthetized monkey. *Curr. Biol.* **11,** 846 854 (2001).
29. Salinas, E. & Abbott, L. F. Vector reconstruction from firing rates. *J. Comput. Neurosci.* **1,** 89 107 (1994).
30. Zhang, K., Ginzburg, I., McNaughton, B. L. & Sejnowski, T. J. Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. *J. Neurophysiol.* **79,** 1017 1044 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by a National Defense Science and Engineering Graduate fellowship (K.N.K.), the National Institutes of Health, and University of California, Berkeley intramural funds. We thank B. Inglis for assistance with MRI, K. Hansen for assistance with retinotopic mapping, D. Woods and X. Kang for acquisition of whole brain anatomical data, and A. Rokem for assistance with scanner operation. We also thank C. Baker, M. D'Esposito, R. Ivry, A. Landau, M. Merolle and F. Theunissen for comments on the manuscript. Finally, we thank S. Nishimoto, R. Redfern, K. Schreiber, B. Willmore and B. Yu for their help in various aspects of this research.

Author Contributions K.N.K. designed and conducted the experiment and was first author on the paper. K.N.K. and T.N. analysed the data. R.J.P. provided mathematical ideas and assistance. J.L.G. provided guidance on all aspects of the project. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.L.G. (gallant@berkeley.edu).

METHODS

Stimuli. The stimuli consisted of sequences of natural photographs. Photographs were obtained from a commercial digital library (Corel Stock Photo Libraries from Corel Corporation), the Berkeley Segmentation Dataset (http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/) and the authors' personal collections. The content of the photographs included animals, buildings, food, humans, indoor scenes, manmade objects, outdoor scenes, and textures. Photographs were converted to greyscale, downsampled so that the smaller of the two image dimensions was 500 pixels, linearly transformed so that the 1/10th and 99 9/10th percentiles of the original pixel values were mapped to the minimum (0) and maximum (255) pixel values, cropped to the central 500 pixels × 500 pixels, masked with a circle, and placed on a grey background (Supplementary Fig. 1a). The luminance of the background was set to the mean luminance across photographs, and the outer edge of each photograph (10% of the radius of the circular mask) was linearly blended into the background.

The size of the photographs was 20° × 20° (500 pixels × 500 pixels). A central white square served as the fixation point, and its size was 0.2° × 0.2° (4 pixels × 4 pixels). Photographs were presented in successive 4 s trials; in each trial, a photograph was presented for 1 s and the grey background was presented for 3 s. Each 1 s presentation consisted of a photograph being flashed ON OFF ON OFF ON where ON corresponds to presentation of the photograph for 200 ms and OFF corresponds to presentation of the grey background for 200 ms (Supplementary Fig. 1b). The flashing technique increased the signal to noise ratio of voxel responses relative to that achieved by presenting each photograph continuously for 1 s (data not shown).

Visual stimuli were delivered using the VisuaStim goggles system (Resonance Technology). The display resolution was 800 × 600 at 60 Hz. A PowerBook G4 computer (Apple Computer) controlled stimulus presentation using software written in MATLAB 5.2.1 (The Mathworks) and Psychophysics Toolbox 2.53 (http://psychtoolbox.org).

MRI parameters. The experimental protocol was approved by the University of California, Berkeley Committee for the Protection of Human Subjects. MRI data were collected at the Brain Imaging Center at University of California, Berkeley using a 4 T INOVA MR scanner (Varian, Inc.) and a quadrature transmit/receive surface coil (Midwest RF, LLC). Data were acquired using coronal slices that covered occipital cortex: 18 slices, slice thickness 2.25 mm, slice gap 0.25 mm, field of view 128 mm × 128 mm. (In one scan session, a slice gap of 0.5 mm was used.) For functional data, a T2* weighted, single shot, slice interleaved, gradient echo EPI pulse sequence was used: matrix size 64 × 64, TR 1 s, TE 28 ms, flip angle 20°. The nominal spatial resolution of the functional data was 2 mm × 2 mm × 2.5 mm. For anatomical data, a T1 weighted gradient echo multislice sequence was used: matrix size 256 × 256, TR 0.2 s, TE 5 ms, flip angle 40°.

Data collection. Data for the model estimation and image identification stages of the experiment were collected in the same scan sessions. Two subjects were used: S1 (author T.N., age 33) and S2 (author K.N.K., age 25). Subjects were healthy and had normal or corrected to normal vision.

Five scan sessions of data were collected from each subject. Each scan session consisted of five model estimation runs and two image identification runs. Model estimation runs (11 min each) were used for the model estimation stage of the experiment. Each model estimation run consisted of 70 distinct images presented two times each. Image identification runs (12 min each) were used for the image identification stage of the experiment. Each image identification run consisted of 12 distinct images presented 13 times each. Images were randomly selected for each run and were mutually exclusive across runs. The total number of distinct images used in the model estimation and image identification runs was 1,750 and 120, respectively. (For additional details on experimental design, see Supplementary Methods.)

Three additional scan sessions of data were collected from subject S1. Two of these were held approximately two months after the main experiment, and consisted of five image identification runs each. The third was held approximately 14 months after the main experiment, and consisted of one image identification run. The images used in these additional scan sessions were randomly selected and were distinct from the images used in the main experiment.

Data pre processing. Functional brain volumes were reconstructed and then co registered to correct differences in head positioning within and across scan sessions. Next, voxel specific response time courses were estimated and deconvolved from the time series data. This produced, for each voxel, an estimate of the amplitude of the response (a single value) to each image used in the model estimation and image identification runs. Finally, voxels were assigned to visual areas based on retinotopic mapping data[17] collected in separate scan sessions. (Details of these procedures are given in Supplementary Methods.)

Model estimation. A receptive field model was estimated for each voxel based on its responses to the images used in the model estimation runs. The model was based on a Gabor wavelet pyramid[11–13]. In the model, each image is represented by a set of Gabor wavelets differing in size, position, orientation, spatial frequency and phase (Supplementary Fig. 2). The predicted response is a linear function of the contrast energy contained in quadrature wavelet pairs (Supplementary Fig. 3). Because contrast energy is a nonlinear quantity, this is a linearized model[10]. The model was able to characterize responses of voxels in visual areas V1, V2 and V3 (Supplementary Table 1), but it did a poor job of characterizing responses in higher visual areas such as V4.

Alternative receptive field models were also used, including the retinotopy only model and several constrained versions of the Gabor wavelet pyramid model. Details of these models and model estimation procedures are given in Supplementary Methods.

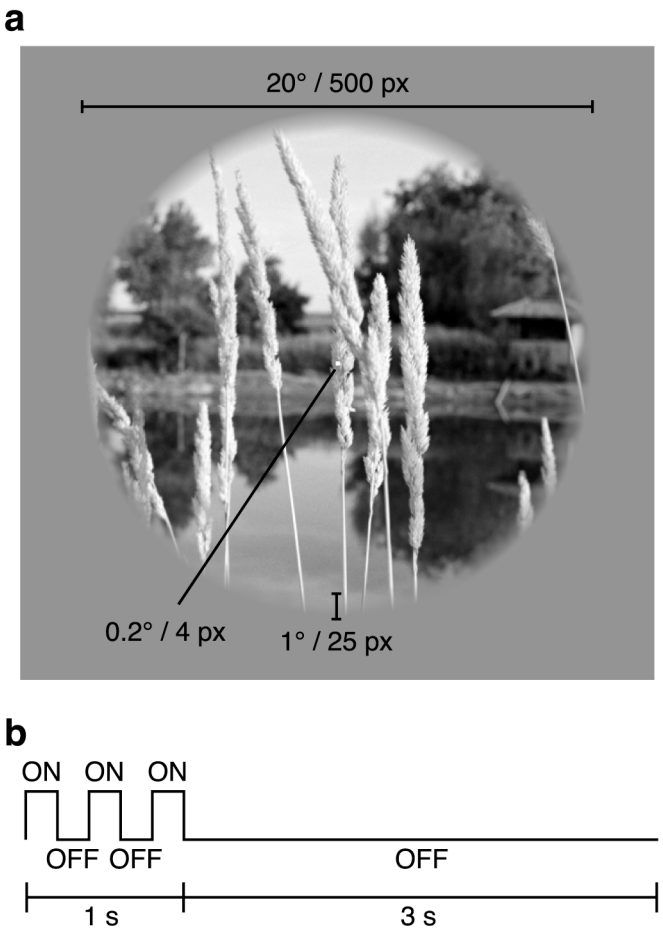
Image identification. Voxel activity patterns were constructed from voxel responses evoked by the images used in the image identification runs. For each voxel activity pattern, the estimated receptive field models were used to identify which specific image had been seen. The identification algorithm is described in the main text. See Supplementary Fig. 4 and Supplementary Methods for details of voxel selection, performance for different set sizes, and noise ceiling estimation. See Supplementary Discussion for a comparison of identification with the decoding problems of classification and reconstruction.

SUPPLEMENTARY INFORMATION

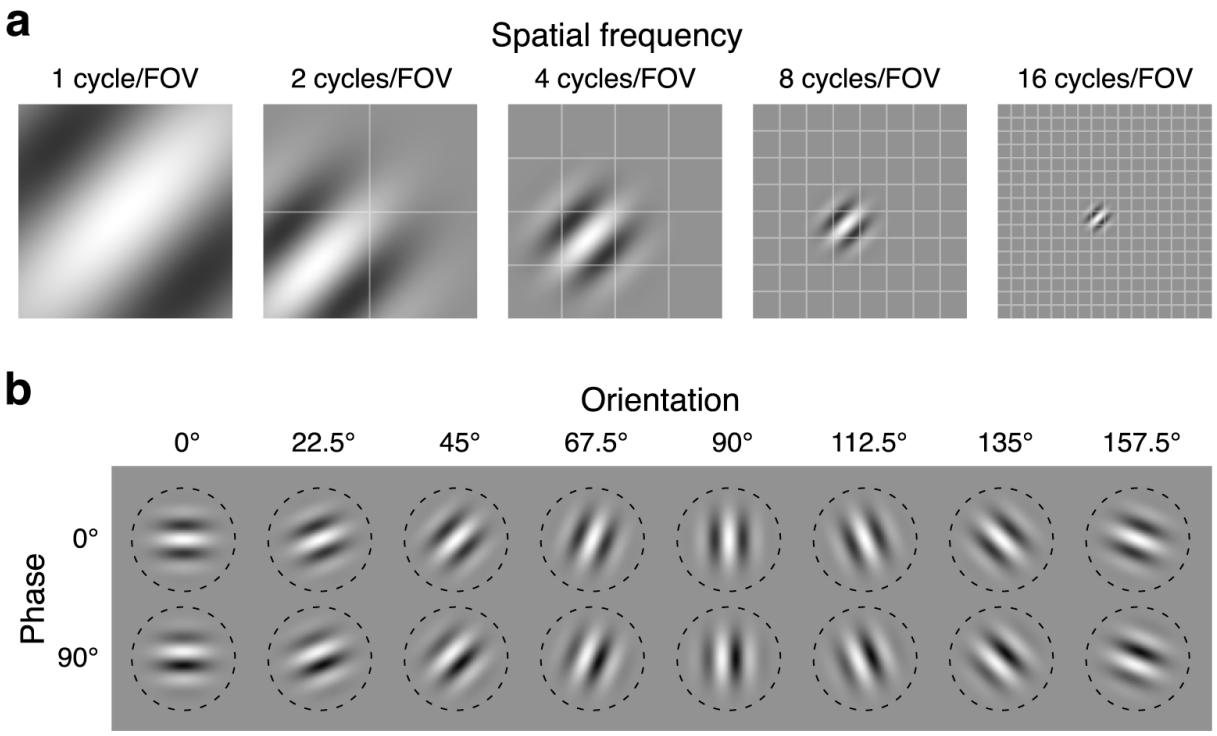
Identifying natural images from human brain activity

Kendrick N. Kay[1], Thomas Naselaris[2], Ryan J. Prenger[3] & Jack L. Gallant[1,2]
[1]Department of Psychology, [2]Helen Wills Neuroscience Institute, [3]Department of Physics,
University of California, Berkeley, California 94720, USA

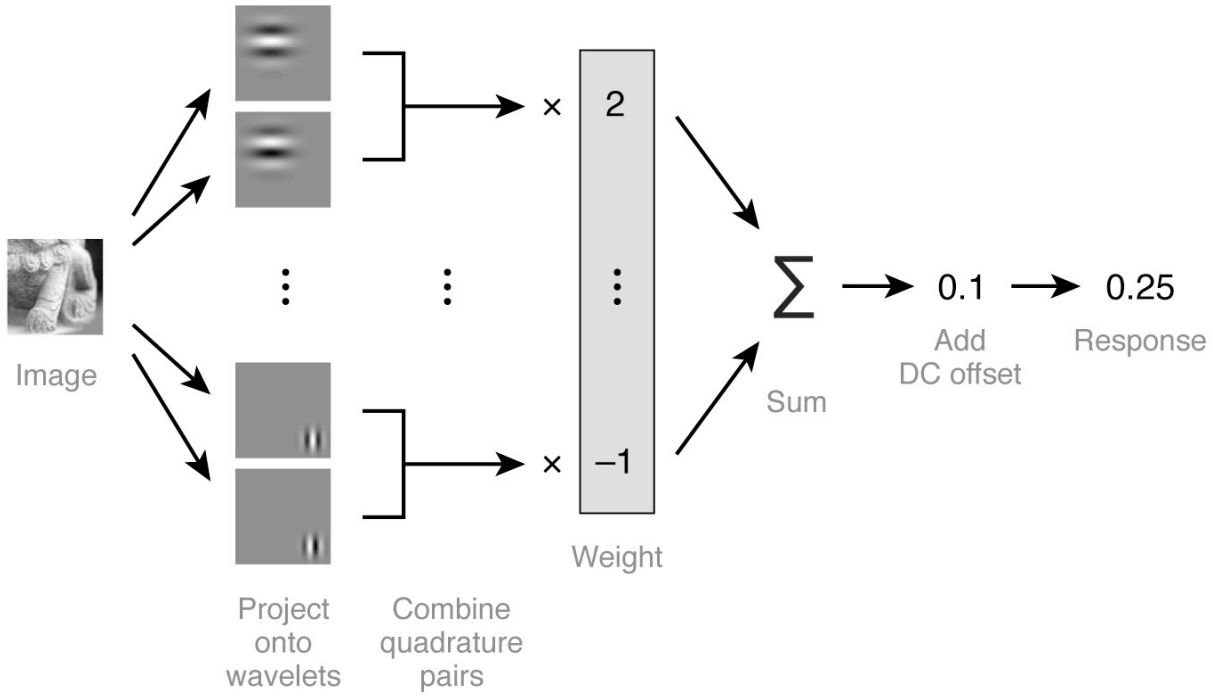
Overview



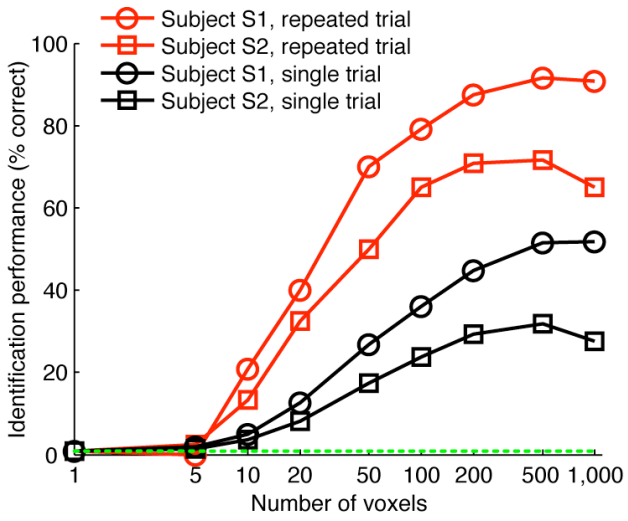
Supplementary Figure 1. Stimulus design. The stimuli consisted of sequences of grayscale natural photographs. **a**, Spatial characteristics. The photographs were masked with a circle (20° diameter) and placed on a gray background. The outer edge of each photograph (1° width) was linearly blended into the background. A central white square (0.2° side length) served as the fixation point. **b**, Temporal characteristics. The photographs were presented for 1 s with a delay of 3 s between successive photographs. Each 1-s presentation consisted of a photograph being flashed ON–OFF–ON–OFF–ON where ON corresponds to presentation of the photograph for 200 ms and OFF corresponds to presentation of the gray background for 200 ms.



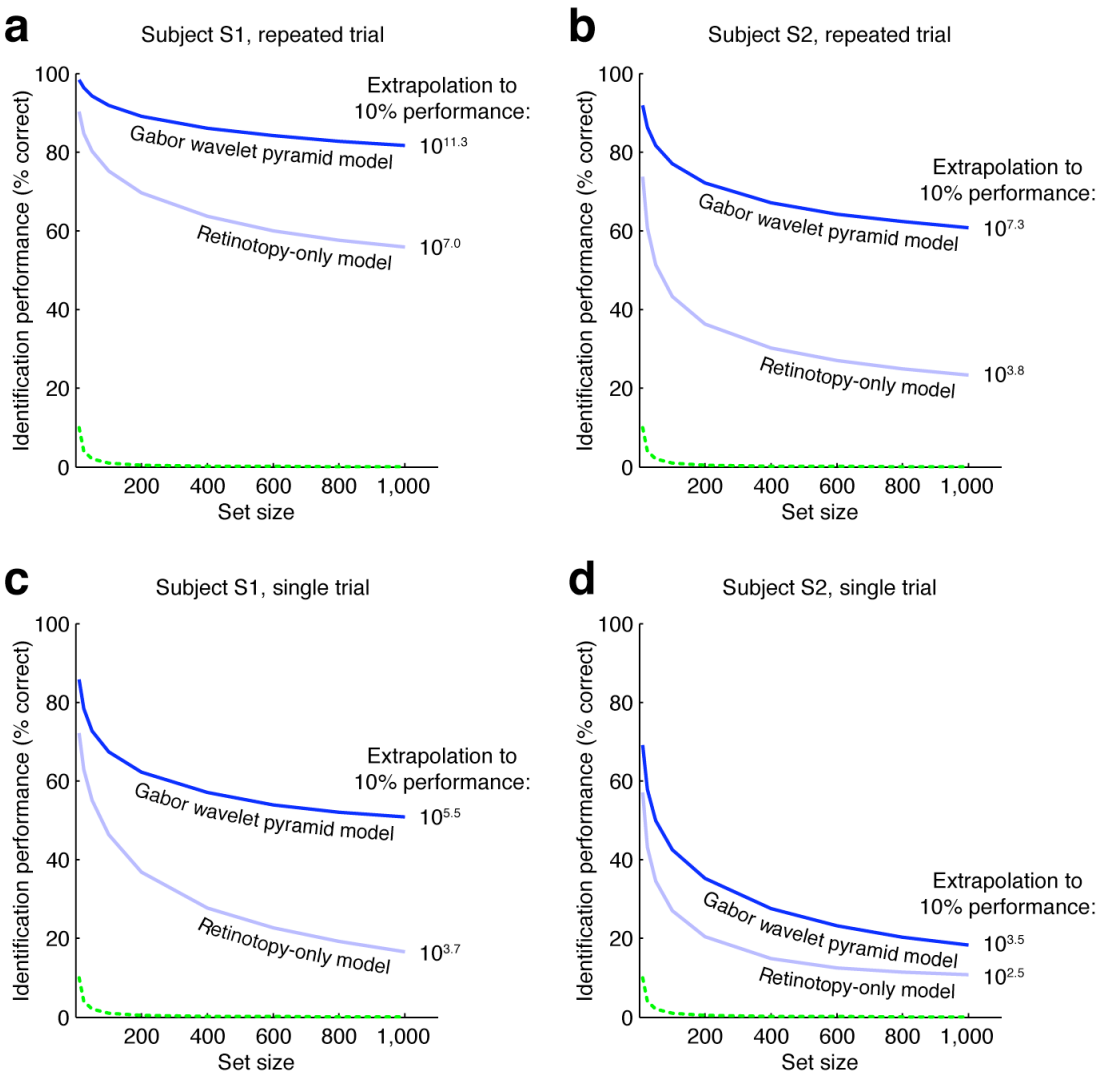
Supplementary Figure 2. Gabor wavelet pyramid design. The receptive-field model used in the present study is based on a Gabor wavelet pyramid[11–13]. **a**, Spatial frequency and position. Wavelets occur at five (or, in some cases, six) spatial frequencies. This panel depicts one wavelet at each of the first five spatial frequencies. At each spatial frequency f cycles per field-of-view (FOV), wavelets are positioned on an $f \times f$ grid, as indicated by the translucent lines. **b**, Orientation and phase. At each grid position, wavelets occur at eight orientations and two phases. This panel depicts a complete set of wavelets for a single grid position. Dashed lines indicate the bounds of the mask associated with each wavelet.



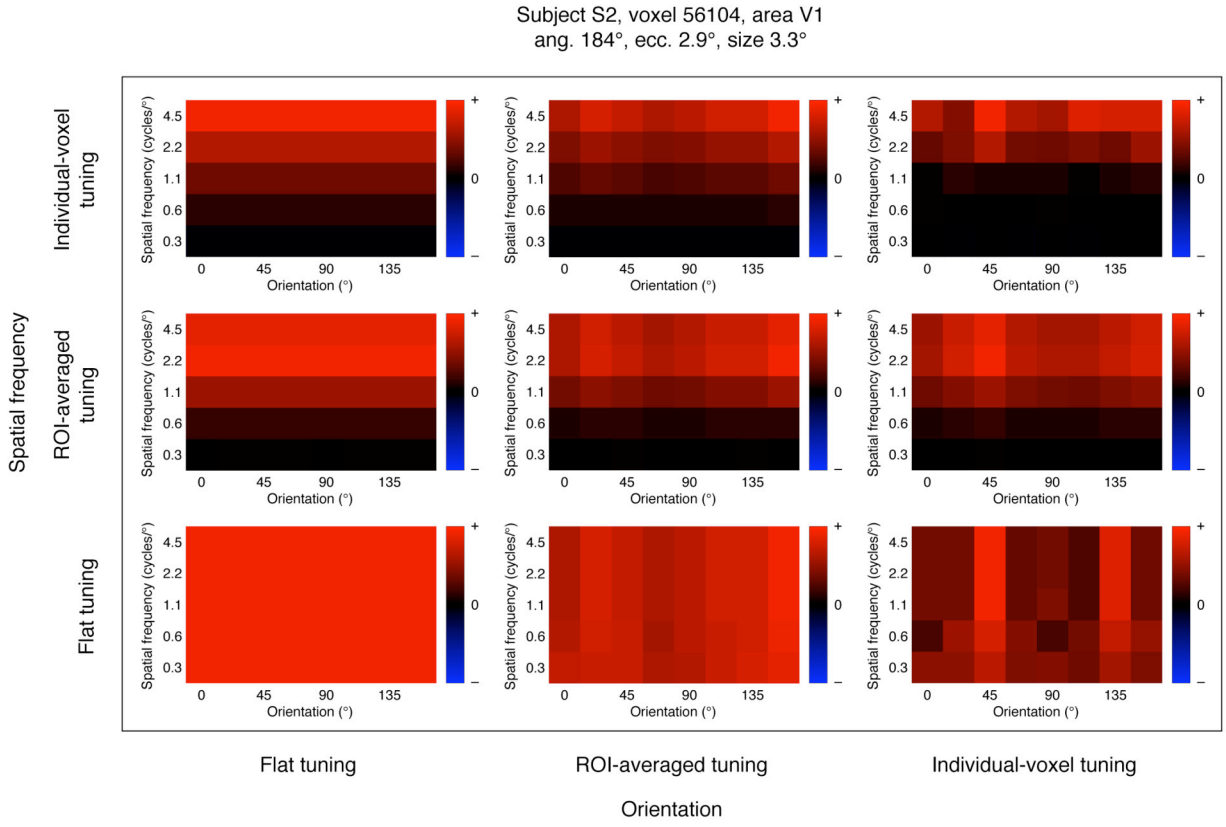
Supplementary Figure 3. Gabor wavelet pyramid model. Each image is projected onto the individual Gabor wavelets comprising the Gabor wavelet pyramid (see Supplementary Fig. 2). The projections for each quadrature pair of wavelets are squared, summed, and square-rooted, yielding a measure of contrast energy. The contrast energies for different quadrature wavelet pairs are weighted and then summed. Finally, a DC offset is added. The weights are determined by gradient descent with early stopping (see Supplementary Methods 6).



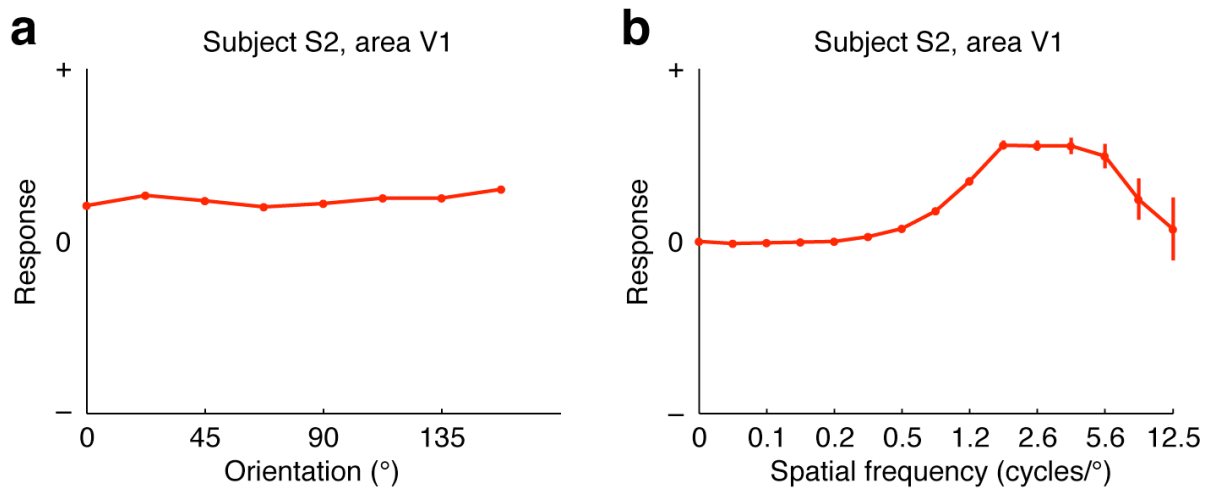
Supplementary Figure 4. Effect of number of voxels on identification performance. To optimize performance of the identification algorithm, we preferentially selected voxels whose receptive-field models had the highest predictive power (see Supplementary Methods 7). In this figure the *x* axis indicates the number of voxels selected and the *y* axis indicates identification performance. The dashed green line indicates chance performance, and results were obtained for a set size of 120 images. In all cases optimal performance was achieved using about 500 voxels. Therefore, all identification results in this study were obtained using 500 voxels.



Supplementary Figure 5. Identification performance for the retinotopy-only model. To determine whether identification is a mere consequence of the retinotopic organization of early visual areas, we evaluated an alternative retinotopy-only model that captures the location and size of each voxel's receptive field but discards orientation and spatial frequency information. **a**, Comparison of identification performance for the retinotopy-only (RO) model and the Gabor wavelet pyramid (GWP) model (results for subject S1 and repeated trials). The *x* axis indicates set size and the *y* axis indicates identification performance. The number to the right of each line gives the estimated set size at which performance declines to 10% correct, and the dashed green line indicates chance performance. Performance for the RO model was substantially lower than for the GWP model. **b**, Results for subject S2 and repeated trials. Once again the RO model performed substantially worse than the GWP model. **c–d**, Single-trial results for subjects S1 and S2. Although identification performance was poorer overall when single trials were used, the GWP model still outperformed the RO model. These results collectively indicate that spatial tuning alone does not yield optimal identification performance; identification improves substantially when orientation and spatial frequency tuning are included in the model.

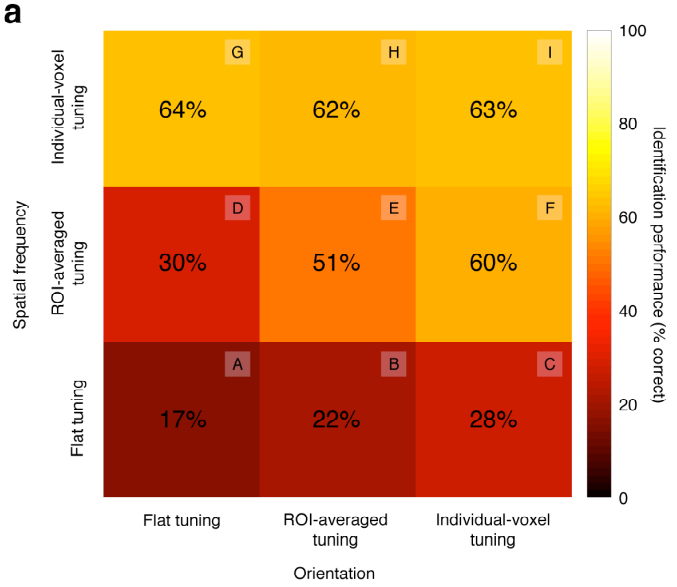


Supplementary Figure 6. Example of constraints on orientation and spatial frequency tuning. To assess the individual contributions of orientation and spatial frequency tuning to identification performance, we evaluated several constrained versions of the Gabor wavelet pyramid model. These models were constructed by fixing the spatial envelope of each voxel and then imposing different constraints on orientation and spatial frequency tuning (see Supplementary Methods 9 for details). This figure illustrates the tuning of one representative voxel under the various models. Nine plots are arranged in three columns and three rows. Each plot depicts the joint orientation and spatial frequency tuning obtained under one specific model (format is the same as in Fig. 2b). The three columns represent different constraints on orientation tuning: in the left column it is constrained to be flat; in the middle column it is constrained to match the mean orientation tuning across voxels in the corresponding region-of-interest (i.e. V1, V2, or V3); in the right column it is unconstrained (the model is allowed full flexibility in orientation tuning). The three rows represent different constraints on spatial frequency tuning: in the bottom row it is constrained to be flat; in the middle row it is constrained to match the mean spatial frequency tuning across voxels in the corresponding region-of-interest; in the top row it is unconstrained. These plots demonstrate that the models successfully incorporate the intended tuning constraints. (In the bottom-right plot orientation tuning at low spatial frequencies is not perfectly matched to the marginal orientation tuning. This is a consequence of the fact that the lowest-frequency wavelets are truncated by the field-of-view, effectively increasing their spectral bandwidth.)



Supplementary Figure 7. Example of ROI-averaged tuning curves. Several of the constrained versions of the Gabor wavelet pyramid model involve fixing the orientation or spatial frequency tuning curve of a voxel to match the mean tuning curve across voxels in the corresponding region-of-interest (i.e. V1, V2, or V3). **a**, Example ROI-averaged orientation tuning curve for area V1. The *x* axis indicates orientation and the *y* axis indicates predicted response. Error bars indicate ± 1 s.e.m. across voxels (bootstrap procedure). The orientation tuning curve is nearly flat. **b**, Example ROI-averaged spatial frequency tuning curve for area V1. The format is the same as panel a, except that the *x* axis indicates spatial frequency. The spatial frequency tuning curve is band-pass.

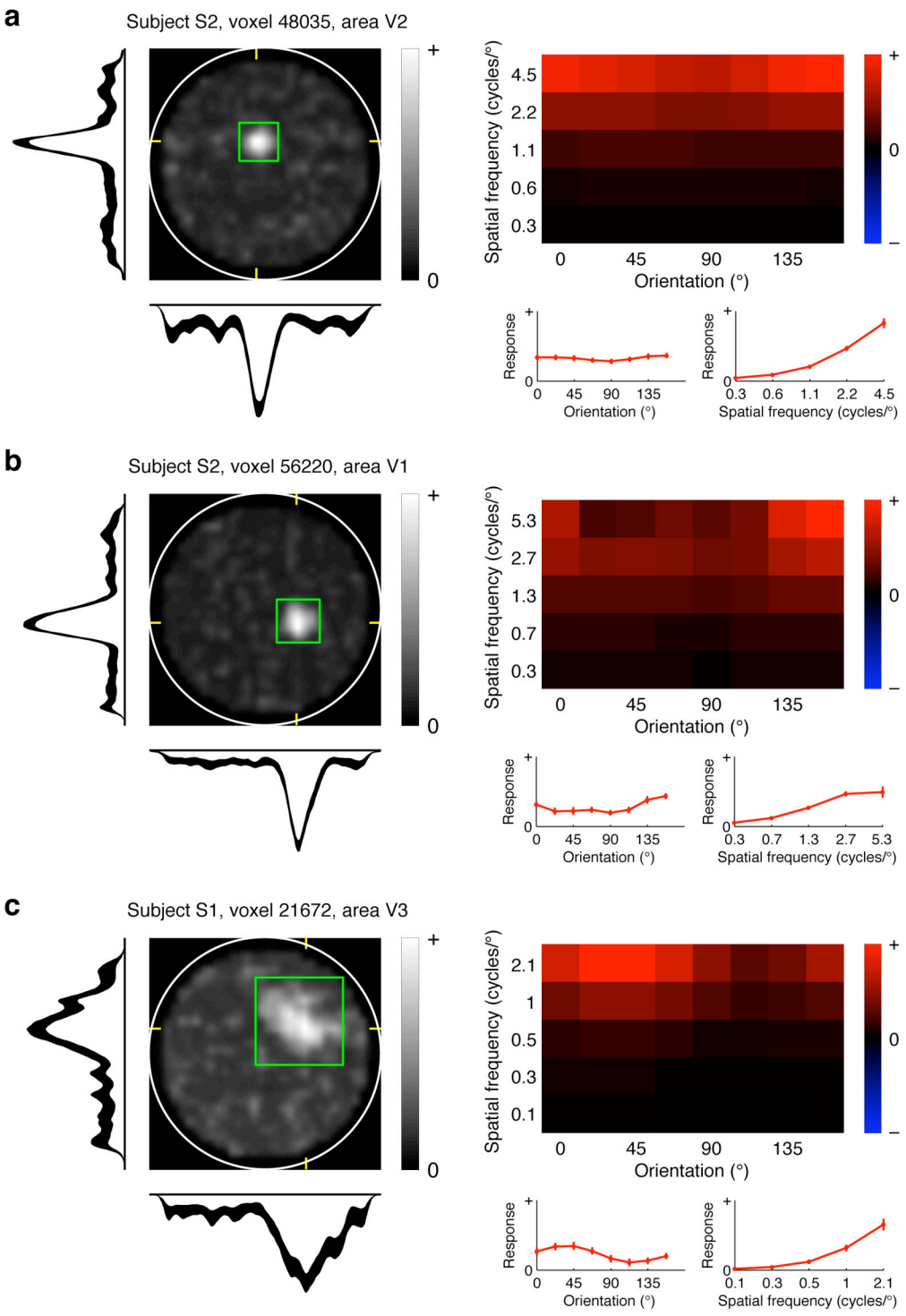
a



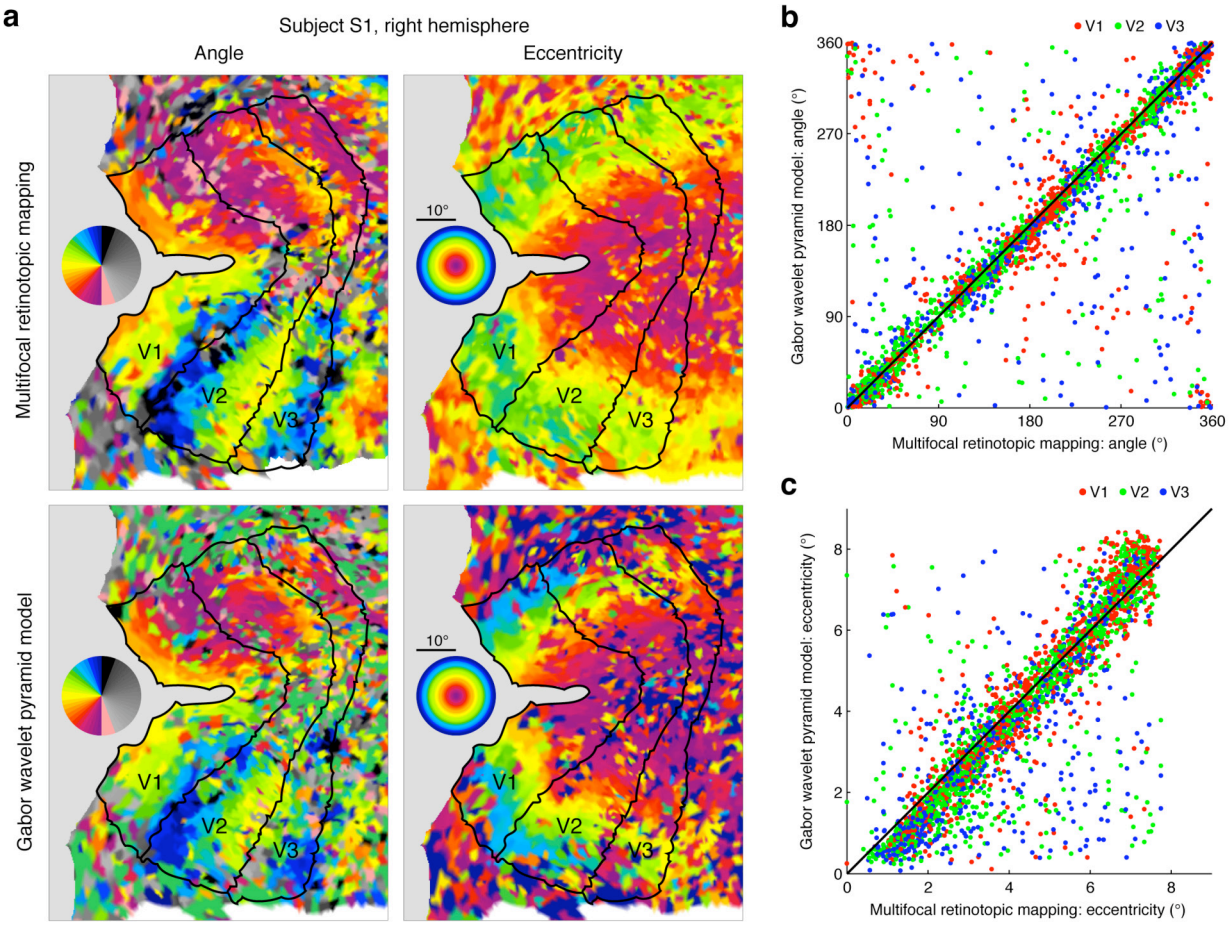
b

	A	B	C	D	E	F	G	H	I
A	–	0.03	8e-6	2e-4	7e-16	6e-21	2e-22	1e-21	2e-21
B	–	–	3e-3	0.03	4e-14	2e-18	5e-20	7e-19	3e-18
C	–	–	–	0.4	4e-10	9e-17	2e-16	2e-15	3e-15
D	–	–	–	–	1e-9	2e-16	3e-16	7e-18	3e-19
E	–	–	–	–	–	4e-4	5e-6	4e-5	8e-5
F	–	–	–	–	–	–	0.08	0.3	0.2
G	–	–	–	–	–	–	–	–	–
H	–	–	–	–	–	–	0.3	–	0.5
I	–	–	–	–	–	–	0.4	–	–

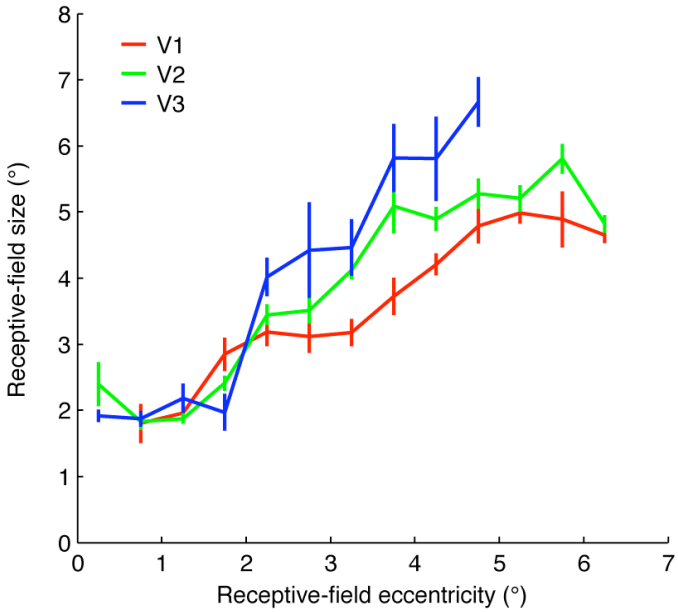
Supplementary Figure 8. Contribution of orientation and spatial frequency tuning to identification performance. Constrained versions of the Gabor wavelet pyramid model were used to investigate the individual contributions of orientation and spatial frequency tuning to identification performance (see Supplementary Fig. 6). **a**, Summary of identification performance under each model. The nine models are labeled by capital letters, and are arranged in three columns and three rows. Different columns represent different constraints on orientation tuning, and different rows represent different constraints on spatial frequency tuning (as in Supplementary Fig. 6). Colors and percentages denote identification performance achieved under each model (repeated trial, 1,000 images, performance averaged across subjects). Both orientation and spatial frequency tuning contribute to identification performance (C > A and G > A), but spatial frequency tuning is relatively more important (G > C). Voxel-to-voxel variation in orientation and spatial frequency tuning also contributes to identification performance (F > E and H > E). **b**, Statistical comparisons of identification performance. This table provides *p*-values for all pairwise model comparisons (one-tailed paired sign test, *p*-values rounded up). A red *p*-value indicates that the model in the corresponding column performed significantly better than the model in the corresponding row ($p < 0.05$), while a black *p*-value indicates that the improvement was not statistically significant ($p \geq 0.05$). The symbol '–' indicates that performance for the column model was less than or equal to that for the row model. The differences in identification performance noted in panel a are all statistically significant.



Supplementary Figure 9. Additional examples of receptive-field models. a–c, Receptive-field models for three representative voxels. The format of each panel is the same as that of Fig. 2. Receptive-field (RF) location, size, orientation tuning, and spatial frequency tuning all vary substantially across voxels. The RFs also vary in reliability; for example, the RF shown in panel c exhibits less reliable spatial tuning than the RFs shown in panels a–b.



Supplementary Figure 10. Validation of retinotopic information derived from receptive-field models. Since retinotopy is a well-established property of voxels in early visual areas[14,16,18], one way to validate the Gabor wavelet pyramid (GWP) model is to confirm that it produces reasonable estimates of voxel receptive-field location. In this figure we compare angle and eccentricity estimates obtained from the GWP model with those obtained from the multifocal (MF) retinotopic mapping technique[17,31] (see Supplementary Methods 11). Note that the data used for the MF technique were completely independent of the data used for the GWP model. **a**, Comparison of retinotopic maps for a representative hemisphere. Voxel data were assigned to surface vertices using nearest neighbor interpolation, and the maps were not smoothed or thresholded. Black lines indicate the boundaries of visual areas V1, V2, and V3. (The same boundaries are replicated on each map.) Overall, the GWP maps are similar to the MF maps and exhibit the typical retinotopic organization[32,33]. The GWP maps are somewhat noisier than the MF maps, which is expected given that the MF technique is specifically optimized to provide retinotopic information. **b**, Quantitative comparison of angle estimates. Dots represent individual voxels taken across subjects (voxels for which the predictive power of the GWP model was not statistically significant at $p < 0.01$ are omitted). Notice that the MF and GWP angle estimates are well matched. **c**, Quantitative comparison of eccentricity estimates (format same as panel b). The MF and GWP eccentricity estimates are generally well matched, but there appear to be systematic discrepancies at the lowest and highest eccentricities. The likely cause of the discrepancies is the spatial granularity of the stimuli used for MF mapping[32].



Supplementary Figure 11. Relationship between receptive-field size and eccentricity. In the course of fitting the Gabor wavelet pyramid model, estimates of the location and size of each voxel's receptive field (RF) were obtained. We examined the relationship between RF size and eccentricity to see if the expected pattern of results could in fact be demonstrated. In this figure the *x* axis indicates RF eccentricity and the *y* axis indicates RF size. (RF size is defined as ± 2 s.d. of a fitted two-dimensional Gaussian; see Supplementary Methods 5.) Voxels were pooled across subjects and then binned by eccentricity. (To ensure robust results, voxels for which RF predictive power was not statistically significant at $p < 0.01$ or for which estimated RF location was not completely within the stimulus bounds were omitted before pooling.) For each bin with at least 10 voxels, the median RF size is plotted, with error bars indicating ± 1 s.e. (bootstrap procedure). RF size increases with eccentricity and across visual areas, consistent with previous fMRI studies[15,19,34–36]. The fact that our model estimation approach uncovers differences in RF size across areas suggests that it could potentially reveal other area differences.

Subject	*Visual area*	*Total number of voxels*	*High SNR (% of total)*	*High predictive power (% of total)*	*High SNR and high predictive power (% of high SNR)*
S1	V1	1331	431 (32%)	533 (40%)	406 (94%)
	V2	2208	659 (30%)	677 (31%)	558 (85%)
	V3	1973	425 (22%)	343 (17%)	260 (61%)
S2	V1	1513	275 (18%)	382 (25%)	256 (93%)
	V2	1982	369 (19%)	426 (21%)	291 (79%)
	V3	1780	223 (13%)	224 (13%)	138 (62%)

Supplementary Table 1. Signal-to-noise ratio of voxel responses and predictive power of receptive-field models. The column *High SNR (% of total)* gives the number of voxels with a signal-to-noise ratio (SNR) greater than 1.5; *High predictive power (% of total)* gives the number of voxels for which the predictive power of the best initial model was statistically significant ($p < 0.01$, bootstrap procedure); and *High SNR and high predictive power (% of high SNR)* gives the number of voxels that satisfied both criteria. (See Supplementary Methods 3 and 6 for details concerning SNR and predictive power, respectively.) Although SNR varied greatly across subjects, SNR was fairly consistent for areas V1, V2, and V3 within each subject. Predictive power generally decreased from V1 to V2 to V3, likely reflecting the fact that the Gabor wavelet pyramid model is not optimal for visual areas beyond V1.

Supplementary Discussion 1. Classification-based decoding methods cannot be used to identify novel images

Previous classification-based studies did not identify novel images

Several fMRI studies of visual cortex[4,5,37,38] have shown that classification-based decoding methods can be used to determine the category of an image seen by an observer, even if the image is a novel instance of the category. In addition, one neurophysiological study of inferotemporal neurons[7] showed that classification methods can be used to determine which object was seen by an observer, even if the object was presented at novel positions or scales. At a superficial level these results may seem to contradict our claim that classification methods cannot be used to identify novel images. However, there are two key differences between these previous studies and the present study. First, the previous studies achieved decoding for only specific kinds of novel images (e.g. novel images drawn from fixed categories). In contrast the present study achieves decoding for arbitrary novel natural images.

Second, the previous studies demonstrated classification, not identification. The goal of classification is to discriminate images belonging to a given category from those belonging to other categories. Classification thus aggregates over the individual images belonging to a given category. In contrast, the goal of identification is to discriminate an individual image from a number of other images. Identification thus treats each image as a distinct entity. To illustrate these ideas, consider a hypothetical experiment that measures brain activity evoked by an image of a dog. The goal of classification is to assign the image to one of several pre-defined categories such as *dog* or *cat*; the goal of identification is to discriminate the specific dog image from a number of other images (regardless of category membership).

Limitations of classification-based decoding methods

Classification-based decoding methods are inherently limited by the fixed set of categories that are used in training. For example, suppose a classifier is trained to discriminate brain activity evoked by dogs from that evoked by cats; without additional training the classifier would be unable to discriminate brain activity evoked by birds from that evoked by dogs or cats. This limited generality entails that classification methods cannot be used to identify novel images. To illustrate: suppose we adapt the classification framework to the problem of identification by treating each individual image as if it defines a unique category[3,7,8]. If previous measurements of brain activity evoked by each image are available for training purposes, standard classification procedures can achieve identification. However, in the case of novel images (i.e. no previous measurements of brain activity evoked by the images are available), we are faced with a critical problem: how do we perform classification for categories we have not trained for? (For additional discussion of the limitations of classification methods, see ref. 3.)

An extension of classification-based decoding methods yields poor identification performance

Is it possible to extend classification-based decoding methods to achieve identification of novel images? To address this question we developed a straightforward extension of classification methods. In this analysis we treated each image used in the model estimation stage of the

experiment as if it defined a unique category (similar to refs. 3, 7, 8). Thus, the 1,750 voxel activity patterns measured in the model estimation stage of the experiment were taken to represent 1,750 unique categories. We call these the *category activity patterns*.

For each of the 120 voxel activity patterns measured in the image identification stage of the experiment, we attempted to identify which specific image had been seen. This was accomplished by taking a given voxel activity pattern *m* and finding the category activity pattern most similar to *m* (similarity was quantified by Pearson's *r*). We call the image associated with the found category activity pattern the *matched image*. (Intuitively, the matched image is the image from the model estimation stage of the experiment that is "brain-wise" most similar to the image seen by the subject.) The matched image was then compared with each of the 120 images used in the image identification stage of the experiment, and the image most similar to the matched image was selected. Two metrics for image similarity were tested: correlation of pixel luminance and correlation of local contrast. (To calculate the local contrast of a given image, the image was divided into $n° \times n°$ blocks and root-mean-square contrast was calculated for each block. The results reported below were obtained using the value of *n* that yielded the best performance, $n = 0.6$.)

Identification performance using the pixel luminance metric was 1.7% (2/120) and 0.8% (1/120) for subjects S1 and S2, respectively (repeated trial). These values were not significantly above chance ($p \geq 0.05$, one-tailed binomial test). Identification performance using the local contrast metric was 5% (6/120) and 5.8% (7/120) for subjects S1 and S2, respectively (repeated trial). These values were above chance ($p < 0.0001$, one-tailed binomial test) but far below the performance levels achieved by the identification algorithm described in the main text (92% and 72% for subjects S1 and S2, respectively). These results suggest that classification methods cannot be easily extended to achieve accurate identification of novel images.

Supplementary Discussion 2. Comparison of classification, identification, and reconstruction

The problems of classification, identification, and reconstruction can be defined formally. Let x_1, x_2, x_3, ... represent different images. (There may be an infinite number of images.) Let l represent a function that maps images to a certain set of labels. For example, $l(x_i)$ is the label assigned to image x_i. Let p_i represent an activity pattern evoked by image x_i on a given trial. We define the following problems:

- *Classification*: given activity pattern p_i, determine $l(x_i)$.
- *Identification*: given activity pattern p_i and a finite set of images (e.g. $\{x_2\ x_7\ x_3\}$) such that x_i is a member of the set, determine x_i.
- *Reconstruction*: given activity pattern p_i, determine x_i.

At the most general level the three problems are similar: in each case the goal is to infer certain information based on a given activity pattern. In theory, identification can be considered a special case of classification where the label assigned to an image is simply the index of that image in the given set of images. However, classification normally refers to the case where a single label is assigned to multiple images, so in practice identification is distinct from classification. Furthermore, although the goal of both identification and reconstruction is to determine the specific image that had evoked a given activity pattern, in identification a set of potential images is provided whereas in reconstruction no such set is provided.

Note that these definitions do not specify what information is available to train a decoder, though this is an important issue in the present study. Unlike classification-based methods, our decoding method can achieve accurate identification of an image even when that image is novel, i.e. even when brain activity evoked by the image is not available for training.

Supplementary Discussion 3. Previous research on voxel tuning properties

The receptive-field model used in the present study is based on a Gabor wavelet pyramid (GWP). The GWP has long been regarded as the standard model of how primary visual cortex (V1) represents shape[11–13]. Under the assumption that fMRI activity reflects local pooled neural activity[1,2,39–41], it is reasonable to suppose that the GWP model is appropriate for describing voxels in early visual areas. Indeed, previous results suggest that fMRI activity in V1 reflects the average activation of a population of Gabor filters[28]. The GWP model used in the present study describes tuning along the dimensions of space, orientation, and spatial frequency. Each of these dimensions has been previously investigated in fMRI.

Spatial tuning has received considerable attention from many laboratories. The phase-encoded retinotopic mapping technique was introduced in the early days of fMRI[14,16,18] and continues to be widely used. This method provides an estimate of the location of each voxel's receptive field. Recent studies have demonstrated that estimates of voxel receptive-field size can be extracted from phase-encoded data through the use of a spatial tuning model[3,15,19,35] such as a two-dimensional Gaussian. An alternative method for estimating spatial tuning is the multifocal retinotopic mapping technique where the stimulus consists of spatial elements (e.g. wedges, rings, sectors) flashed pseudorandomly across the visual field[17,31]. This method provides a more direct estimate of the spatial envelope of a voxel receptive field, but is limited by the granularity of the stimuli and by the assumption of linear spatial summation[17].

Orientation tuning has typically been investigated in fMRI by using adaptation-based techniques[42–47] or by pooling signals across many voxels[20,48]. However, recent classification-based studies have shown that individual voxels have a slight orientation bias[1,2]. These studies are also noteworthy since they demonstrate that multivariate analysis techniques can increase the amount of information extracted from fMRI data compared to conventional univariate analysis techniques.

Spatial frequency is the final dimension represented in the GWP model. Of the various dimensions, spatial frequency has been the least studied in fMRI. A few studies have shown that fMRI signals pooled across entire visual areas exhibit some spatial frequency tuning[21,22,49]. However, these studies did not investigate potential voxel-to-voxel variation in tuning.

Most fMRI experiments measure tuning along one dimension at a time. This approach assumes that stimulus dimensions are separable and that they can be measured independently of one another. In addition, fMRI experiments usually measure tuning using artificial stimuli such as gratings and checkerboard patterns (but see exceptions[21,28,50]). In the present study the GWP model is fit to voxel responses evoked by natural images. This approach measures tuning along multiple dimensions simultaneously, and produces a unified description of how images are mapped onto fMRI activity.

Supplementary Methods 1. Design of model estimation and image identification runs

The experiment consisted of two distinct stages, model estimation and image identification. Model estimation runs and image identification runs were conducted in the same fMRI scan sessions. Each estimation run used 70 distinct images presented 2 times each. Each run consisted of 168 trials, and had a duration of 168 trials × 4 s = 11.2 min. The first four and last four trials were null trials (no images presented). For the remaining 160 trials, every 8th trial was also a null trial. The presentation order of the images was determined by randomly generating a large number of sequences under the constraint that same image could not be presented on consecutive trials, and then choosing the sequence that yielded the greatest estimation efficiency[51].

Each identification run used 12 distinct images presented 13 times each. The presentation order of the images was determined by an m-sequence[52] of level 13, order 2, and length $13^2 - 1 = 168$. The m-sequence included 12 null trials (no images presented). Code for m-sequence generation was provided by T. Liu (http://fmriserver.ucsd.edu/ttliu/mttfmri_toolbox.html). During stimulus presentation the first 6 trials were repeated at the end of the 168-trial sequence. In the repeated-trial analysis, data collected during the initial 6 trials were ignored[53,54]. In the single-trial analysis, all data were used. Each run had a duration of 174 trials × 4 s = 11.6 min.

Supplementary Methods 2. Reconstruction and co-registration of brain volumes

Functional and anatomical brain volumes were reconstructed using the ReconTools software package (https://cirl.berkeley.edu/view/BIC/ReconTools). For functional volumes, a phase correction was applied to reduce Nyquist ghosting and image distortion, and differences in slice acquisition times were corrected by sinc interpolation.

All functional volumes acquired for a given subject were registered to a single spatial reference frame. Automated motion correction procedures (SPM99, http://www.fil.ion.ucl.ac.uk/spm/) were used to correct differences in head positioning within scan sessions by rigid-body transformations. Manual co-registration procedures (in-house software) were used to correct differences in head positioning across scan sessions by affine transformations. Each functional volume was resampled only once (by sinc interpolation); this minimized interpolation errors that could accumulate over multiple resamplings. No additional spatial filtering was applied to the functional volumes.

Supplementary Methods 3. Time-series pre-processing

The time-series data for each voxel were pre-processed prior to the model estimation and image identification stages of the experiment. The primary purpose of the pre-processing was to estimate and deconvolve voxel-specific response timecourses from the time-series data. This decreased the computational requirements of subsequent analyses by reducing the effective number of data points. Pre-processing was based on the basis-restricted separable (BRS) model (see Supplementary Methods 4). In brief, the BRS model uses a set of basis functions to characterize the shape of the response timecourse and a set of parameters to characterize the amplitudes of responses to different images.

During pre-processing the time-series data were analyzed both as repeated trials and as single trials. The repeated-trial analysis produced, for each voxel, an estimate of the amplitude of the response (a single value) evoked by each distinct image used in the model estimation and image identification runs. In this case each estimate reflects data from multiple image presentations. The single-trial analysis produced, for each voxel, an estimate of the amplitude of the response (a single value) evoked by each trial of the model estimation and image identification runs. In this case each estimate reflects data from a single image presentation.

Repeated-trial analysis

The following procedure was performed for each voxel in each scan session. First, the BRS model was fit to the time-series data from the model estimation runs. A set of Fourier basis functions was used to characterize the shape of the response timecourse, and a separate parameter was used to characterize the amplitude of the response to each distinct image. Fitting the BRS model produced an estimated timecourse and a set of estimated response amplitudes. If necessary, the estimated timecourse and estimated response amplitudes were multiplied by −1 so that the estimated timecourse had a positive value at a time lag of 5 s. (This prevented ambiguity with respect to the sign of the response amplitudes.) We refer to the estimated timecourse as the *hemodynamic response function* (HRF), and the estimated response amplitudes as the *model estimation responses*.

Second, the BRS model was fit to the time-series data from the image identification runs. One basis function was used to characterize the shape of the response timecourse; this basis function was simply the HRF calculated in step 1. A separate parameter was used to characterize the amplitude of the response to each distinct image. Fitting the BRS model produced a set of estimated response amplitudes. We refer to the estimated response amplitudes as the *image identification responses*.

Third, the model estimation responses were standardized, and the same transformation (i.e. the same mean and standard deviation) was applied to the image identification responses. Standardization improved the consistency of responses across scan sessions (data not shown).

After this procedure was performed for each voxel in each scan session, model estimation responses and image identification responses were aggregated across scan sessions. For each model estimation response, the ratio between the absolute value of the response and its standard

error was calculated. For a given voxel the median ratio across model estimation responses was taken as the signal-to-noise ratio (SNR) of that voxel.

Single-trial analysis

The following procedure was performed for each voxel in each scan session. First, the BRS model was fit to the time-series data from the model estimation and image identification runs. One basis function was used to characterize the shape of the response timecourse; this basis function was simply the HRF calculated in the repeated-trial analysis. A separate parameter was used to characterize the amplitude of the response to each trial. Fitting the BRS model produced a set of estimated response amplitudes. We refer to the estimated response amplitudes for the model estimation and image identification runs as the *single-trial model estimation responses* and *single-trial image identification responses*, respectively. Next, the single-trial model estimation responses were standardized, and the same transformation (i.e. the same mean and standard deviation) was applied to the single-trial image identification responses. After this procedure was performed for each voxel in each scan session, single-trial model estimation responses and single-trial image identification responses were aggregated across scan sessions.

Analysis for additional scan sessions

In addition to the scan sessions for the main experiment, three additional scan sessions were conducted (see Methods in the main text). Each of these scan sessions consisted solely of image identification runs. To analyze the time-series data from these scan sessions, the following procedure was performed for each voxel in each scan session. First, the BRS model was fit to the time-series data from the image identification runs using the procedure described in step 1 of the repeated-trial analysis. This produced a set of image identification responses. Second, the image identification responses were standardized. Third, the BRS model was fit to the time-series data from the image identification runs using the procedure described in step 1 of the single-trial analysis. This produced a set of single-trial image identification responses. Fourth, the single-trial image identification responses were standardized. After this procedure was performed for each voxel in each scan session, image identification responses and single-trial identification responses were aggregated across scan sessions.

Construction of voxel activity patterns

The results of the repeated-trial and single-trial analyses were used to construct the voxel activity patterns used in the image identification stage of the experiment. Each voxel activity pattern represents the ensemble voxel response to an image. *Repeated-trial activity patterns* reflect data from multiple image presentations, and were constructed by concatenating individual voxels' estimated response amplitudes for an image. *Single-trial activity patterns* reflect data from single image presentations, and were constructed by concatenating individual voxels' estimated response amplitudes for a single trial.

Supplementary Methods 4. Basis-restricted separable model

The basis-restricted separable (BRS) model was used to pre-process the time-series data for each voxel (see Supplementary Methods 3). The BRS model assumes that each distinct image evokes a fixed response and that responses to different images sum over time. In addition, the model assumes that the response timecourses elicited by different images differ by only a scale factor[53]. To account for stimulus-related effects, the BRS model uses a set of basis functions to characterize the shape of the response timecourse[51] and a set of parameters to characterize the amplitudes of responses to different images. To account for noise-related effects, the model uses a set of polynomials[53] of degrees 0 through 3 and a first-order autoregressive noise model[55].

Let t be the number of time-series data points, e be the number of distinct images or trials, l be the number of points in the response timecourse, m be the number of timecourse basis functions, and p be the number of polynomial regressors. The time-series data for a given voxel are modeled as

$$\mathbf{y} = (\mathbf{X} * (\mathbf{Lc}))\mathbf{h} + \mathbf{Sb} + \mathbf{n}$$

where $\mathbf{y}$ is the data ($t \times 1$), $\mathbf{X}$ is the stimulus matrix ($t \times e$), $\mathbf{L}$ is the set of timecourse basis functions ($l \times m$), $\mathbf{c}$ is a set of parameters ($m \times 1$), $*$ denotes convolution, $\mathbf{h}$ is a set of response amplitudes ($e \times 1$), $\mathbf{S}$ is the set of polynomial regressors ($t \times p$), $\mathbf{b}$ is a set of parameters ($p \times 1$), and $\mathbf{n}$ is a noise term ($t \times 1$).

For the analysis of the time-series data as repeated trials, $\mathbf{X}$ consisted of one column per distinct image, where each column was a binary sequence with ones indicating the onsets of an image. For the analysis of the time-series data as single trials, $\mathbf{X}$ consisted of one column per trial, where each column was a binary sequence with a one indicating the onset of a trial. In cases where the shape of the timecourse was unknown, $\mathbf{L}$ was a set of Fourier basis functions consisting of a constant function and sine and cosine functions with 1, 2, and 3 cycles. These basis functions extended from 1 to 16 s after image onset. In cases where an estimate of the shape of the timecourse was available, $\mathbf{L}$ was simply taken to be that estimate.

Model parameters were estimated using a variant of the Cochrane-Orcutt procedure[55]. After initializing $\mathbf{h}$ to all ones, iterations alternated between ordinary least-squares estimation of $\mathbf{c}$ and $\mathbf{b}$ while holding $\mathbf{h}$ fixed and ordinary least-squares estimation of $\mathbf{h}$ and $\mathbf{b}$ while holding $\mathbf{c}$ fixed. After each iteration autoregressive noise parameters were estimated from the residuals of the model fit. These autoregressive noise parameter estimates were used to transform the data and design matrix prior to the next iteration. Fitting proceeded until convergence of parameter estimates.

Supplementary Methods 5. Model estimation

In the model estimation stage of the experiment, a receptive field was estimated for each voxel using the Gabor wavelet pyramid (GWP) model. The model estimation procedure is complicated because it involves multiple uses of the GWP model. For this reason we provide a high-level description of the procedure in this section, and present specific details of the GWP model in Supplementary Methods 6.

Rough localization of the receptive field

The first step of the model estimation procedure was to obtain a rough localization of the receptive field (RF). This was accomplished by fitting several initial models to the data. Each of the initial models covered a specific region of the stimulus (called the *field-of-view*), and was an instantiation of the GWP model at a resolution of 128 px × 128 px. One model covered the full 20° × 20° extent of the stimulus. In this case performance was limited by the fact that the maximum wavelet spatial frequency was 1.6 cycles/°. To better characterize voxels tuned to higher spatial frequencies, two additional models were used. One covered the central 10.1° × 10.1° of the stimulus, and the other covered the central 5.2° × 5.2° of the stimulus. In these cases the maximum wavelet spatial frequencies were 3.2 cycles/° and 6.2 cycles/°, respectively. (Voxels tuned to higher spatial frequencies tended to be found in more central regions of the visual field; data not shown.)

For each of the initial models, the RF was constrained to be orientation invariant. This was accomplished by summing over groups of input channels that differ in orientation but share the same spatial frequency and position, prior to fitting the model. The orientation invariance constraint reduced the number of free parameters and improved predictive power (data not shown). (Note that the final model was not constrained to be orientation invariant; see below.) There were a total of 1,367 free parameters for each initial model.

Precise localization of the receptive field

The second step of the model estimation procedure was to obtain a more precise estimate of the RF location. This was accomplished by fitting an isotropic two-dimensional Gaussian function to the spatial envelope associated with each initial model. The RF location was estimated as the region bounded by ± 2 s.d. of the fitted Gaussian. (The RF size was taken to be the size of this region.) For the 10.1° × 10.1° and 5.2° × 5.2° models, the estimated RF location was considered valid only if the 2-s.d. region was completely within the field-of-view of the model. This criterion excluded models artificially truncated by the field-of-view.

Of all the initial models that yielded a valid estimate of RF location, the model that achieved the least squared error on a separate stopping set was chosen (see Supplementary Methods 6). We refer to this model as the *best initial model*, and it was taken as providing the best estimate of the RF location. To reduce computational demands, subsequent analyses included only those voxels for which the predictive power of the best initial model was statistically significant (see Supplementary Methods 6).

Final estimate of the receptive field

The last step of the model estimation procedure was to obtain a final estimate of the RF. This was accomplished by fitting a GWP model that was specifically tailored to the estimated RF location. This model had a resolution of 64 px × 64 px, and was not constrained to be orientation invariant. There were a total of 2,730 free parameters in this final model.

Supplementary Methods 6. Gabor wavelet pyramid model

In Supplementary Methods 5 we described how the Gabor wavelet pyramid (GWP) model was used to estimate the receptive field of each voxel. In this section we provide specific details of the GWP model, such as how model parameters are determined.

Basic framework

The GWP model is applied to a specific region of the stimulus, called the *field-of-view* (FOV). The resolution is typically 64 px × 64 px, though in some cases, a resolution of 128 px × 128 px is used. The model describes how the portion of the stimulus within the FOV (henceforth simply referred to as the *image*) is transformed into a predicted response. Note that the GWP model does not include a temporal component because voxel-specific response timecourses are removed from the time-series data in pre-processing.

Stimulus pre-processing

To accommodate a variety of FOVs and resolutions, the stimuli used in the experiment were pre-processed at multiple resolutions. The dimensions of the pre-processed stimuli were given by $\min(500,\text{round}(2^{9-x/8}))$ where x ranges from 0 to 24. For example, for $x = 0$ the stimuli were left at the original resolution of 500 px × 500 px, and for $x = 10$ the stimuli were downsampled to a resolution of 215 px × 215 px. Stimuli were converted to luminance values using the measured luminance response of the goggles (see below). The mean luminance across all stimuli was then subtracted.

The luminance response of the goggles was measured with a Minolta LS-110 photometer (Konica Minolta Photo Imaging, Mahwah, NJ). The luminance response of the left-eye display was slightly different from that of the right-eye display; for analysis, the average of the two luminance responses was assumed. The minimum, maximum, and mean luminance was 0.8 cd/m^2, 11.1 cd/m^2, and 6.3 cd/m^2, respectively.

Design of the wavelet pyramid

The Gabor wavelet pyramid is illustrated in Supplementary Fig. 2. For the 64 px × 64 px model resolution, wavelets occur at five spatial frequencies: 1, 2, 4, 8, and 16 cycles per FOV. (For the 128 px × 128 px model resolution, wavelets occur at six spatial frequencies: 1, 2, 4, 8, 16, and 32 cycles per FOV.) At each spatial frequency f cycles per FOV, wavelets are positioned on an $f \times f$ grid. At each grid position wavelets occur at eight orientations, 0, 22.5°, 45°, ..., and 157.5°, and two quadrature phases, 0° and 90°. An isotropic Gaussian mask is used for each wavelet, and its size relative to spatial frequency is such that all wavelets have a spatial frequency bandwidth of 1 octave and an orientation bandwidth of 41°. A luminance-only wavelet that covers the entire image is also included.

Wavelets are truncated to lie within the bounds of the image, and are restricted in spatial extent by setting to zero the portions of the masks whose values are less than 0.01 of the peak value

(Supplementary Fig. 2b). Each wavelet is made zero-mean and unit-length within the bounds of its associated mask.

Transformation from image to predicted response

The following steps transform a given image into the predicted response from the GWP model (Supplementary Fig. 3). First, the image is projected onto the set of Gabor wavelets. The projections for each quadrature pair of wavelets are then squared, summed, and square-rooted, yielding a set of *input channels*. These input channels reflect the contrast energy contained in quadrature wavelet pairs. (For the luminance-only wavelet, the projection is squared, multiplied by 2, and square-rooted.) Next, the input channels are weighted by a set of values called the *kernel* and then summed. Finally, a DC offset is added to the result.

Wavelets positioned near the edge of the circular stimulus mask (see Supplementary Fig. 1) yield artifactually small projections. To avoid instability in parameter estimation, the projection for a given wavelet is set to zero if more than half of its associated mask lies beyond 90% of the stimulus radius.

The quantification of contrast energy is a nonlinear operation that transforms the stimulus into a space where the relationship between stimulus and response is more linear; for this reason, the GWP model is termed a *linearized model*[10]. A purely linear model that characterizes the voxel response as a weighted sum of the raw wavelet projections yields very poor fits (data not shown).

Estimation of model parameters

Responses to the images used in the model estimation runs of the experiment are used to fit the GWP model. Formally, let p be the number of images, and q be the number of input channels. The voxel responses were modeled as

$$\mathbf{y} = \mathbf{X}\mathbf{h} + c\mathbf{1} + \mathbf{n}$$

where $\mathbf{y}$ is the set of responses ($p \times 1$), $\mathbf{X}$ is the set of input channels ($p \times q$), $\mathbf{h}$ is the kernel ($q \times 1$), c is the DC offset (1×1), $\mathbf{1}$ is a vector of ones ($p \times 1$), and $\mathbf{n}$ is a noise term ($p \times 1$). Model parameters were estimated using gradient descent with early stopping[56]. Gradient descent is an iterative fitting technique where the difference between the model fit and the data is gradually reduced. Early stopping is a form of regularization[10] where the magnitude of model parameter estimates are shrunk in order to prevent overfitting.

The specific procedure was as follows. A randomly selected 20% of the responses were removed and kept as a stopping set. The mean of the remaining responses $\mathbf{y}_\mu$ (1×1) was subtracted, yielding responses $\tilde{\mathbf{y}}$ ($p \times 1$). The mean of each input channel $\mathbf{X}_\mu$ ($1 \times q$) was subtracted and the standard deviation of each input channel $\mathbf{X}_\sigma$ ($1 \times q$) was divided out, yielding input channels $\tilde{\mathbf{X}}$ ($p \times q$). The kernel was initialized to all zeros ($\mathbf{h}_1 = \mathbf{0}$) and then iteratively updated using gradient descent:

$$\mathbf{h}_{i+1} = \mathbf{h}_i - \varepsilon \mathbf{g}_i$$

where $\mathbf{h}_i$ is the kernel at iteration i, $\mathbf{g}_i$ is the normalized error gradient at iteration i, and $\varepsilon = 0.001$ is the step size. The normalized error gradient is given by

$$\mathbf{g}_i = \left[\left[\tilde{\mathbf{X}}^{\mathrm{T}}\left(\tilde{\mathbf{X}}\mathbf{h}_i - \tilde{\mathbf{y}}\right)\right] + \alpha\mathbf{g}_{i-1}\right]$$

where $[\mathbf{x}] = \mathbf{x} / \|\mathbf{x}\|$ represents vector length normalization, $\alpha = 0.9$ is a momentum parameter[57], and $\mathbf{g}_0 = \mathbf{0}$. Iterations proceeded until the squared error on the stopping set no longer decreased, or until the squared error on the responses no longer decreased. The final estimate of the kernel was calculated as

$$\hat{\mathbf{h}} = \mathbf{h}_{final} ./ \mathbf{X}_{\sigma}^{\mathrm{T}}$$

where $\mathbf{h}_{final}$ is the kernel at the last iteration and ./ denotes element-by-element division. The final estimate of the DC offset was calculated as

$$\hat{c} = \tilde{\mathbf{y}} - \mathbf{X}_{\mu}\hat{\mathbf{h}}$$

where the symbols are as defined earlier.

Estimation of variance

One hundred bootstrap samples were drawn from the original set of responses, and parameter estimates were obtained for each bootstrap sample. (The size of each bootstrap sample was equal to the number of responses, and the stopping set was selected after each bootstrap sample was drawn.) Standard errors on parameter estimates were calculated as the standard deviation across bootstraps. Final parameter estimates were calculated as the mean across bootstraps.

To prevent artificially high variance of parameter estimates, the number of fitting iterations was held constant across bootstrap samples. This was accomplished as follows. Prior to bootstrapping, parameter estimates were obtained using gradient descent with early stopping on the original set of responses. The number of fitting iterations n was recorded. Then, for each bootstrap sample, parameter estimates were obtained using gradient descent for n iterations.

Quantification of predictive power

An objective measure of the quality of a receptive-field model is how well the model predicts responses to images not used for model estimation[10]. Here, the predictive power of a receptive-field model was calculated as the correlation (Pearson's r) between measured and predicted responses for the images used in the image identification runs of the experiment. (There were 120 images used in the image identification runs, and these were distinct from the 1,750 images used in the model estimation runs; see Methods in the main text.) A bootstrap procedure was used to estimate statistical significance of predictive power ($r > 0$, one-tailed p-values).

Calculation of tuning curves

Tuning curves for space, orientation, and spatial frequency were calculated for each receptive-field (RF) model. To calculate the spatial tuning curve, i.e. spatial envelope, of an RF, the wavelet mask associated with each input channel was normalized to sum to 1, and was then scaled by the absolute value of the kernel weight associated with that input channel. The spatial

envelope was obtained by summing all wavelet masks. To calculate the orientation and spatial frequency tuning curves of an RF, a set of sinusoidal gratings were constructed at the same orientations and spatial frequencies used in the GWP model. At each combination of orientation and spatial frequency, gratings were constructed at multiple phases. The response of the RF to each grating was calculated, and tuning curves were obtained by averaging responses over one or more of the dimensions of orientation, spatial frequency, and phase.

Supplementary Methods 7. Image identification

Voxel selection

In our experiment approximately 5000 voxels were located in the stimulated portions of visual areas V1, V2, and V3 (see Supplementary Fig. 10 and Supplementary Table 1). There was substantial variation in the predictive power of the receptive-field models obtained for different voxels. Therefore, to optimize performance of the identification algorithm, we preferentially selected voxels whose receptive-field models had the highest predictive power. (Predictive power was quantified as how well a given model predicts responses to images not used to estimate the model; see Supplementary Methods 6). Note that the image to be identified was not included in the calculation of predictive power; this prevented voxel selection bias.

All identification results in this study were obtained using 500 voxels, as that number yields optimal performance (Supplementary Fig. 4). Most of these voxels were located in area V1 where predictive power was highest (Supplementary Table 1).

For measurement of identification performance under the Gabor wavelet pyramid and retinotopy-only models, voxels were selected based on the predictive power of the specific model under consideration. This ensured that each model had the best possible chance at performing well. For measurement of identification performance under the various constrained versions of the Gabor wavelet pyramid model, a single, fixed set of voxels was used. (The voxels in this set were selected based on the predictive power of the model that imposed no constraints on orientation and spatial frequency tuning.) Fixing the set of voxels used ensured that differences in identification performance directly reflect the different constraints imposed by the models.

Identification performance for different set sizes

To measure identification performance for set sizes up to 1,000 images, the following procedure was used. First, a library of 999 images was constructed. These images were randomly selected and were different from the images used in the model estimation and image identification stages of the experiment. Then, for set size s and measured voxel activity pattern m, identification performance was calculated as the probability that the predicted voxel activity pattern for the correct image is more correlated with m than the predicted voxel activity patterns for $s - 1$ images drawn randomly from the library:

$$f(m,s) = \prod_{i=1}^{s-1} \frac{1{,}000 - g(m) - i}{1{,}000 - i}$$

where $f(m,s)$ is identification performance and $g(m)$ is the number of library images whose predicted voxel activity patterns were more correlated with m than with the correct image. Finally, identification performance was averaged over all measured voxel activity patterns m.

To measure identification performance for larger set sizes, an extrapolation method was used. First, the correlation between the measured voxel activity pattern m and the predicted voxel activity pattern for each library image was calculated. This produced a distribution of 999 correlation values. Next, the distribution was smoothed using a Gaussian kernel. Kernel width

was chosen by pseudo-likelihood cross-validation[58] using code provided by A. Ihler (http://ttic.uchicago.edu/~ihler/code/). (Smoothing in this way produces a better estimate of the true underlying distribution, and is reasonable given that the library images were randomly selected.) Identification performance was then calculated as

$$f(m,s) = \left(1 - h(m)\right)^{s-1}$$

where $f(m,s)$ is identification performance and $h(m)$ is the fraction of the smoothed distribution larger than the correlation between m and the predicted voxel activity pattern for the correct image. This equation quantifies the probability that the predicted voxel activity pattern for the correct image is more correlated with m than with the predicted voxel activity patterns for $s - 1$ images drawn randomly from all possible images. Finally, identification performance was averaged over all measured voxel activity patterns m. To validate the described extrapolation method, we calculated empirical performance levels for a set size of six million images, and confirmed that these values are accurately estimated by extrapolation.

Estimation of the noise ceiling

The noise ceiling on identification performance was estimated in order to determine whether differences in identification performance across subjects could be attributed to differences in signal-to-noise ratio (see Fig. 4a). The noise ceiling is the theoretical maximum performance that could ever be achieved, given the level of noise in the data. To estimate the noise ceiling, 25 bootstrap-like simulations were conducted for each of the 120 images used in the image identification stage of the experiment. In each simulation, the first step was to generate a measured voxel activity pattern for the correct image. This was accomplished by taking the mean of a random sample drawn from the single-trial activity patterns evoked by the correct image. The next step was to generate a predicted voxel activity pattern for each potential image the subject could have seen. This was accomplished by taking the mean of a random sample drawn from the single-trial activity patterns evoked by each potential image. (The intuition here is that the quality of the predicted voxel activity patterns is limited only by intrinsic measurement variability, not by the predictive power of receptive-field models.) Finally, the image whose predicted voxel activity pattern was most correlated with the measured voxel activity pattern was selected. The noise ceiling was calculated as the percentage of simulations where identification was successful.

Supplementary Methods 8. Retinotopy-only model

The retinotopy-only (RO) model characterizes the response of each voxel as a function of the luminance and contrast of a specific region of the stimulus (this region is henceforth simply referred to as the *image*). There are two input channels. The luminance channel represents absolute deviation from mean luminance. (It has been shown that changes in uniform illumination evoke fMRI activity in early visual areas[27].) The contrast channel represents the total energy contained in the image excluding overall luminance. Note that the RO model is invariant to the particular orientations and spatial frequencies present in the image.

The RO model provides a plausible functional description of a voxel in early visual areas, and it is similar to recently proposed models of phase-encoded retinotopic mapping data[3,15,19,35]. Since the RO model captures only spatial tuning, it serves as a way of testing whether the additional orientation and spatial frequency tuning captured by the Gabor wavelet pyramid (GWP) model have a significant impact on identification performance. If orientation and spatial frequency tuning are irrelevant for identification or if they cannot be estimated reliably from voxel responses, then performance for the RO model should be at least as good as performance for the GWP model.

To ensure that the RO and GWP models are compared fairly, the RO model was applied to the same estimated receptive-field location as used for the GWP model (see Supplementary Methods 5), and both models were fit using the same gradient descent method (see Supplementary Methods 6).

Spatially-weighted metrics for luminance and contrast

To implement the RO model we must choose metrics for luminance and contrast. The standard metrics for luminance and contrast are the mean and standard deviation of the pixel luminance values, respectively. These metrics are spatially homogenous in the sense that all portions of the image contribute equally. However, it is reasonable to presume that the receptive field of a voxel exhibits spatial gradation such that portions of the image near the center of the receptive field contribute more strongly to the response than portions of the image near the periphery of the receptive field. Indeed, previous studies[3,15,35] have proposed a two-dimensional Gaussian model of the spatial envelope of a voxel receptive field. Moreover, the receptive fields obtained in the present study under the GWP model do appear to be spatially graded (see Fig. 2 and Supplementary Fig. 9).

To accommodate spatial gradation the RO model uses the following metrics. The *spatially-weighted luminance* of an image is given by

$$L = \frac{\sum_i w_i x_i}{\sum_i w_i}$$

where L is the spatially-weighted luminance, w_i is the weight on pixel i, and x_i is the luminance of pixel i. The *spatially-weighted contrast* of an image is given by

$$C = \sqrt{\frac{\sum_i w_i \left(x_i - L\right)^2}{\sum_i w_i}}$$

where C is the spatially-weighted contrast and the other symbols are as defined earlier. These metrics are calculated at the original stimulus resolution (downsampling is not necessary). Note that in the case where all weights are equal to one, the spatially-weighted metrics for luminance and contrast reduce to the standard metrics for luminance and contrast.

Transformation from image to predicted response

Let G represent the two-dimensional Gaussian fit to the spatial envelope associated with the best initial model (as described in Supplementary Methods 5). The following steps transform a given image into the predicted response from the RO model. First, the spatially-weighted luminance of the image is calculated using the weights provided by G. The absolute value of the result constitutes the first input channel. (This full-wave rectification parallels how luminance is treated in the GWP model.) Next, the spatially-weighted contrast of the image is calculated using the weights provided by G. The result constitutes the second input channel. The two input channels are then weighted by a set of values and summed. Finally, a DC offset is added to the result.

Validation of the spatially-weighted metrics

To verify that the spatially-weighted metrics yield reasonable results, we compared identification performance achieved using the spatially-weighted metrics with that achieved using the standard metrics. (The standard metrics were calculated using the region of the stimulus bounded by ± 2 s.d. of the Gaussian function G.) Identification performance was 55% and 42% for the spatially-weighted metrics and standard metrics, respectively (repeated trial, 120 images, performance averaged across subjects). This validates the spatially-weighted metrics and indicates that we have cast the RO model in the best possible light.

Supplementary Methods 9. Constrained versions of the Gabor wavelet pyramid model

Several constrained versions of the Gabor wavelet pyramid (GWP) model were constructed in order to assess the individual contributions of orientation and spatial frequency tuning to identification performance. These models are based on the GWP model instantiated at a resolution of 64 px × 64 px, and impose various constraints on orientation and spatial frequency tuning. The models were applied to the same estimated receptive-field location as used for the GWP model (see Supplementary Methods 5) and were fit using the same gradient descent method (see Supplementary Methods 6).

Model simplification

To facilitate manipulation of orientation and spatial frequency tuning, the spatial envelope of the GWP model was first fixed. This was accomplished by weighting the image with the two-dimensional Gaussian associated with the estimated receptive-field location (see Supplementary Methods 5), and summing over input channels that differ in position but share the same orientation and spatial frequency. (Note that different voxels had different spatial envelopes.) To facilitate imposition of tuning constraints, input channels were linearly transformed such that weights on input channels directly reflect how the model responds to sinusoidal gratings (details of the transformation are provided in a later section).

Constraints on orientation and spatial frequency tuning

Systematic constraints were imposed on orientation and spatial frequency tuning. Three different constraints were used for each dimension, yielding a total of 3 × 3 = 9 different models. Under the constraint of *flat tuning*, the tuning curve of each voxel is constrained to be entirely flat. Under *ROI-averaged tuning*, the tuning curve of each voxel is constrained to match the mean tuning curve across voxels in the corresponding region-of-interest (i.e. V1, V2, or V3), and any voxel-to-voxel variation in tuning is ignored. Under *individual-voxel tuning*, each voxel is allowed full flexibility in tuning, so voxel-to-voxel variation in tuning is captured. (For an illustrative example, see Supplementary Fig. 6.)

Tuning constraints were achieved by applying marginalization operations to input channels. To achieve flat tuning, input channels were summed across the relevant dimension; to achieve ROI-averaged tuning, input channels were multiplied by the appropriate ROI-averaged tuning curve (see Supplementary Fig. 7) and then summed across the relevant dimension; and to achieve individual-voxel tuning, input channels were left as-is. To illustrate, consider the model that imposes flat orientation tuning and ROI-averaged spatial frequency tuning. This model was constructed by summing over input channels that differ in orientation but share the same spatial frequency, multiplying the resulting channels by the ROI-averaged spatial frequency tuning curve, and summing the results.

Comparison with the retinotopy-only model

Like the retinotopy-only (RO) model, the constrained versions of the GWP model help assess the contribution of orientation and spatial frequency tuning to identification performance. The RO

model serves as a simple, plausible alternative to the GWP model, and assesses the overall importance of orientation and spatial frequency. In contrast, the constrained versions of the GWP model examine the individual contributions made by orientation and spatial frequency, and constitute a more direct investigation of the orientation and spatial frequency information conveyed by the GWP model.

The model that imposes flat orientation and spatial frequency tuning is similar to the RO model in that both models capture spatial tuning but discard orientation and spatial frequency information. However, the models are not equivalent: they are constructed from different image bases (Gabor wavelet basis vs. pixel basis) and incorporate different kinds of nonlinearities. The models also differ in how they handle overall luminance, so their predicted responses can diverge substantially at very low spatial frequencies.

Technical detail on the transformation of input channels

Before transformation, weights on input channels do not reflect how the model responds to sinusoidal gratings. This is due to the fact that the Gabor wavelets have overlapping spectra and the fact that there are different numbers of wavelets at different spatial frequency levels of the Gabor wavelet pyramid.

The crux of the transformation lies in simulating the response of the model to gratings. Gratings are constructed at 16 equally spaced phases at each combination of orientation and spatial frequency used in the GWP model. This yields a total of 8 orientations × 5 spatial frequencies × 16 phases = 640 gratings. The gratings are then used to construct a set of input channels $\mathbf{G}$ ($n \times q$) where $n = 640$ is the number of gratings and $q = 41$ is the number of input channels.

Suppose that responses evoked by the gratings were actually measured. These responses could be modeled as

$$\mathbf{r} = \mathbf{G}\mathbf{k} + \mathbf{n}$$

where $\mathbf{r}$ ($n \times 1$) is the set of grating responses, $\mathbf{k}$ is the kernel ($q \times 1$), and $\mathbf{n}$ is a noise term ($n \times 1$). Then, ordinary least-squares estimation could be used to determine the kernel that minimizes the squared error between the model prediction and the measured responses:

$$\hat{\mathbf{k}} = \left(\mathbf{G}^{\mathrm{T}}\mathbf{G}\right)^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{r}$$

where the symbols are as defined earlier. Intuitively, $\hat{\mathbf{k}}$ can be viewed as the kernel that best achieves the measured grating responses $\mathbf{r}$ under the least-squares (LS) criterion. (In practice, $\hat{\mathbf{k}}$ achieves very good approximations of $\mathbf{r}$—see Supplementary Fig. 6.)

Now, observe that the images shown in the actual experiment can be used to construct a set of input channels $\mathbf{X}$ ($p \times q$) where p is the number of images. Under the assumption that the kernel is equal to $\hat{\mathbf{k}}$, the predicted responses to the images are given by $\mathbf{X}\hat{\mathbf{k}} + c\mathbf{1}$ where c is a DC offset (1×1) and $\mathbf{1}$ is a vector of ones ($p \times 1$). This expression can be rewritten as $\tilde{\mathbf{X}}\mathbf{r} + c\mathbf{1}$ where $\tilde{\mathbf{X}} = \mathbf{X}\left(\mathbf{G}^{\mathrm{T}}\mathbf{G}\right)^{-1}\mathbf{G}^{\mathrm{T}}$ is a set of transformed input channels ($p \times n$). Thus, implicit here is the following model of the responses to the images:

$$\mathbf{y} = \tilde{\mathbf{X}}\mathbf{r} + c\mathbf{1} + \mathbf{n}$$

where $\mathbf{y}$ is the set of image responses ($p \times 1$) and $\mathbf{n}$ is a noise term ($p \times 1$).

The above considerations demonstrate that under the LS criterion, responses to the images shown in the actual experiment can be modeled using the transformed input channels $\tilde{\mathbf{X}}$ and the set of weights $\mathbf{r}$. Thus, the transformation of $\mathbf{X}$ into $\tilde{\mathbf{X}}$ achieves the desired condition that weights on input channels directly reflect the response of the model to gratings. The final step is to sum over the input channels of $\tilde{\mathbf{X}}$ that represent different grating phases but the same grating orientation and spatial frequency. This is reasonable since phase information is discarded when quadrature pairs of wavelets are combined.

Supplementary Methods 10. Visual area localization

Construction of cortical surface representation

High-resolution anatomical data were acquired on a 1.5 T Philips Eclipse MR scanner (Philips Medical Systems, N.A., Bothell, WA). A T1-weighted MPRAGE pulse sequence was used: TR 15 ms, TE 4.47 ms, flip angle 35°, field-of-view 240 mm × 240 mm × 275.6 mm, matrix size 256 × 256 × 212, resolution 0.9375 mm × 0.9375 mm × 1.3 mm. Two anatomical volumes were acquired for each subject. The volumes were resampled to isotropic 1 mm × 1 mm × 1 mm voxels, manually co-registered using a rigid-body transformation, and averaged together to increase the contrast-to-noise ratio. The SureFit BETA v4.45 software package[59] was used to construct a triangulated mesh at the boundary between white and gray matter. The Caret v5.1 software package[59] was used to flatten this surface representation using a cut along the calcarine sulcus. (See http://brainmap.wustl.edu/caret/ for more information on SureFit and Caret.)

Registration of functional volumes

In the main experiment, an in-plane anatomical volume was acquired in the spatial reference frame to which all functional volumes for a given subject were registered. This in-plane anatomical volume was manually registered to the high-resolution anatomical volume (described above) using a rigid-body transformation. The parameters for this transformation were then used as an initial guess for the registration of the functional volumes to the high-resolution anatomical volume. This registration was subsequently improved by manually adjusting scaling and translation along the in-plane image dimensions. This resulted in an affine transformation that described the registration of the functional volumes to the cortical surface representation.

Localization of visual areas

In separate scan sessions fMRI data were collected using the multifocal retinotopic mapping technique[17,31] (see Supplementary Methods 11). These data were used to generate flattened maps of receptive-field angle and eccentricity. Visual areas V1, V2, and V3 were selected on these surface maps, and were represented as mutually disjoint sets of vertices. For assignment of voxels to visual areas, only voxels within 4 mm of surface vertices were considered. Each voxel was assigned to the visual area associated with the vertex closest to the voxel. Voxels outside of areas V1, V2, and V3 were discarded and not used in this study.

Supplementary Methods 11. Multifocal retinotopic mapping

The multifocal retinotopic mapping technique[17,31] was used to localize visual areas and to validate retinotopic information derived from the Gabor wavelet pyramid model. Estimates of retinotopic tuning provided by the multifocal technique have been shown to be similar to those provided by the more conventional phase-encoded technique[31,60,61].

Stimulus

The stimulus size was 20° × 20° (500 px × 500 px). A central white square served as the fixation point, and its size was 0.2° × 0.2° (4 px × 4 px). The stimulus was composed of 33 spatial components: a central circle and surrounding sectors defined by the intersections of 8 wedges and 4 rings. The boundaries of the wedges were positioned at angles of 0°, 45°, 90°, ..., and 315°, and the boundaries of the rings were positioned at eccentricities of 0.5°, 1.3°, 2.8°, 5.4°, and 10°. Each spatial component had one of two states. In the ON state, the spatial component was filled with a grayscale texture composed of non-Cartesian gratings[32]. The texture switched to different random configurations at a rate of 4 Hz. In the OFF state, the spatial component was filled with the gray background. The luminance of the gray background was set to the mean luminance of the texture.

The ON/OFF patterns for the spatial components were determined by an m-sequence[52] of level 5, order 4, and length $5^4 - 1 = 624$. Code for m-sequence generation was provided by T. Liu (http://fmriserver.ucsd.edu/ttliu/mttfmri_toolbox.html). One level of the m-sequence was associated with the ON state, and the other levels were associated with the OFF state. The m-sequence was repeatedly cyclically shifted by four elements to produce the ON/OFF pattern for each spatial component. Each element was assigned a duration of 4 s, and the total stimulus duration was 624 elements × 4 s = 41.6 min. For the purposes of data collection, the stimulus was divided into three consecutive segments (13.9 min each).

Data collection

Retinotopic mapping data were collected in one scan session from each subject. The same stimulus presentation setup and MRI parameters were used as in the main experiment. Each scan session consisted of three runs (13.9 min each), corresponding to the three segments of the stimulus.

Data analysis

Functional brain volumes were reconstructed and co-registered as in the main experiment. The time-series data for each voxel were then analyzed using the basis-restricted separable model (see Supplementary Methods 4). A set of basis functions was used to characterize the shape of the response timecourse, and a free parameter was used to characterize the amplitude of the response to each spatial component. Note that this model assumes linear spatial summation[17] in the sense that the response of a voxel to a combination of spatial components is assumed to equal the sum of the responses of the voxel to each individual spatial component.

For each voxel the estimated response amplitudes to each spatial component were used to calculate estimates of the angle and eccentricity of the voxel's receptive field. For angle, a vector summation procedure[32] was used:

$$A = \arg\left(\sum_i |a_i|^+ e^{j\theta_i} \right)$$

where A is the estimated receptive-field angle, i ranges over each spatial component except the central circle, a_i is the estimated response amplitude to spatial component i, $|\ |^+$ represents positive half-wave rectification, and θ_i is the mean angle of spatial component i. For eccentricity, a center-of-mass weighting procedure[32] was used:

$$E = \frac{\sum_i |a_i|^+ k_i}{\sum_i |a_i|^+}$$

where E is the estimated receptive-field eccentricity, i ranges over each spatial component, k_i is the mean eccentricity of spatial component i, and other symbols are as defined earlier.

Supplementary Note 1. Additional references

31. Vanni, S., Henriksson, L. & James, A. C. Multifocal fMRI mapping of visual cortical areas. *Neuroimage* **27**, 95–105 (2005).
32. Hansen, K. A., Kay, K. N. & Gallant, J. L. Topographic organization in and near human visual area V4. *J. Neurosci*. **27**, 11896–11911 (2007).
33. Wandell, B. A., Dumoulin, S. O. & Brewer, A. A. Visual field maps in human cortex. *Neuron* **56**, 366–383 (2007).
34. Kraft, A. *et al*. fMRI localizer technique: efficient acquisition and functional properties of single retinotopic positions in the human visual cortex. *Neuroimage* **28**, 453–463 (2005).
35. Larsson, J. & Heeger, D. J. Two retinotopic visual areas in human lateral occipital cortex. *J. Neurosci*. **26**, 13128–13142 (2006).
36. Tootell, R. B. *et al*. Functional analysis of V3A and related areas in human visual cortex. *J. Neurosci*. **17**, 7060–7078 (1997).
37. O'Craven, K. M. & Kanwisher, N. Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *J. Cogn. Neurosci*. **12**, 1013–1023 (2000).
38. O'Toole, A. J., Jiang, F., Abdi, H. & Haxby, J. V. Partially distributed representations of objects and faces in ventral temporal cortex. *J. Cogn. Neurosci*. **17**, 580–590 (2005).
39. Boynton, G. M., Engel, S. A., Glover, G. H. & Heeger, D. J. Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci*. **16**, 4207–4221 (1996).
40. Heeger, D. J., Huk, A. C., Geisler, W. S. & Albrecht, D. G. Spikes versus BOLD: what does neuroimaging tell us about neuronal activity? *Nature Neurosci*. **3**, 631–633 (2000).
41. Rees, G., Friston, K. & Koch, C. A direct quantitative relationship between the functional properties of human and macaque V5. *Nature Neurosci*. **3**, 716–723 (2000).
42. Boynton, G. M. & Finney, E. M. Orientation-specific adaptation in human visual cortex. *J. Neurosci*. **23**, 8781–8787 (2003).
43. Engel, S. A. Adaptation of oriented and unoriented color-selective neurons in human visual areas. *Neuron* **45**, 613–623 (2005).
44. Fang, F., Murray, S. O., Kersten, D. & He, S. Orientation-tuned fMRI adaptation in human visual cortex. *J. Neurophysiol*. **94**, 4188–4195 (2005).
45. Larsson, J., Landy, M. S. & Heeger, D. J. Orientation-selective adaptation to first- and second-order patterns in human visual cortex. *J. Neurophysiol*. **95**, 862–881 (2006).
46. Murray, S. O., Olman, C. A. & Kersten, D. Spatially specific fMRI repetition effects in human visual cortex. *J. Neurophysiol*. **95**, 2439–2445 (2006).
47. Tootell, R. B. *et al*. Functional analysis of primary visual cortex (V1) in humans. *Proc. Natl Acad. Sci. USA* **95**, 811–817 (1998).
48. Furmanski, C. S. & Engel, S. A. An oblique effect in human primary visual cortex. *Nature Neurosci*. **3**, 535–536 (2000).
49. Goodyear, B. G., Nicolle, D. A., Humphrey, G. K. & Menon, R. S. BOLD fMRI response of early visual areas to perceived contrast in human amblyopia. *J. Neurophysiol*. **84**, 1907–1913 (2000).
50. Sereno, M. I. & Huang, R. S. A human parietal face area contains aligned head-centered visual and tactile maps. *Nature Neurosci*. **9**, 1337–1343 (2006).
51. Dale, A. M. Optimal experimental design for event-related fMRI. *Hum. Brain Mapp*. **8**, 109–114 (1999).
52. Buracas, G. T. & Boynton, G. M. Efficient design of event-related fMRI experiments

using m-sequences. *Neuroimage* **16**, 801–813 (2002).
53. Kay, K. N., David, S. V., Prenger, R. J., Hansen, K. A. & Gallant, J. L. Modeling low-frequency fluctuation and hemodynamic response timecourse in event-related fMRI. *Hum. Brain Mapp.* **29**, 142–156 (2008).
54. Kellman, P., van Gelderen, P., de Zwart, J. A. & Duyn, J. H. Method for functional MRI mapping of nonlinear response. *Neuroimage* **19**, 190–199 (2003).
55. Bullmore, E. *et al*. Colored noise and computational inference in neurophysiological (fMRI) time series analysis: resampling methods in time and wavelet domains. *Hum. Brain Mapp.* **12**, 61–78 (2001).
56. Skouras, K., Goutis, C. & Bramson, M. J. Estimation in linear models using gradient descent with early stopping. *Stat. Comput.* **4**, 271–278 (1994).
57. Qian, N. On the momentum term in gradient descent learning algorithms. *Neural Networks* **12**, 145–151 (1999).
58. Cao, R., Cuevas, A. & Manteiga, W. G. A comparative study of several smoothing methods in density estimation. *Comput. Stat. Data An.* **17**, 153–176 (1994).
59. Van Essen, D. C. *et al*. An integrated software suite for surface-based analyses of cerebral cortex. *J. Am. Med. Inform. Assn.* **8**, 443–459 (2001).
60. Fukunaga, M., van Gelderen, P., de Zwart, J. A., Jansma, J. M. & Duyn, J. H. Retinotopic fMRI mapping with pseudo-random stimulus presentation using the m-sequence paradigm. *Soc. Neurosci. Abstr.* 693.2 (2004).
61. Kay, K. N., Hansen, K. A., David, S. V. & Gallant, J. L. Artifacts in phase-encoded fMRI retinotopic mapping. *Soc. Neurosci. Abstr.* 508.12 (2005).