Behavioral/Systems/Cognitive

Receptive Field Properties of Neurons in the Early Visual Cortex Revealed by Local Spectral Reverse Correlation

Shinji Nishimoto, Tsugitaka Ishida, and Izumi Ohzawa
Graduate School of Frontier Biosciences, Osaka University, Osaka 560-8531, Japan

We introduce a novel class of white-noise analyses, named local spectral reverse correlation (LSRC), which is capable of revealing various aspects of visual receptive field profiles that were undetectable previously in a single simple measurement. The method is based on spectral analyses in a two-dimensional spatial frequency domain for spatially localized areas within and around their receptive fields. Extracellular single-unit recordings were performed for area 17 and 18 neurons in anesthetized cats. A dynamic dense noise pattern was presented in which the pattern covered an area two to three times larger than the classical receptive field. Spike trains were then cross-correlated with frequency spectra of localized noise pattern to obtain spatially localized selectivity maps in the two-dimensional frequency domain. Our findings are as follows. (1) The new LSRC method allows measurements of two-dimensional frequency tunings and their spatial extent even for cells with substantial nonlinearity. (2) A small subset of neurons shows spatial inhomogeneity in the two-dimensional frequency tunings. (3) In addition to facilitatory response profiles, we can also visualize suppressive profiles localized both in space and spatial frequency domains. Our results suggest that the new analysis technique can be a powerful tool for measuring visual response profiles that contain inhomogeneity in space, as well as for studying neurons with substantial nonlinearities. These features make the method particularly suitable for studying response profiles of neurons in early as well as intermediate extrastriate visual areas.

Key words: **cat; areas 17 and 18; early visual cortex; shape selectivity; local spectral reverse correlation; spatial frequency**

Introduction

Various mapping methods, in particular those that use a reverse correlation analysis, have been very effective in providing detailed receptive fields of neurons in early stages of the visual pathway (Jones and Palmer, 1987; DeAngelis et al., 1993; Reid et al., 1997). However, some aspects of receptive field properties cannot be measured easily by the currently available methods. These include the so-called cross-orientation suppression (Morrone et al., 1982; Bonds, 1989), surround suppression (Hubel and Wiesel, 1968; Dreher, 1972), and possible local variations of tuning properties within a receptive field. Once we go beyond these early areas, mapping receptive fields of neurons in the extrastriate visual areas are expected to be even more difficult. How would one go about measuring receptive fields of downstream cells that collect input from early visual cortical neurons? To acquire selectivity to complex visual features, the extrasriate neurons might collect spatially inhomogeneous inputs from these early-stage filters, meaning that the filter properties (e.g., orientation and spatial frequency tunings) are not uniform over their receptive fields but are different greatly to define selectivities to curved contours (Gallant et al., 1993, 1996; Pasupathy and Connor, 2001). To extend the reverse correlation method for studying details of the early visual cortical areas and for possible uses in extrastriate areas, we have developed a new class of white-noise analysis, named local spectral reverse correlation (LSRC). The purposes of this study are to validate the new LSRC analysis and to examine details of previously invisible response properties including the extent and nature of spatial inhomogeneity, if any, of cells in the early visual cortex.

The LSRC method uses a wide-area, two-dimensional dynamic white-noise sequence similar to those used in previous studies (Reid et al., 1997). However, the key novel idea is in calculating the cross-correlations between the spike train and amplitude spectra of spatially windowed (hence localized in analysis) noise sequence. By doing this, we can acquire the response profiles of the cell in a two-dimensional frequency domain for subfields within and around the receptive field. We have applied this procedure for cells in the early visual cortex of cats and found the following. (1) The LSRC method allows measurements of two-dimensional frequency tunings and their spatial extent for both simple and complex cells, whereas conventional space-domain reverse correlation with dense noise does not reveal first-order responses for complex cells. (2) A small subset of neurons

Received Oct. 25, 2005; revised Feb. 5, 2006; accepted Feb. 6, 2006.

This work was supported by Grants 15029230 and 15700258 and the Project on Neuroinformatics Research in Vision through special coordination funds for promoting science and technology from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) and by Grant 13308048, the 21st Century COE Program, and a France–Japan Joint Research Program Grant from the Japan Society for the Promotion of Science. We thank our laboratory members, Hiroki Tanaka, Takahisa Sanada, Rui Kimura, Kota Sasaki, Masayuki Fukui, Miki Arai, Masashi Iida, and Taihei Ninomiya, for their help in experiments and valuable discussions. We also thank Dr. Jack Gallant and his colleagues for their support.

Correspondence should be addressed to Dr. Izumi Ohzawa, Graduate School of Frontier Biosciences and School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531 Japan. E-mail: ohzawa@fbs.osaka-u.ac.jp.

S. Nishimoto's present address: Helen Wills Neuroscience Institute, University of California at Berkeley, Berkeley, CA 94720-1650.

DOI:10.1523/JNEUROSCI.4558-05.2006

 0270-6474/06/263269-12$15.00/0

exhibits spatial inhomogeneity in the two-dimensional frequency tunings. (3) In addition to facilitatory response profiles, we can also visualize suppressive profiles localized both in space and spatial frequency domains, which cannot be revealed by a standard reverse correlation procedure. Computational investigations also reveal that the new method is highly effective even in cases in which high thresholds prevent a cell from responding to individual local optimal stimuli alone. Our results suggest that the new analysis technique can be a powerful tool for measuring visual receptive filed profiles that contain inhomogeneity in space, as well as for studying neurons with substantial nonlinearities.

Materials and Methods

All recordings were made from adult cats weighing between 1.5 and 3.7 kg. All animal care and experimental guidelines conformed to those established by the National Institutes of Health and were approved by the Osaka University Animal Care and Use Committee.

Surgical procedure. Detailed procedures have been described in our recent publication (Nishimoto et al., 2005). Briefly, each cat was anesthetized with isoflurane (2.5–3.5% in O_2) after initial preanesthetic doses of hydroxyzine (atarax; 2.5 mg) and atropine (0.05 mg). Electrocardiogram electrodes and a rectal temperature probe were inserted, and a femoral vein was catheterized. Then, cefotiam hydrochloride (Panspolin; 8.3 mg) and dexamethasone sodium phosphate (Decadron; 0.4 mg) were administered. Subsequently, a tracheostomy was performed, and a tracheal tube was inserted. Then, the animal's head was secured in a stereotaxic device with the use of ear and mouth bars and clamps on the orbital rim. Tips of the ear bars were coated with local anesthetic gel (Lidocaine). Anesthesia was then switched to sodium thiopental (Ravonal; given continuously at 1.0–1.5 mg/kg/h). After stabilization of anesthesia, paralysis was induced with a loading dose of gallamine triethiodide (10–20 mg), and the animal was placed under artificial respiration at the rate of 20–30 strokes per minute. The respiration rate and stroke volume were adjusted to maintain end-tidal CO_2 between 3.5 and 4.3%. Artificial respiration was performed with a gas mixture of 70% N_2O and 30% O_2. The infusion fluid thereafter contained Ravonal, gallamine triethiodide (10 mg/kg/h), and glucose (40 mg/kg/h) in Ringer's solution. A craniotomy was then performed directly above the central representation of the visual field in the visual area 17 or 18 (Horsley-Clarke P4 L2.5 for recordings of A17 and A3 L3 for recordings of A18). The dura was dissected away to allow insertion of microelectrodes. We used tungsten microelectrodes (5 MΩ; A-M Systems, Everett, WA) for recording spike activity extracellularly. Typically, two electrodes were used to increase the chance of encountering cells, and they were mounted in parallel in a single protective guide tube and driven by a common microelectrode drive (Narishige, Tokyo, Japan). After lowering the electrodes to the cortical surface, agar was used to protect the cortex, and melted wax was applied over the agar to create a sealed chamber for stabilization. Body temperature was maintained near 38.3°C with the use of a servo-controlled heating pad. Pupils were dilated with atropine (1%), and nictitating membranes were retracted with phenylephrine hydrochloride (Neosynesin; 5%). Contact lenses of appropriate power with 4 mm artificial pupils were positioned on each cornea.

To record the activity of single units, electrical signals from the microelectrodes were amplified (10,000×) and bandpass filtered (300–5000 Hz). Then spike sorting was achieved using a custom-built spike sorter (Ohzawa et al., 1996), in which each spike was sorted by their waveforms and time-stamped with 40 μs resolution.

Visual stimulation. All of the experiment control functions and generations of visual stimuli were performed using custom-written software on two Windows personal computers. Visual stimuli were generated by a dedicated personal computer and displayed on a cathode ray tube display (GDM-FW900; a resolution of 1600 × 1024 pixels, refreshed at 76 Hz; Sony, Tokyo, Japan). The animal saw the display through a custom-built haploscope, which allowed dichoptic presentations of visual stimuli to the left and right eyes separately using 800 × 1024 pixel areas of the display. The distances (total length of light paths) between the screen and the eyes were set to 57 cm, subtending the visual field of 23° (horizontal) × 30° (vertical) for each eye. All measurements were performed for the dominant eye.

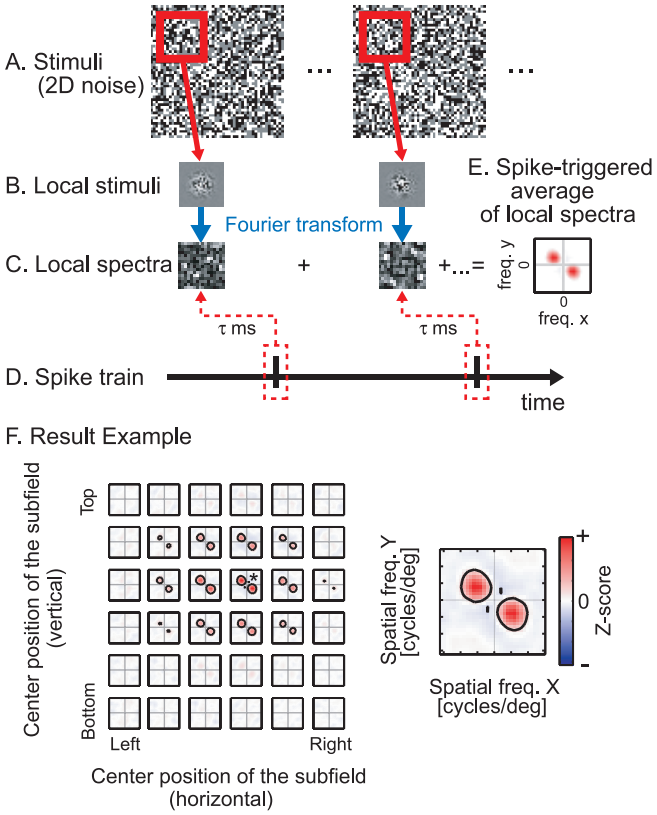


Figure 1. ***A***, Schematic diagram of the LSRC procedure (see Materials and Methods for details). ***C–E***, By calculating a cross-correlation between the spike train (***D***) and spectra of Gaussian-windowed stimuli (***C***), we obtain a two-dimensional frequency tuning for the given subfield (***E***). ***F***, By changing iteratively the center position of the Gaussian window, we can obtain a spatial matrix of the two-dimensional frequency tunings, corresponding to the matrix of localized areas of analysis shown in ***B***. The strongest local spectral tuning map is indicated by an asterisk and is shown enlarged on the right. In each of these spectral maps, facilitatory and suppressive responses are shown by red and blue, respectively, according to the scale bar (suppression is essentially absent for this neuron). These conventions are used for subsequent figures. By nature of the Fourier transform, the tunings are symmetric about the origin. 2D, Two-dimensional; freq., frequency; deg, degree.

For each cell we encountered, we have presented a dynamic two-dimensional noise array (Fig. 1*A*). The area covered by the noise array is typically two to three times larger than the classical receptive fields in width and height (typical ranges are from 12 × 12° to 20 × 20°). The noise array consists of 51 × 51 elements, in which the luminance of each element is bright (~90 cd/m^2), dark (~3 cd/m^2), or equal to the mean luminance of the display (~47 cd/m^2). The noise array is redrawn with a new noise pattern every 26 ms (two video frames). Typically, 10 blocks of the noise arrays (a total of 68,400 frames, or 30 min) are presented to obtain sufficient number of spikes for data analysis.

Data analysis. To obtain two-dimensional frequency tunings for spatially localized areas, we have performed a LSRC. LSRC is an application of the standard spike-triggered average techniques (de Boer and Kuyper, 1968; Jones and Palmer, 1987). In the conventional space–time receptive field mapping, a spike-triggered average of stimuli itself (Fig. 1*A*) is calculated in the space–time domain. In LSRC, instead, we calculate a spike-triggered average of the amplitude spectra of a given subfield of the noise array (Fig. 1*C*) to obtain a two-dimensional frequency tuning for the given subfield (Fig. 1*E*). By interpreting the two-dimensional frequency tuning (Fig. 1*E*) as a polar coordinate representation, we obtain a joint spatial frequency and orientation profile. The distance from the origin to the peak of the excitation (shown in red in Fig. 1*E*) indicates the

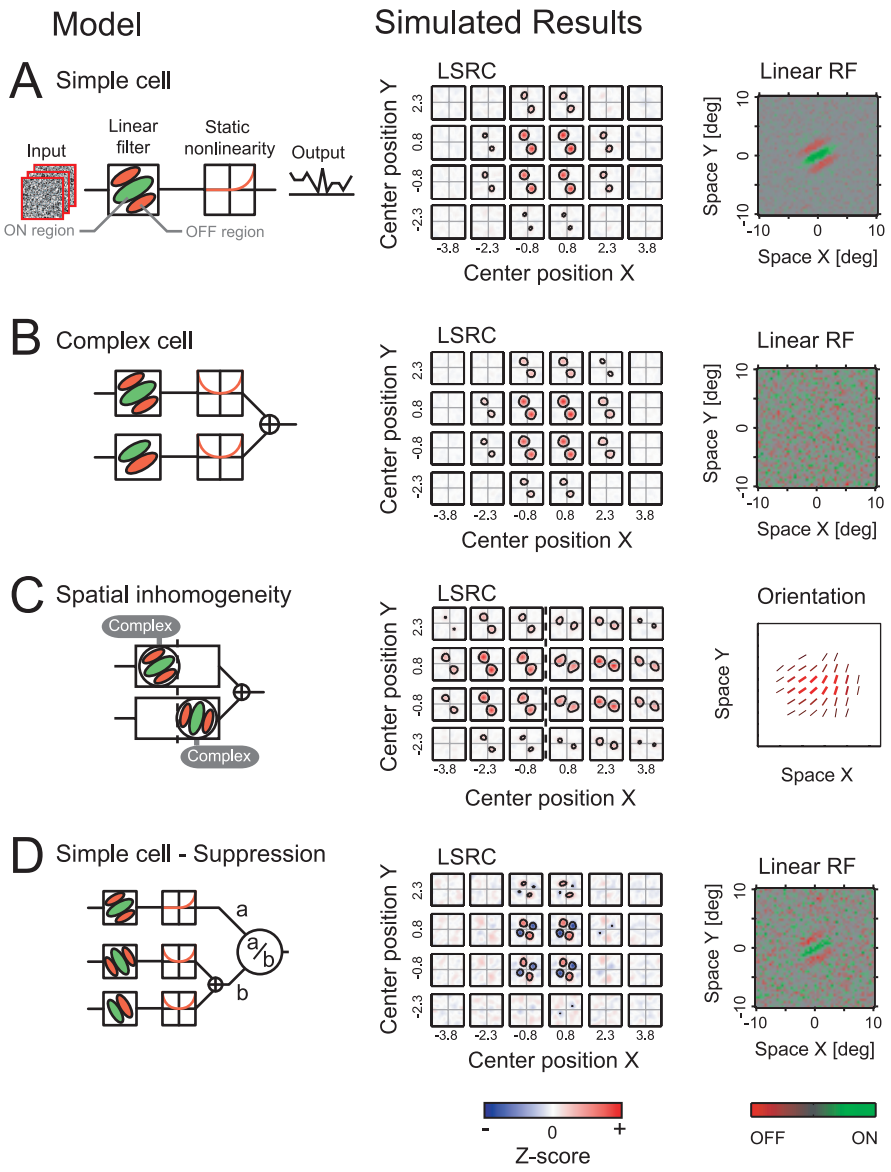


Figure 2. Results are shown of simulations conducted for four types of model neurons. The leftmost drawings depict structures of the models, and the right drawings show the simulated results from the LSRC and the standard reverse correlation analyses. In linear receptive field (RF) maps shown in the rightmost column, ON and OFF subregions are indicated by green and red, respectively, according to the scale bar at the bottom. The linear receptive fields in these simulations and experiments are obtained from the same data set as that used for LSRC. The vertical dashed lines in ***C*** indicate the border between the two complex cell components comprising the final output. ***D***, a and b represent facilitatory and suppressive inputs, respectively. deg, Degree.

optimal spatial frequency for the local subfield of the receptive field. Similarly, the angle of the line connecting the origin and the excitation peak (with the horizontal axis) depicts the optimal orientation for the local subfield.

By systematically changing positions of the subfield for calculating the spectra, we can obtain a matrix of two-dimensional frequency tunings (Fig. 1*F*), in which each element of the matrix contains the two-dimensional frequency tunings for the given subfield. Therefore, the final matrix of frequency tunings describes the tuning profile of the cell as a function of position (x, y) as well as spatial frequency and orientation in a joint manner. Note that we use Z-score values for representing the response strength in these spectral receptive field profiles throughout this report, to take variability and statistical significance of responses into account (see below). Z-score values may be negative, which may be interpreted as a reduction of activities below the baseline level.

The subfields were windowed by a two-dimensional Gaussian function, and the frequency spectra were calculated by the standard fast Fourier transform algorithm with zero padding (Press et al., 1992). The center of the window was stepped typically by σ of the Gaussian function, where σ is the SD.

We have calculated spike-triggered averages of stimulus local spectra for correlation delays from 0 to 150 ms in 15 ms steps. Then, the optimal correlation delay was determined as the delay for which the signal amplitude was maximal. Typical optimal correlation delays were 45 or 60 ms.

The average number of spikes for our population of cells was 7421 spikes per 30 min stimulation. The SD of the mean is 8225 spikes per 30 min. The minimum and maximum were 1073 and 51,013 spikes, respectively.

Statistical examination. To evaluate the significance of the spike-triggered signals, we calculated the average and SD (noise level) of signals using shuffled correlations. We obtained the shuffled correlations by calculating cross-correlations between spike trains and shifted (unpaired) stimulus blocks. The mean and the SD of the shuffled correlations were then used to normalize the original spike-triggered signals into Z-score representations. To reduce a computational burden, we assume that the noise level is identical for a sequence of random pattern of any given subfield and spatial frequency. Therefore, for each neuron, we calculate a set of mean and SD values of the shuffled correlations and use it as parameters for normalizing all spike-triggered signals for the neuron. The statistical significance of signals was examined by the Z-score, corrected for multiple comparisons by the Bonferroni's method. The degree of freedom for the Bonferroni's correction is set to the number of subfields multiplied by the number of noise elements within $\pm 1\ \sigma$ of the analyzing Gaussian window. Black curves in the LSRC plot indicate contours for $p = 0.05$.

Results

Here, we present two sets of results. One is from a set of simulation studies of the LSRC method, conducted to validate our new method itself and to examine how to interpret the results. The other is an experimental result of the LSRC analysis applied for visual neurons in the early visual cortex.

Simulation study

To examine whether the LSRC method works correctly to reveal spatially localized selectivity, results from a set of simulation studies are presented. We have calculated responses of several kinds of model neurons (Fig. 2) to white-noise stimuli and analyzed the response profile of these neurons by the LSRC method as well as a standard reverse correlation procedure (Jones and Palmer, 1987; DeAngelis et al., 1993; Reid et al., 1997).

Simple and complex cells

Figure 2*A* shows the structure of a model simple cell, together with results of the LSRC analysis and a linear receptive field profile obtained by the standard reverse correlation analysis. The instantaneous responses of the model simple cell are calculated by a linear filtering stage (a Gabor function), followed by a static

nonlinearity (a power-law, half-wave rectification). For this model simple cell and other cell types shown in Figure 2, simulations are performed by using a rate-coding model (Troyer et al., 1998) in which the output of the model is a scalar value that corresponds to the firing rate of the model neuron. The output value of the cell in Figure 2*A* is thus given by the following:

$$L(t) = \mathrm{Pos}^2[\int LF(x,y)S_t(x,y)dxdy)], \quad (1)$$

where $LF(x,y)$ is a weighting function (linear filter), $S_t(x,y)$ is a two-dimensional stimulus sequence, and $\mathrm{Pos}[v]$ is a half-rectification function, where $\mathrm{Pos}[v] = v$ for $v > 0$ and $\mathrm{Pos}[v] = 0$ otherwise.

Because our scope is limited to the spatial aspect of receptive fields in this study, the temporal dynamics of responses are not considered. Instead, the model was instantaneous and generated one output value for each stimulus frame. Furthermore, instead of generating spikes and computing spike-triggered average of stimuli, equivalent cross-correlation may be computed by multiplying each stimulus frame (or each local spectral amplitude for the case of LSRC) by the output of the model and summing the resulting patterns for all stimulus frames shown to the model cell. This computational procedure applies to both LSRC and standard space-domain reverse correlation analyses.

The results of the LSRC analysis on the model cell responses, shown as a matrix of selectivity in a two-dimensional frequency domain, show the position and spatial extent of the receptive field profile indicated by the limited number of maps with significant excitations. Note also that we can obtain the orientation and spatial frequency selectivity for each localized subfield, allowing us to examine the possible variations of these tuning properties within the receptive field. The recovered linear receptive field profile (Fig. 2*A*, right), as expected, shows a spatial structure consisting of ON and OFF regions as originally set in the linear stage of the model neuron.

Figure 2*B* shows results of another simulation for a model complex cell. Our model complex cell is based on the standard energy model (Adelson and Bergen, 1985; Pollen et al., 1989; Ohzawa et al., 1990; Emerson et al., 1992). Because complex cells do not possess spatially separated ON and OFF subregions, a standard reverse correlation procedure does not reveal any spatial structure (Fig. 2*B*, right). On the other hand, the LSRC method can visualize the position and spatial extent of response profile as well as two-dimensional frequency tunings as in the case of the simple cell. The ability to visualize the response profile of cells even with substantial nonlinearity, like complex cells, is one of the advantages of the LSRC method not available for the standard reverse correlation procedure. Although the spike-triggered average of stimuli (i.e., the output of the standard reverse correlation procedure) will be zero if the underlying nonlinearity is "symmetric" as in the energy model (Simoncelli et al., 2004), LSRC can reveal response profiles for both symmetric and asymmetric types of nonlinearities because the analysis is based on the absolute values of the spectral components.

Spatial inhomogeneity

Can the LSRC method reveal a response profile of neurons, if the selectivities are not homogeneous within their receptive field? This is an important question related to whether LSRC can reveal the profiles of the next-level neurons beyond complex cells, which may be organized to collect from neurons tuned to different parameters. To address this question, we modeled a spatially inhomogeneous neuron in which the orientation selectivity differs depending on the spatial position within the receptive field (Fig. 2*C*). The model neuron sums the output of two model complex cells (Fig. 2*C*, left), in which these two components differ in their preferred orientation by 45° and in their spatial positions of the receptive fields. As shown in the simulated result, LSRC can successfully recover the spatial inhomogeneity of the response profile. The two-dimensional frequency tuning for a subfield centered at (−2.3, 0.8), for example, shows a preference to the orientation of 30°, whereas the preferred orientation for a subfield centered at (2.3, 0.8) is 75°. These are exactly the configurations defined in the model. However, note that in the middle of these two locations, we see a tuning profile that is a mixture of those of the original component units and that gives an appearance as if the orientation tuning shifts smoothly over space. This is attributable to a smoothing or blurring effect resulting from the size of the Gaussian window used to compute the spectra. This is a generally applicable limitation for any localized spectral methods in which there is a trade-off between the resolution in the frequency domain and the original domain. Therefore, one limitation of the LSRC method is that an abrupt boundary in tuning parameters, such as orientation, will not be detected as such without using a smaller window size and consequently sacrificing the spectral resolution.

Suppressive profiles

In Figure 2*D*, we have examined the effect of suppressive components. Our model neuron consists of a simple cell-like facilitative component and a divisive, cross-orientation suppression [a special case of a model by Heeger (1992)]. The suppressive component is modeled as a complex cell-like energy unit because the suppressive effects are known to be phase invariant (DeAngelis et al., 1992). The results show that, although the standard reverse correlation reveals only the linear receptive field (Fig. 2*D*, right), LSRC shows spatial and spectral positions of suppression (blue areas), in which stimulus energy in the corresponding positions reduces the output of the model neuron. We also simulated a subtractive type of the suppression and found that LSRC can also reveal the subtractive type of suppressions (data not shown).

Overcoming high threshold and nonlinearities for studying higher-order neurons

Several previous studies show that cells selective to complex visual stimuli cannot generally be activated by stimulations using only a part of their optimal stimuli (Tanaka et al., 1991; Pasupathy and Connor, 2001; Ito and Komatsu, 2004). A possible neural mechanism that could explain this phenomenon is that the spiking threshold and nonlinearities are so large that individual parts of appropriate stimuli shown alone cannot achieve the excitation necessary for eliciting spikes. Rather, visual stimuli must contain the essential parts simultaneously in order for the summed excitation to overcome the spiking threshold. This type of nonlinearity is thought to be a source of the problem in mapping response profiles in a part-by-part manner using elementary stimuli such as a bar or a patch of grating (Pollen et al., 2002). The LSRC method may, at least partly, overcome this difficulty. Because large visual areas over the receptive field are always stimulated, by random chance, near optimal combinations of multiple local features can appear in the sequences. Because the white noise for different spatial areas are uncorrelated, the response profiles could be mapped for each local area. This means that the knowledge of selectivities of individual parts of the receptive fields may be obtained from a set of stimuli that covers the entire area.

To examine whether LSRC possesses these desired features for mapping response profiles, even in the situation that partial stim-

ulations cannot reveal them, we have conducted an additional simulation study as illustrated in Figure 3. The model neuron (Fig. 3*A*) is similar to the spatially inhomogeneous neuron as in Figure 2*C* but has a high spiking threshold that makes the cell unresponsive unless strong excitations are given. The result demonstrates that, although partial stimulations cannot reveal response profiles in a reliable manner (Fig. 3*C*), the full stimulations covering the entire receptive field reveal significant response profiles for each localized area (Fig. 3*B*). These results indicate that LSRC is a highly promising method for overcoming difficulties of mapping receptive field profiles of neurons that combine output of other neurons in a complex manner, without using assumptions regarding the specific details of the combination.

In summary, the simulation studies confirm that (1) LSRC can reveal two-dimensional frequency tunings and their spatial position and extent for both simple and complex cells, (2) LSRC can recover spatially inhomogeneous receptive field profiles, and (3) LSRC can also visualize suppressive profiles localized both in space and spatial frequency domains. Below, we apply the analysis for visual neurons in the cat early visual cortex.

Physiological study

The LSRC analyses were completed for a total of 193 cells (154 cells from area 17 and 34 cells from area 18, in 20 cats). The cortical area of recording is judged based on the coordinates of the electrode penetrations. Of these 193 cells, additional spatial frequency tuning measurements using drifting grating stimuli could be completed with sufficient reliability for 148 neurons. Seventy-seven of these cells were classified as simple, and 71 cells were classified as complex, according to the standard criteria (Skottun et al., 1991; Li et al., 2003; Priebe et al., 2004).

Local spectral selectivity

Figure 4 shows examples of the results for the LSRC and the standard reverse correlation analyses applied for a simple cell (Fig. 4*A–C*) and a complex cell (Fig. 4*D–F*) in area 17. These two types of analyses are conducted on the same data set for this and other cells. Although the standard reverse correlation procedure applied for the simple cell (Fig. 4*C*) yields a space–domain receptive field profile, that for the complex cell (Fig. 4*F*) shows no structure in this domain. In contrast, results of the LSRC analyses (applied to the same data) show clear response profiles, the two-dimensional frequency tunings as well as the spatial position and extent, for the both simple and complex cells. The spatial extent of the complex cell shows a horizontal elongation that is neither perpendicular nor parallel to the preferred orientation of the cell. The two-dimensional frequency tunings appear similar for all the profiles that contain signals, suggesting that the preferred orientation and spatial frequency is homogeneous throughout the receptive filed for these two cells.

The spatial homogeneity of the two-dimensional profile, however, is not always the case. Figure 5 shows another example for a complex cell in area 17. This cell shows a clear spatial inhomogeneity of the tuning characteristics within the receptive field. As seen in Figure 5, *B* and *C*, the two-dimensional frequency tunings differ substantially depending on the subfield location, in that the optimal orientation and spatial frequency differ substantially depending on the regions in which stimuli are presented. Figure 5*E* depicts a spatial arrangement of orientation tunings for each subfield. Assuming that response amplitude of the cell is dependent on the weighted sum of local features, a stimulus containing a curvature would be optimal for this cell. Although we

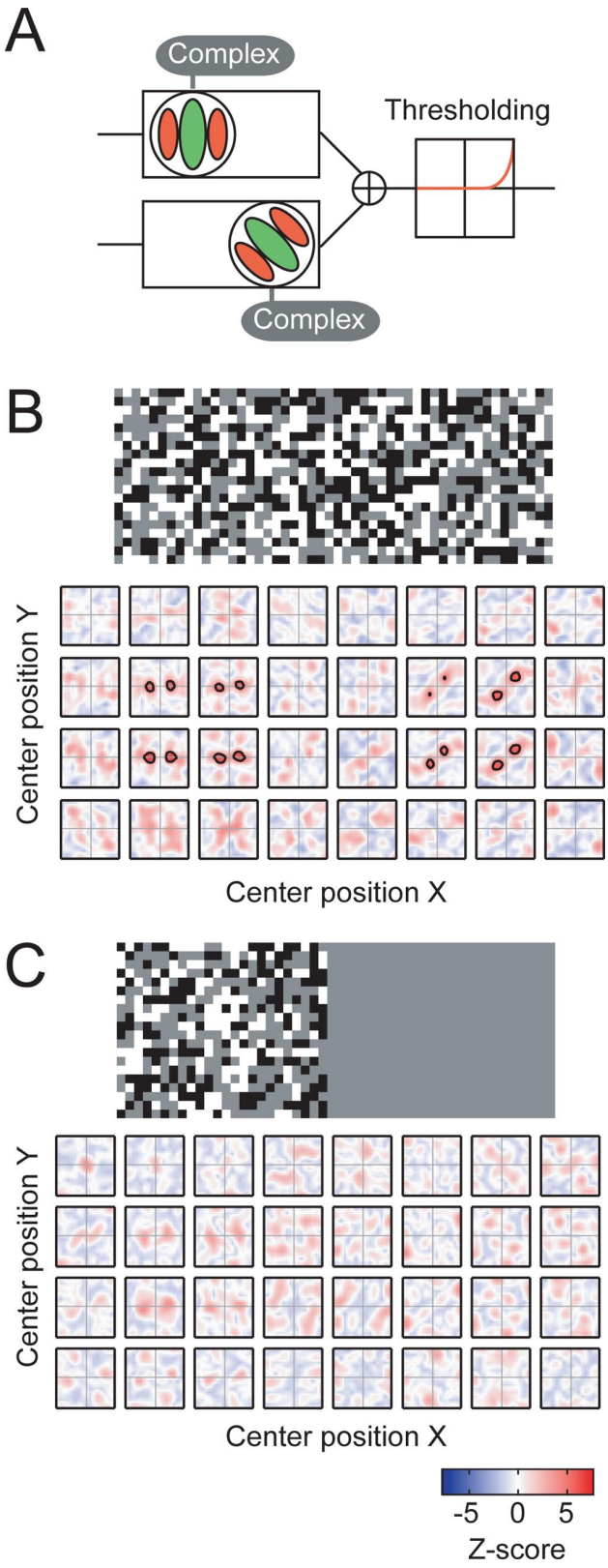


Figure 3. Benefits of the LSRC analysis in studying a higher-order neuron with high threshold are illustrated by simulation. ***A***, Schematic diagram for a model neuron. The model is a spatially inhomogeneous neuron with two complex-type subunits as in Figure 2*C*, but the subunits have a relatively high firing threshold that prevent the cell from firing unless strong excitatory input is given. The simulations of the LSRC analysis were conducted for two different conditions: ***B***, Full-field noise array that stimulates both subunits simultaneously. ***C***, Half-field noise array in which the right half of the stimuli was masked, providing stimulation of only one of the two subunits. The neuron does not respond unless the entire receptive field is stimulated.

did not test the neuron with such stimuli directly, this may mean that visual processing of complex features found in the higher-order areas (Gallant et al., 1993, 1996; Hegde and Van Essen, 2000; Pasupathy and Connor, 2001; Ito and Komatsu, 2004) is started, at least partly, in the early visual cortex.

How prevalent are the neurons with spatial inhomogeneity in the early visual cortex? To address this question, we summarize the degree of spatial inhomogeneity of orientation and spatial frequency tunings for our population of cells (Fig. 6). For each neuron, we have calculated the maximum difference of optimal parameters, both for orientation and spatial frequency, among profiles of subfields that contain significant signals ($p < 0.01$; Bonferroni corrected). Most of the cells in the early visual cortex show generally homogeneous profiles as in Figure 4, and only a small subset of cells shows the spatial inhomogeneity as shown in Figure 5 (this cell is indicated by a black arrow in Fig. 6). There is no significant difference in distributions of maximum differences for both parameters (orientation and spatial frequency) between areas 17 and 18 (two-sample Wilcoxon test; $p > 0.1$).

Care must be used, however, in interpreting the apparent inhomogeneities shown in Figure 6. This is because intrareceptive field variations of filtering properties may arise simply because of our procedure in examining the profile using small analyzing windows. For instance, if an analysis window is too small and covers only a part of an ON region of a simple cell receptive field, the resulting spectrum would primarily be that of the Gaussian analysis window itself, which is low-pass, not bandpass as expected from the entire receptive field. Therefore, such artifacts of the procedure may cause apparent intrareceptive field variations of tuning parameters for both spatial frequency and orientation.

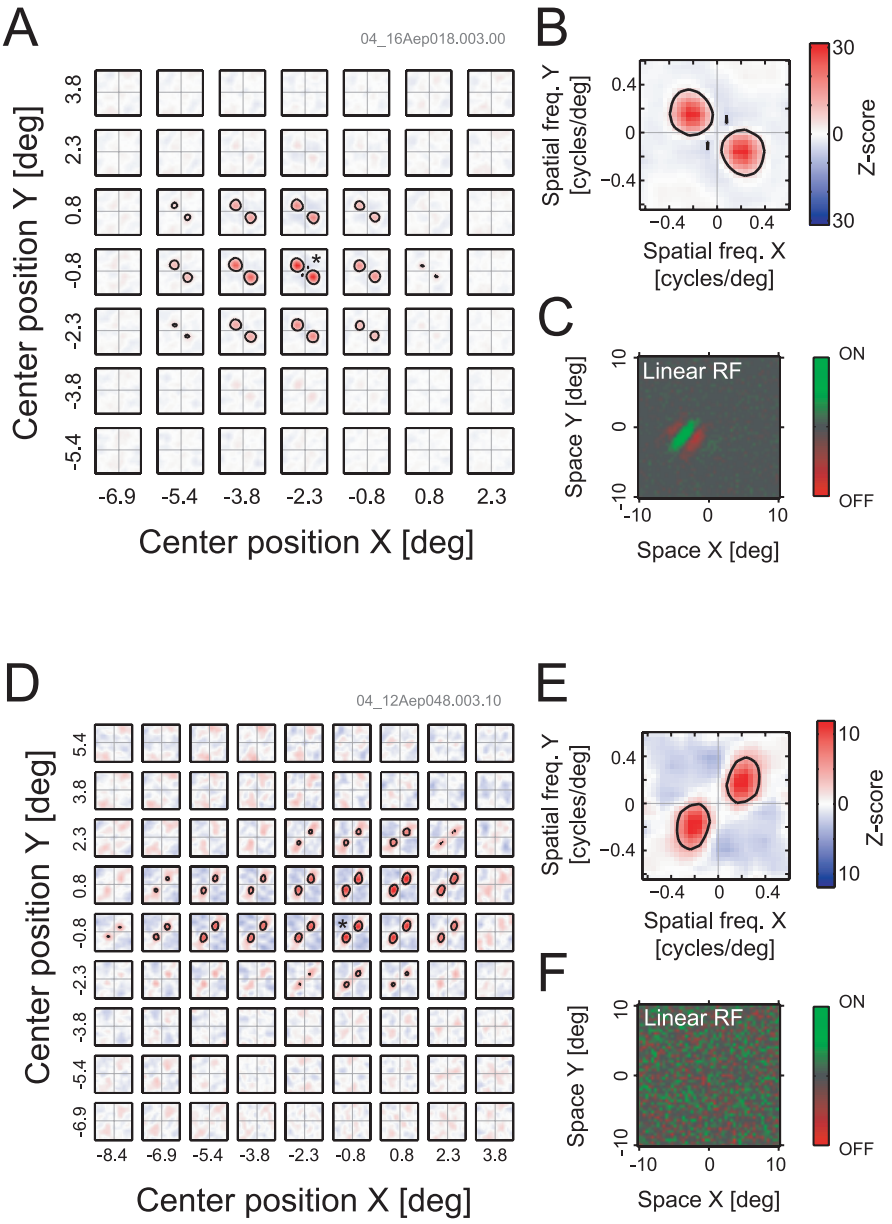


Figure 4. Local spectral selectivities are depicted for a simple (***A–C***) and a complex (***D–F***) cell. ***A***, A spatial matrix of local spectral selectivity maps for a simple cell in area 17. Each individual plot shows a tuning property in a two-dimensional spatial frequency domain for the corresponding spatial subfield. The spectral selectivity maps are arranged to reflect the spatial positions of the corresponding subfields. ***B***, A detailed profile of the most responsive subfield indicated by an asterisk in ***A***. ***C***, A linear receptive field profile calculated from the same data as for ***A***. ***D–F***, Data for a complex cell in area 17 in the same format as that for ***A–C***. deg, Degree; freq., frequency; RF, receptive field.

To examine how much intrareceptive inhomogeneities the LSRC procedure itself might induce, we have performed simulations using a model simple cell as in Figure 2*A*. Specifically, model simple cells with Gabor-shaped spatial receptive field profiles were tested using the LSRC procedure, and the methodologically induced variations were examined. We have simulated 1000 model cells, each of which has different parameters selected randomly from the physiologically realistic ranges. Table 1 shows the range of parameters we have used for the simulations based on our physiological data. For each model cell, we have performed the LSRC procedure and calculated maximum variations of optimal parameters for both orientation and spatial frequency, as we performed for the real data. The mean variations for simulated cells are 7.2° for orientation and 0.32 octaves for spatial frequency. The ellipses in Figure 6 show 95 and 99% confidence limits of variations determined by the simulations. We also performed a similar test using model complex cells, but the trend is essentially identical to the case of the simple cell models (data not shown). The result indicates that the most of the small variations within these ellipses are indistinguishable from variations induced by the LSRC procedure itself. Clearly, however, there were cells, even in the early visual cortex, that exhibited large intrareceptive field variations of tuning parameters, which cannot be attributed to the methodological factors.

Measurement of phase dependency

So far, our primary concern has been the analyses of the spike-triggered averages of the absolute spectral amplitude, and the spatial phase dependency has been ignored. However, Fourier

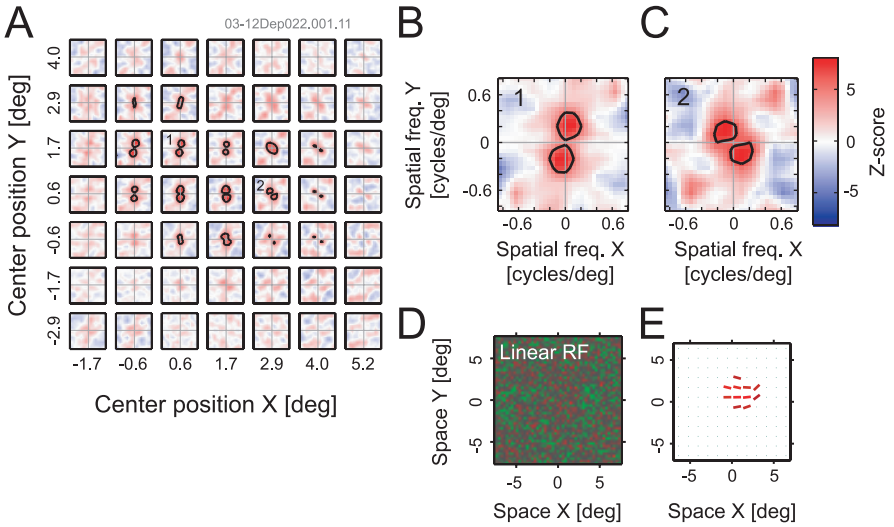


Figure 5. Local spectral tuning data are shown of a cell that exhibits substantial spatial inhomogeneity of orientation and spatial frequency tuning within its receptive field. ***A***, Local spectral selectivities for a complex cell in area 17. ***B***, ***C***, Detailed tuning properties in ***A***, as indicated by numbers (1 and 2). ***D***, A linear receptive field (RF) profile of the cell. ***E***, A spatial arrangement of orientation tunings obtained from ***A***, in which the orientation of the bars indicates the preferred orientation of the corresponding subfields. Only the data for subfields that contain significant signals are shown ($p < 0.01$; Bonferroni corrected). This cell was nondirection selective, based on a test with conventional drifting grating stimuli. deg, Degree; freq., frequency.

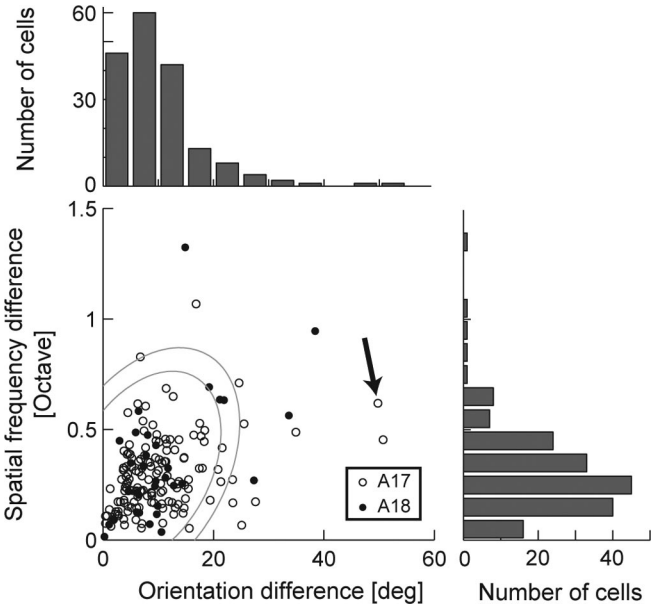


Figure 6. A population summary is shown of the spatial inhomogeneity of tuning parameters across the receptive fields. The horizontal axis indicates the maximum difference of the preferred orientations among multiple local spectral tuning maps. The vertical axis indicates the same for spatial frequency. Open and filled symbols indicate data for cells in area 17 (A17) and area 18 (A18), respectively. The arrow indicates the cell shown in Figure 5. For this figure, 178 neurons with bandpass spatial frequency tuning profiles and with more than two significant frequency domain maps are included for computing differences of tuning parameters. Two ellipses represent 95 and 99% confidence limits of variations from simulations as a control, assuming homogeneous properties (see Results for details). deg, Degree.

transforms contain information not only of spectral amplitude for each frequency component, but also that of phase. To use all of the information available with the LSRC method, we now extend the method to include both the amplitude and the phase. Because the phase dependence is the key feature that separates simple cells from complex cells (Movshon et al., 1978a; De Valois et al., 1982), incorporating phase information to the LSRC method will allow us to analyze these conventional cell types from a new perspective.

To gain an understanding on how phase dependence may be extracted from the data, we must first clarify how phase dependence metrics are defined statistically based on individual spike data and stimulus frames. The red dots in Figure 7, *A* and *D*, show distributions of unaveraged Fourier coefficients for spike-triggered stimuli (for the optimal correlation delay) for the simple and complex cells shown in Figure 4, respectively. These distributions are for the optimal spatial frequency, orientation, and position (i.e., the conditions corresponding to the peaks of Figure 4, *B* and *E*). Because a Fourier coefficient is a complex number, each coefficient is plotted as a point on a two-dimensional complex plane with a real and an imaginary component. In this domain, the distance and angle of the dot from the origin indicate the amplitude and spatial phase, respectively, of a sine wave of the given frequency that is contained in the relevant region of the stimulus.

Note that the centroid of red points in Figure 7*A* is biased toward the bottom left side. This corresponds to the fact that the simple cell tends to respond to stimuli of a particular phase and tends not to respond to the anti-phase stimuli. For comparison, if we plot Fourier components for all stimulus frames for the corresponding condition without regard for spikes from the neuron, we obtain a distribution depicted by gray dots in Figure 7*A* (there are more gray dots than red ones). The distribution is unbiased with respect to the origin, indicating that the noise stimulus sequence contain a homogeneous distribution of Fourier components with respect to spatial phase. On the other hand, the complex cell did not show such phase dependency as illustrated in Figure 7*D*, where the distribution of red dots, the spike-triggered Fourier coefficients, is unbiased and centered nearly exactly at the origin. The distribution of gray points are hidden behind the red dots in Figure 7*D*.

Therefore, the magnitude of the phase dependency can be determined from the bias in the distribution of Fourier coefficients for spike-triggered stimuli in the complex domain. We quantify the bias by calculating a vector sum of the spike-triggered Fourier coefficients and define the phase selectivity index (PSI), for each frequency and position, as follows:

$$\mathrm{PSI} = \frac{|\Sigma f_{\mathrm{spike}}|/n_{\mathrm{spike}}}{\Sigma |f_{\mathrm{total}}|/n_{\mathrm{total}}}, \qquad (2)$$

where f_{spike} are spike-triggered Fourier coefficients, f_{total} are the Fourier coefficients for the entire stimuli, n_{spike} is the number of spikes, and n_{total} is the number of total frames in the entire stimulus sequence. The PSI should be high when a cell responds in a phase-dependent manner as in Figure 7*A* and is close to 0 when there is no phase dependency as Figure 7*D*.

By using the PSI, we can obtain a spatial map of phase dependency together with the signal magnitude. Such a map allows us to examine possible spatial variations, if any, of local phase sensitivity within a given receptive field. Figure 7, *C* and *F*, shows

Table 1. Simulation parameters for estimating inhomogeneities of methodological origin

	Carrier orientation (°)	Envelope orientation (°)	Carrier phase (°)	Spatial frequency (cycles/stimulus area)	Envelope sigma	Envelope aspect ratio
Minimum	0	0	0	3.0	1.5/SF*0.7	0.7
Maximum	180	180	360	4.5	1.5/SF*1.3	1.3

Ranges of parameters are shown for simulations for determining confidence intervals of interreceptive field variations (see Fig. 6), that might be induced because of the LSRC procedure. See Results for details. SF, Spatial frequency.

spatial "amplitude-phase" maps for the simple and complex cells, respectively. In these plots, the signal magnitude for each subfield is represented as luminance, whereas the PSI and the preferred spatial phase are shown as saturation and hue, respectively. In these representations, only the PSI values for a fixed (optimal) spatial frequency were used. Spatial variations of the phase dependency for the simple cell can be seen by its map with highly saturated colors in Figure 7*C*. A similar map for the complex cell (Fig. 7*F*) consists of points with highly desaturated (almost white) colors, because the cell shows little phase dependency.

The PSI should have a close relationship to the conventional classification of the simple and complex cell. To what extent is the PSI correlated with the modulation ratio (Skottun et al., 1991; Li et al., 2003; Priebe et al., 2004), the standard criteria for classifications of simple and complex cells? Figure 8 summarizes the result. In this figure, only the PSI value for the spatial position with maximal Z-score is used for each neuron. There is a significant correlation between the PSI and the modulation ratio ($p \ll 0.01$; test for Spearman's correlation coefficients). Therefore, our result opens a possibility of classifying simple and complex cell types based on the dense noise mapping data alone, which was not possible previously because the linear receptive field profiles typically do not show any structure and are indistinguishable noise for many complex cells (Fig. 7*E*).

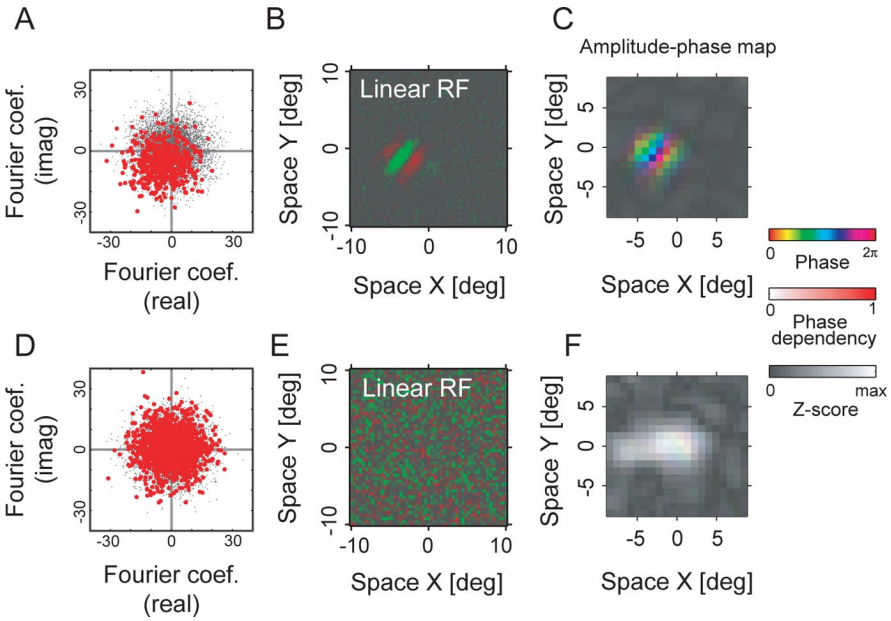


Figure 7. Spatial phase sensitivities are calculated for a simple cell (***A–C***) and a complex cell (***D–F***). ***A***, Fourier coefficients for the maximally effective spatial frequency component in stimuli that led to spikes are shown (red dots) for the optimal correlation delay of 45 ms. The centroid of red dots is offset from the origin indicating selectivity for a given phase. For estimating the distribution for the noise stimuli themselves, Fourier coefficients for the same frequency component are also shown for all frames of the subfield of noise sequence (gray dots). Only the coefficients for one frequency component for the maximally responsive subfield are used. ***B***, Spatial receptive field map obtained by a standard reverse correlation procedure. ***C***, Spatial structure of the phase dependency. The optimal spatial phase, PSI (see Results), and signal amplitude of the LSRC analysis are represented as hue, saturation, and brightness, respectively. ***D–F***, The same as ***A–C***, but for a complex cell. The two cells are the same as those shown in Figure 4. For ***C*** and ***F***, the maximum values for the Z-scores were 34.1 and 17.5, respectively. coef., Coefficient; imag, imaginary; deg, degree; RF, receptive field; max, maximum.

Profiles of suppression

The LSRC analysis can also visualize suppressive profiles of visual neurons, as suggested by the simulation study (Fig. 2*D*). Figure 9*A–D* shows an example cell that exhibits clear suppressive components. Although the linear receptive field (Fig. 9*C*) only captures the facilitatory profile, LSRC reveals the existence of both facilitatory and suppressive components as indicated by the red and blue regions, respectively, in Figure 9, *A* and *B*. The suppression appears to be strongest at an orientation approximately orthogonal to that for facilitation. However, we should use caution in interpreting the results regarding how the suppression is organized as a function of stimulus orientation. Figure 9*D* shows an orientation tuning profile obtained from a conventional drifting grating test. Responses to the off-peak orientation are less than the spontaneous discharge rate of the neuron, which is a reflection of the suppressive effect. However, this suppression seems to be present for virtually all orientations, except for the peak, and not just for the orientation orthogonal to the optimal (DeAngelis et al., 1992). The LSRC analyses calculate the net sum of facilitation or suppression for each frequency and thus can only visualize

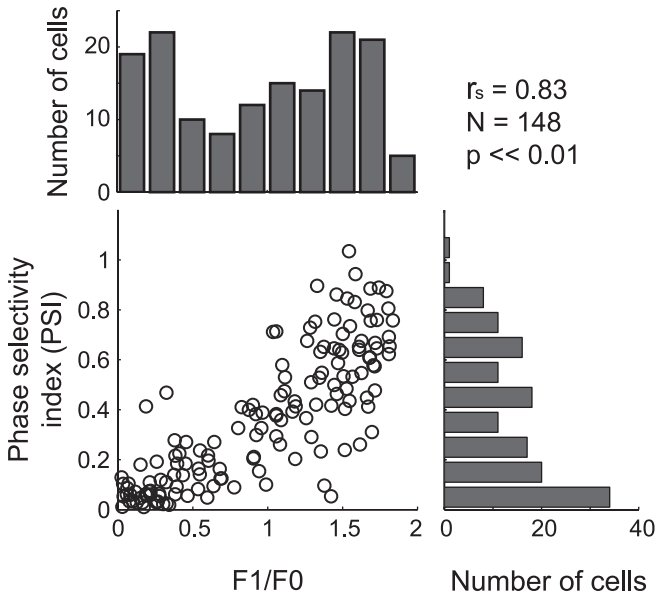


Figure 8. A relationship is shown between the PSI and the conventional modulation ratio (F1/F0). The histograms at the top and the right show distributions of these indices separately. There is a significant correlation between the PSI and F1/F0 ratio ($p < 0.01$; test for the Spearman's correlation coefficient).

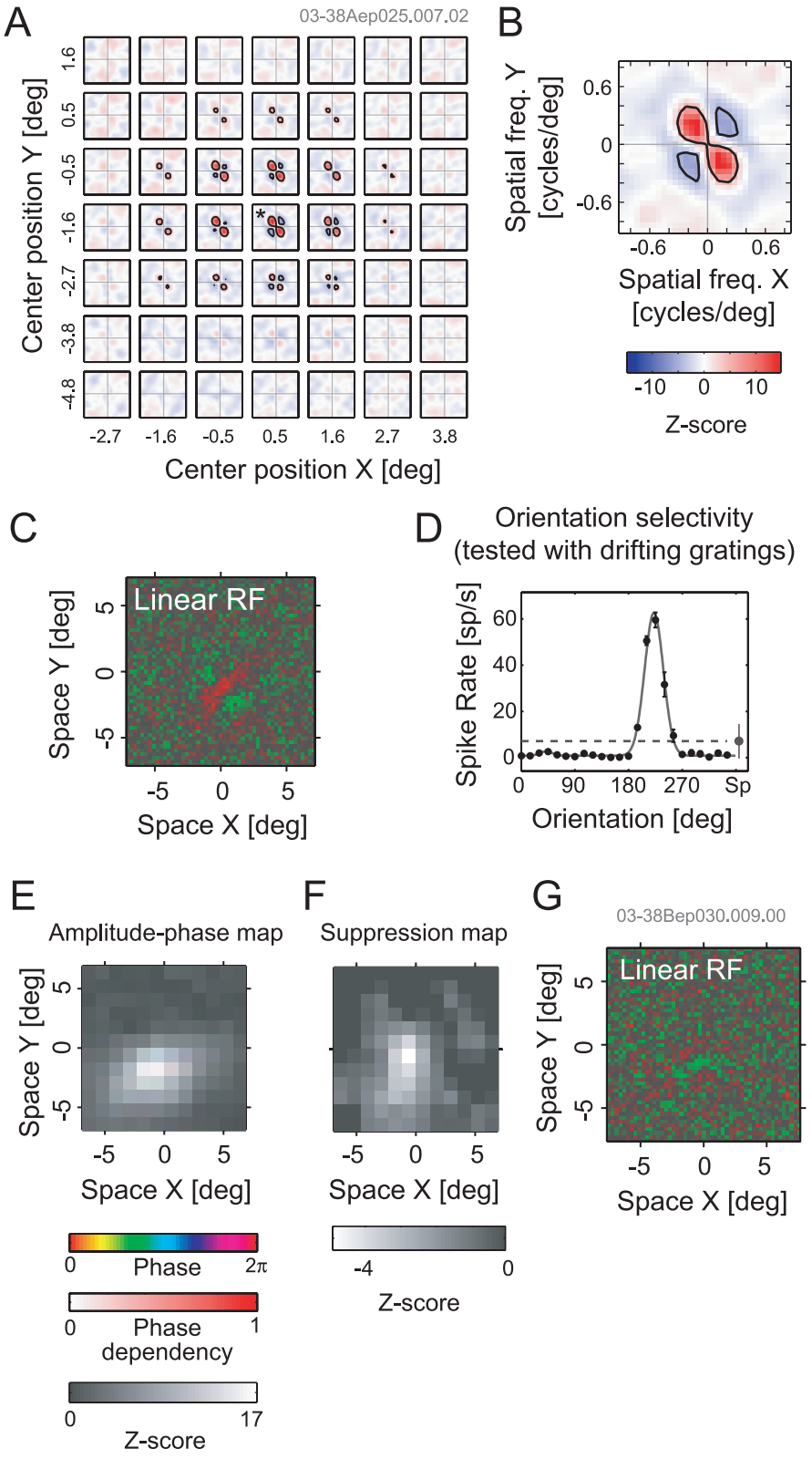


Figure 9. Data from cells with suppressive responses are shown for two complex cells in area 18. ***A***, A result of the LSRC analysis. The reddish regions (Z-score >0) indicate facilitatory components, whereas the bluish regions (Z-score <0) show suppressive components. ***B***, Magnified tuning for a local spectral map, indicated by an asterisk in ***A***. ***C***, Linear receptive field profile. ***D***, An orientation tuning curve obtained by a conventional drifting grating test. The dashed horizontal line indicates the spontaneous firing rate. ***E–G***, Data from another neuron. ***E***, Amplitude and phase map for facilitatory responses in the same format as that of Figure 7, *C* and *F*. The facilitation was evaluated at the spatial frequency and orientation of 0.33 cycles/degree and 18°. ***F***, Suppression map is shown using the Z-score only. The suppression was evaluated at the spatial frequency and orientation of 0.23 cycles/degree and 117°. ***G***, Linear receptive field for the second neuron. deg, Degree; freq., frequency; RF, receptive field.

whichever is stronger. Therefore, it should be noted that we cannot discriminate the following possibilities apart: whether the suppressive effects exist for all orientations of the frequency range overlapped to the facilitatory one or whether they exist just for orientations nearly orthogonal to the optimal.

Even with the limitations noted above, the ability of the LSRC method for mapping the degree and spatial parameters of suppression in addition to facilitation would be useful in general for examining potentially inhomogeneous properties of the response profile of the cell. Figures 9*E–G* shows another example cell exhibiting a form of inhomogeneity, in that spatial areas for facilitation (*E*) and suppression (*F*) are not exactly overlapped. The facilitation and the suppression were evaluated at different spatial frequencies and orientations where each was most predominant. Although the facilitatory area appears to be elongated horizontally, the suppressive area shows a vertical elongation and is smaller than that for the facilitation. The smaller spatial extents for the suppression are consistent with findings of a previous study (DeAngelis et al., 1992).

Figure 10 shows two-dimensional frequency tuning profiles of four cells that exhibit strongest suppressions among our data. The optimal frequency for suppression is neither always orthogonal in the orientation nor identical in their spatial frequency to the facilitatory peak (Fig. 10*B*). Although several previous reports have also pointed out that optimal spatial frequencies for facilitation and suppression are not always identical (Bonds, 1989; DeAngelis et al., 1992), determining parameters for the optimal suppressive stimuli has been quite difficult based on one-dimensional tests. For example, to obtain the suppressive spatial frequency tuning, one had to select an orientation for the suppressive stimulus and vary its spatial frequency. With such tests, even a strong suppression like the one shown in Figure 10*B* could have easily been missed because both the optimal suppressive orientation and spatial frequency are different from the typical values of these parameters. Two-dimensional tests in the joint orientation and spatial frequency domain, such as those used by Ringach et al. (2002) and ours, are generally required for accurately determining optimal parameters of suppression. It is somewhat puzzling that, among our population of cells, only 10 of 193 cells showed significant suppressive profiles (*t* test with Bonferroni's correc-

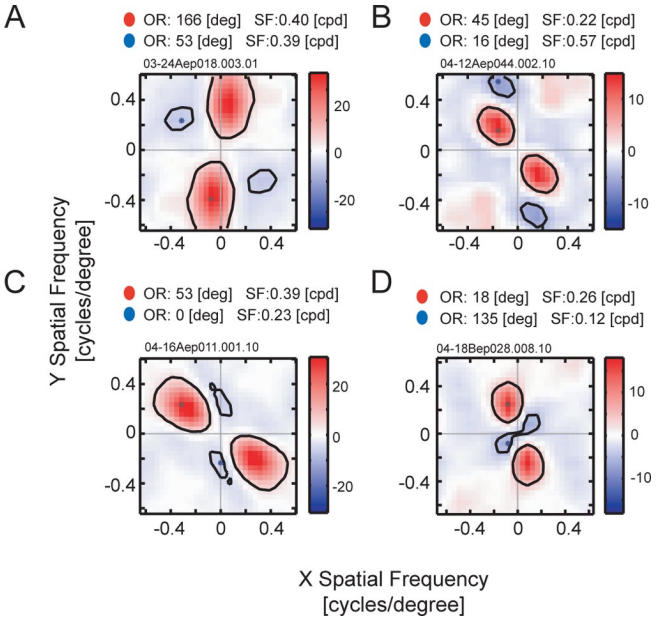


Figure 10. Two-dimensional spatial frequency tunings are shown for four additional cells that exhibit significant suppressions in the LSRC analysis. ***A***, A two-dimensional spatial frequency tuning for the subfield with the strongest suppression for a complex cell in area 17. ***B***, A similar tuning map for a simple cell in area 17. ***C***, ***D***, Similar plots for two complex cells in area 17. OR, Orientation; SF, spatial frequency; deg, degree; cpd, cycles per degree.

tion; $p < 0.01$). Previous work based on superimposed drifting sinusoidal gratings (plaid) stimuli seems to find some degree of cross-orientation suppression for most neurons (DeAngelis et al., 1992). It might be related to the difference in the type of stimuli used (dense noise vs plaids), because it is known that response profiles, especially the suppressive profiles, are different depending on the class of stimuli used to acquire tuning profiles (David et al., 2004). It may also be related to the fact that we use Bonferroni's correction for quantitative estimates of the strength of facilitation and suppression. This correction may have been too conservative. Another point we should consider is that uncovering suppression depends on the mean firing rate of the cell to the noise stimulus. In any case, our sample size does not allow summaries of the relationship between the facilitatory and suppressive parameters of these neurons. Resolution of these issues requires comparative studies of suppression using both dense noise and grating stimuli.

Discussion

Relationship to previous studies

Previous methods for mapping receptive fields and stimulus selectivities of neurons have certain advantages but also suffer from various shortcomings. For example, standard dense noise receptive field mapping generally allows measurements of first-order receptive fields, making it suitable for mapping simple cell receptive fields (Alonso et al., 2001). However, mapping attempts generally fail for complex cells and neurons in higher-order visual areas with substantial nonlinearities. Although it is theoretically possible to measure second- and higher-order receptive field maps, the amount of time required for measurements generally becomes prohibitively long in practice. Therefore, despite the advantage of having the least number of assumptions about what visual features neurons may be sensitive to, the white-noise stimuli have only been moderately effective. Alternative approaches, adopted by the majority of recent studies, have been based on a finite set of computationally generated complex stimuli rich in curved elements, such as non-Cartesian gratings (Gallant et al., 1993, 1996) and curvature-direction stimuli (Pasupathy and Connor, 2001). Although these analyses have been very effective in revealing key stimulus features that excite neurons, the stimulus sets are inherently finite, and assumptions about possible domains of selectivities are built into the stimuli implicitly. Ideal stimulus sets therefore would be those (1) with infinite possible configurations, (2) minimum assumptions, and (3) applicability for all cell types and possible neural connections. LSRCs have many of such ideal properties and could provide an experimental framework for revealing selectivity to the complex visual features.

Recently, several groups have shown that neurons in the primary visual cortex can be described as a set of spatiotemporal linear filters, and these underlying filters can be estimated by conducting a spike-triggered covariance (STC) technique (Touryan et al., 2002; Rust et al., 2004, 2005). The STC and LSRC analyses seem to share several desirable features, especially in that both of these techniques attempt to reveal filtering profiles underlying the responses of neurons and that both of them use white-noise sequences. A notable advantage of LSRC would be its efficiency. Although the LSRC procedure is essentially a first-order approximation of filtering profiles, the STC procedure belongs to a class of second-order approximations and thus need more spikes to map underlying properties. Practically speaking, whereas STC requires several tens of thousands of spikes to generate response profiles of reasonable signals (Rust et al., 2005), LSRC needs only a few thousands of spikes to obtain significant signals, and we could obtain excellent profiles with as few as ~5000 spikes. In our recording sessions (two-dimensional white-noise stimulation in anesthetized cats), we rarely encounter cells that elicit several tens of thousands of spikes within a typical recording time of ~30 min. Therefore, LSRC may be applicable to a wider variety of visual areas than STC, especially the areas where the white-noise sequences could elicit relatively a small number of spikes.

Inhomogeneity of response profiles

One of our key motivations for developing LSRC was to find possible spatial inhomogeneities of response profiles, such as local variations of preferred orientations within a given receptive field, that could serve to produce selectivities to complex visual features, particularly those with curved elements. By applying LSRC to cells in the early visual cortex, we found two types of spatial inhomogeneities: (1) a small subset of neurons shows local variations of preferred orientation and spatial frequency within their receptive field; and (2) spatial extents for facilitatory and suppressive components are not always overlapped exactly. Although the proportion of neurons that possess spatial inhomogeneities is not large in the early visual cortex, the existence of these classes of cells may, nevertheless, mean that the processing of complex object features begins at this stage. It would also supply baseline data for selectivity variations within the receptive fields in the early visual cortex, which would be valuable in constraining possibilities of building up downstream neurons that are much more selective to local feature combinations.

Parameter selections in the LSRC analysis

In this study, we have performed the LSRC measurements with relatively high-density stimuli (dot size, 0.2–0.4°) to ensure the ability to reveal receptive field structures tuned to high spatial frequency components. The Nyquist frequency for our typical configurations is 1.25–2.5 cycles/degree, which is reasonably

higher than the frequency that is known to be signaled by cells in the early visual cortex of cats (Movshon et al., 1978b; Zhou and Baker, 1994). Although, theoretically, we could use even smaller dots to increase the Nyquist frequency, the average power within each frequency band (thus ability to elicit spikes) would decrease for stimuli consisting of small dots. Trials would be needed to determine reasonable ranges of dot density when applying LSRC to other visual areas.

In the LSRC procedure, we are able to choose arbitrarily the position, size, and steps of the Gaussian window (i.e., the area over which the spectrum is computed) after the experiments are completed. This is one of the advantages of the LSRC method, because we need not be concerned about the exact position and boundary of the receptive field during the experiments. In other words, the spatial parameters of the analysis may be optimized *post hoc* via trials on the data. These features make LSRC particularly suitable for recordings via large multielectrode arrays. Receptive fields of many cells recorded from such electrodes may span a substantial area of the visual field as well as being tuned to a wide range of parameters. However, in the analyses, we should choose carefully the size of the window. If the size is too small, we cannot acquire proper response profiles for low spatial frequency spectra. On the other hand, if the window size is too large, we lose spatial resolutions. We selected our size of analysis such that the analysis window covers at least a half of the period of the optimal spatial frequency within 1 σ of the Gaussian if the cell shows clear bandpass profiles in their spatial frequency tunings. In rare cases in which neurons had a low-pass spatial frequency tuning, we used the σ value corresponding to one-fifth of the mapped area.

Possible applications and limitations

In this study, we have limited our scope only to the spatial aspects of receptive field profiles, and the LSRC analyses were performed only in the two-dimensional spatial frequency domain. However, the LSRC analysis can naturally be extended to include the temporal domain for studying response profiles in a joint three-dimensional (two-dimensional spatial and a temporal) frequency domain. Because there is physiological evidence that temporal property of cross-orientation inhibition is different from the facilitatory profile (Allison et al., 2001), it is of interest to examine whether the suppression and facilitation could be mapped separately in the three-dimensional frequency domain.

It is natural to think of the applications of LSRC for cells in the higher-order visual areas, such as V2 and V4. Because the cells in these areas are known to respond to more complex stimuli (Gallant et al., 1993, 1996; Pasupathy and Connor, 2001; Ito and Komatsu, 2004), it is of interest to examine how local spectral response profiles are organized for cells in these areas. However, the current LSRC method is probably not applicable to neurons with strong position invariance such as those in the inferotemporal cortical area (Ito et al., 1995; Tanaka, 1996), because tunings for given oriented segments are not tied to specific locations within the receptive field in these areas.

The area MT is another candidate for the application of LSRC. There is a well known model for MT pattern motion-selective neurons (Movshon et al., 1986) by Simoncelli and Heeger (1998). In their model, MT neurons are constructed by summing output of V1 neurons satisfying the constraints for a given velocity. The spatiotemporal LSRC analysis, as describe above, should be able to provide response profiles in the three-dimensional joint domain (spatial frequency, orientation, and temporal frequency) and thus could be used to assess the validity of the model directly, extending the work by Perrone and Thiele (2001). Position invariance is expected for responses of MT neurons. However, it is not a limitation in this case, because LSRC is not used to detect spatial inhomogeneity. Instead, it is used to determine whether the spatiotemporal frequency domain receptive field of MT neurons has a planar organization as proposed by Simoncelli and Heeger (1998).

In conclusion, LSRC is a highly general and efficient method for characterizing neurons in intermediate-stage visual areas beyond the primary visual cortex. The use of random noise stimuli, with the minimum of assumptions about what the cells might be "looking for," makes the LSRC method particularly suitable for multielectrode, multicell recordings. This is because, at least for initial bulk characterizations, it is desirable not to optimize stimulus parameters only for a selected set of neurons. Therefore, our study provides a basis on which the results from other areas may be compared with respect to inhomogeneities of tuning properties within receptive fields.

References

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

Allison JD, Smith KR, Bonds AB (2001) Temporal-frequency tuning of cross-orientation suppression in the cat striate cortex. Vis Neurosci 18:941–948.

Alonso JM, Usrey WM, Reid RC (2001) Rules of connectivity between geniculate cells and simple cells in cat primary visual cortex. J Neurosci 21:4002–4015.

Bonds AB (1989) Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. Vis Neurosci 2:41–55.

David SV, Vinje WE, Gallant JL (2004) Natural stimulus statistics alter the receptive field structure of v1 neurons. J Neurosci 24:6991–7006.

DeAngelis GC, Robson JG, Ohzawa I, Freeman RD (1992) Organization of suppression in receptive fields of neurons in cat visual cortex. J Neurophysiol 68:144–163.

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.

de Boer R, Kuyper P (1968) Triggered correlation. IEEE Trans Biomed Eng 15:169–179.

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

Dreher B (1972) Hypercomplex cells in the cat's striate cortex. Invest Ophthalmol 11:355–356.

Emerson RC, Bergen JR, Adelson EH (1992) Directionally selective complex cells and the computation of motion energy in cat visual cortex. Vision Res 32:203–218.

Gallant JL, Braun J, Van Essen DC (1993) Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. Science 259:100–103.

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.

Heeger DJ (1992) Normalization of cell responses in cat striate cortex. Vis Neurosci 9:181–197.

Hegde J, Van Essen DC (2000) Selectivity for complex shapes in primate visual area V2. J Neurosci 20:RC61(1–6).

Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. J Physiol (Lond) 195:215–243.

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

Ito M, Tamura H, Fujita I, Tanaka K (1995) Size and position invariance of neuronal responses in monkey inferotemporal cortex. J Neurophysiol 73:218–226.

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

Li B, Peterson MR, Freeman RD (2003) Oblique effect: a neural basis in the visual cortex. J Neurophysiol 90:204–217.

Morrone MC, Burr DC, Maffei L (1982) Functional implications of cross-orientation inhibition of cortical visual cells. I. Neurophysiological evidence. Proc R Soc Lond B Biol Sci 216:335–354.

Movshon JA, Thompson ID, Tolhurst DJ (1978a) Spatial summation in the

receptive fields of simple cells in the cat's striate cortex. J Physiol (Lond) 283:53–77.
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 (Lond) 283:101–120.
Movshon JA, Adelson EH, Gizzi MS, Newsome WT (1986) The analysis of moving visual patterns. In: Experimental brain research supplementum II: pattern recognition mechanisms (Chagas C, Gattass R, Gross C, eds), pp 117–151. New York: Springer.
Nishimoto S, Arai M, Ohzawa I (2005) Accuracy of subspace mapping of spatiotemporal frequency domain visual receptive fields. J Neurophysiol 93:3524–3536.
Ohzawa I, DeAngelis GC, Freeman RD (1990) Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. Science 249:1037–1041.
Ohzawa I, DeAngelis GC, Freeman RD (1996) Encoding of binocular disparity by simple cells in the cat's visual cortex. J Neurophysiol 75:1779–1805.
Pasupathy A, Connor CE (2001) Shape representation in area V4: position-specific tuning for boundary conformation. J Neurophysiol 86:2505–2519.
Perrone JA, Thiele A (2001) Speed skills: measuring the visual speed analyzing properties of primate MT neurons. Nat Neurosci 4:526–532.
Pollen DA, Gaska JP, Jacobson LD (1989) Physiological constraints on models of visual cortical function. In: Models of brain function (Cotterill RMJ, ed), pp 115–135. Cambridge, UK: Cambridge UP.
Pollen DA, Przybyszewski AW, Rubin MA, Foote W (2002) Spatial receptive field organization of macaque V4 neurons. Cereb Cortex 12:601–616.
Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) Numerical recepes in C. Cambridge, UK: Cambridge UP.
Priebe NJ, Mechler F, Carandini M, Ferster D (2004) The contribution of spike threshold to the dichotomy of cortical simple and complex cells. Nat Neurosci 7:1113–1122.
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, Bredfeldt CE, Shapley RM, Hawken MJ (2002) Suppression of neural responses to nonoptimal stimuli correlates with tuning selectivity in macaque V1. J Neurophysiol 87:1018–1027.
Rust NC, Schwartz O, Movshon JA, Simoncelli EP (2004) Spike-triggered characterization of excitatory and suppressive stimulus dimensions in monkey V1. Neurocomputing 58–60:793–799.
Rust NC, Schwartz O, Movshon JA, Simoncelli EP (2005) Spatiotemporal elements of macaque v1 receptive fields. Neuron 46:945–956.
Simoncelli E, Pillow J, Paninski L, Schwartz O (2004) Characterization of neural responses with stochastic stimuli. In: The new cognitive neuroscience (Gazzaniga M, ed), pp 327–338. Cambridge, MA: MIT.
Simoncelli EP, Heeger DJ (1998) A model of neuronal responses in visual area MT. Vision Res 38:743–761.
Skottun BC, De Valois RL, Grosof DH, Movshon JA, Albrecht DG, Bonds AB (1991) Classifying simple and complex cells on the basis of response modulation. Vision Res 31:1079–1086.
Tanaka K (1996) Inferotemporal cortex and object vision. Annu Rev Neurosci 19:109–139.
Tanaka K, Saito H, Fukada Y, Moriya M (1991) Coding visual images of objects in the inferotemporal cortex of the macaque monkey. J Neurophysiol 66:170–189.
Touryan J, Lau B, Dan Y (2002) Isolation of relevant visual features from random stimuli for cortical complex cells. J Neurosci 22:10811–10818.
Troyer TW, Krukowski AE, Priebe NJ, Miller KD (1998) Contrast-invariant orientation tuning in cat visual cortex: thalamocortical input tuning and correlation-based intracortical connectivity. J Neurosci 18:5908–5927.
Zhou YX, Baker Jr CL (1994) Envelope-responsive neurons in areas 17 and 18 of cat. J Neurophysiol 72:2134–2150.

Current Biology 21, 1641–1646, October 11, 2011 ©2011 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2011.08.031

Report

Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies

Shinji Nishimoto,[1] An T. Vu,[2] Thomas Naselaris,[1] Yuval Benjamini,[3] Bin Yu,[3] and Jack L. Gallant[1,2,4,*]
[1]Helen Wills Neuroscience Institute
[2]Joint Graduate Group in Bioengineering
[3]Department of Statistics
[4]Department of Psychology
University of California, Berkeley, Berkeley, CA 94720, USA

Summary

Quantitative modeling of human brain activity can provide crucial insights about cortical representations [1, 2] and can form the basis for brain decoding devices [3–5]. Recent functional magnetic resonance imaging (fMRI) studies have modeled brain activity elicited by static visual patterns and have reconstructed these patterns from brain activity [6–8]. However, blood oxygen level-dependent (BOLD) signals measured via fMRI are very slow [9], so it has been difficult to model brain activity elicited by dynamic stimuli such as natural movies. Here we present a new motion-energy [10, 11] encoding model that largely overcomes this limitation. The model describes fast visual information and slow hemodynamics by separate components. We recorded BOLD signals in occipitotemporal visual cortex of human subjects who watched natural movies and fit the model separately to individual voxels. Visualization of the fit models reveals how early visual areas represent the information in movies. To demonstrate the power of our approach, we also constructed a Bayesian decoder [8] by combining estimated encoding models with a sampled natural movie prior. The decoder provides remarkable reconstructions of the viewed movies. These results demonstrate that dynamic brain activity measured under naturalistic conditions can be decoded using current fMRI technology.

Results

Many of our visual experiences are dynamic: perception, visual imagery, dreaming, and hallucinations all change continuously over time, and these changes are often the most compelling and important aspects of these experiences. Obtaining a quantitative understanding of brain activity underlying these dynamic processes would advance our understanding of visual function. Quantitative models of dynamic mental events could also have important applications as tools for psychiatric diagnosis and as the foundation of brain machine interface devices [3–5].

Modeling dynamic brain activity is a difficult technical problem. The best tool available currently for noninvasive measurement of brain activity is functional magnetic resonance imaging (fMRI), which has relatively high spatial resolution [12, 13]. However, blood oxygen level-dependent (BOLD) signals measured using fMRI are relatively slow [9], especially when compared to the speed of natural vision and many other mental processes. It has therefore been assumed that fMRI data would not be useful for modeling brain activity evoked during natural vision or by other dynamic mental processes.

Here we present a new motion-energy [10, 11] encoding model that largely overcomes this limitation. The model separately describes the neural mechanisms mediating visual motion information and their coupling to much slower hemodynamic mechanisms. In this report, we first validate this encoding model by showing that it describes how spatial and temporal information are represented in voxels throughout visual cortex. We then use a Bayesian approach [8] to combine estimated encoding models with a sampled natural movie prior, in order to produce reconstructions of natural movies from BOLD signals.

We recorded BOLD signals from three human subjects while they viewed a series of color natural movies ($20^\circ \times 20^\circ$ at 15 Hz). A fixation task was used to control eye position. Two separate data sets were obtained from each subject. The training data set consisted of BOLD signals evoked by 7,200 s of color natural movies, where each movie was presented just once. These data were used to fit a separate encoding model for each voxel located in posterior and ventral occipitotemporal visual cortex. The test data set consisted of BOLD signals evoked by 540 s of color natural movies, where each movie was repeated ten times. These data were used to assess the accuracy of the encoding model and as the targets for movie reconstruction. Because the movies used to train and test models were different, this approach provides a fair and objective evaluation of the accuracy of the encoding and decoding models [2, 14].

BOLD signals recorded from each voxel were fit separately using a two-stage process. Natural movie stimuli were first filtered by a bank of neurally inspired nonlinear units sensitive to local motion-energy [10, 11]. L1-regularized linear regression [15, 16] was then used to fit a separate hemodynamic coupling term to each nonlinear filter (Figure 1; see also Supplemental Experimental Procedures available online). The regularized regression approach used here was optimized to obtain good estimates even for computational models containing thousands of regressors. In this respect, our approach differs from the regression procedures used in many other fMRI studies [17, 18].

To determine how much motion information is available in BOLD signals, we compared prediction accuracy for three different encoding models (Figures 2A–2C): a conventional static model that includes no motion information [8, 19], a nondirectional motion model that represents local motion energy but not direction, and a directional model that represents both local motion energy and direction. Each of these models was fit separately to every voxel recorded in each subject, and the test data were used to assess prediction accuracy for each model. Prediction accuracy was defined as the correlation between predicted and observed BOLD signals. The averaged accuracy across subjects and voxels in early visual areas (V1, V2, V3, V3A, and V3B) was 0.24, 0.39, and 0.40 for the static, nondirectional, and directional encoding models, respectively (Figures 2D and 2E; see Figure S1A for subject- and area-wise comparisons). This

*Correspondence: gallant@berkeley.edu

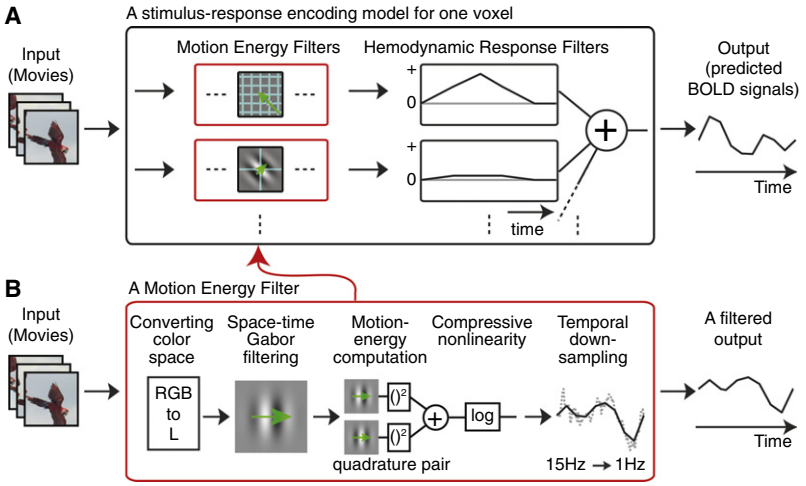


Figure 1. Schematic Diagram of the Motion-Energy Encoding Model

(A) Stimuli pass first through a fixed set of nonlinear spatiotemporal motion-energy filters (shown in detail in B) and then through a set of hemodynamic response filters fit separately to each voxel. The summed output of the filter bank provides a prediction of BOLD signals.

(B) The nonlinear motion-energy filter bank consists of several filtering stages. Stimuli are first transformed into the Commission Internationale de l'Éclairage L*A*B* color space, and the color channels are stripped off. Luminance signals then pass through a bank of 6,555 spatiotemporal Gabor filters differing in position, orientation, direction, spatial, and temporal frequency (see Supplemental Experimental Procedures for details). Motion energy is calculated by squaring and summing Gabor filters in quadrature. Finally, signals pass through a compressive nonlinearity and are temporally downsampled to the fMRI sampling rate (1 Hz).

difference in prediction accuracy was significant ($p < 0.0001$, Wilcoxon signed-rank test). An earlier study showed that the static model tested here recovered much more information from BOLD signals than had been obtained with any previous model [8, 19]. Nevertheless, both motion models developed here provide far more accurate predictions than are obtained with the static model. Note that the difference in prediction accuracy between the directional and nondirectional motion models, though significant, was small (Figure 2E; Figure S1A). This suggests that BOLD signals convey spatially localized but predominantly nondirectional motion information. These results show that the motion-energy encoding model predicts BOLD signals evoked by novel natural movies.

To further explore what information can be recovered from these data, we estimated the spatial, spatial frequency, and temporal frequency tuning of the directional motion-energy encoding model fit to each voxel. The spatial receptive fields of individual voxels were spatially localized (Figures 2F and 2G, left) and were organized retinotopically (Figures 2H and 2I), as reported in previous fMRI studies [12, 19–23]. Voxel-based receptive fields also showed spatial and temporal frequency tuning (Figures 2F and 2G, right), as reported in previous fMRI studies [24, 25].

To determine how motion information is represented in human visual cortex, we calculated the optimal speed for each voxel by dividing the peak temporal frequency by the peak spatial frequency. Projecting the optimal speed of the voxels onto a flattened map of the cortical surface (Figure 2J) revealed a significant positive correlation between eccentricity and optimal speed: relatively more peripheral voxels were tuned for relatively higher speeds. This pattern was observed in areas V1, V2, and V3 and for all three subjects ($p < 0.0001$, t test for correlation coefficient; see Figure S1B for subject- and area-wise comparisons). To our knowledge, this is the first evidence that speed selectivity in human early visual areas depends on eccentricity, though a consistent trend has been reported in human behavioral studies [26–28] and in neurophysiological studies of nonhuman primates [29, 30]. These results show that the motion-energy encoding model describes tuning for both spatial and temporal information at the level of single voxels.

To further characterize the temporal specificity of the estimated motion-energy encoding models, we used the test data to estimate movie identification accuracy. Identification accuracy [7, 19] measures how well a model can correctly associate an observed BOLD signal pattern with the specific stimulus that evoked it. Our motion-energy encoding model could identify the specific movie stimulus that evoked an observed BOLD signal 95% of the time (464 of 486 volumes) within ± one volume (1 s; subject S1; Figures 3A and 3B). This is far above what would be expected by chance (<1%). Identification accuracy (within ± one volume) was >75% for all three subjects even when the set of possible natural movie clips included 1,000,000 separate clips chosen at random from the internet (Figure 3C). This result demonstrates that the motion-energy encoding model is both valid and temporally specific. Furthermore, it suggests that the model might provide good reconstructions of natural movies from brain activity measurements [5].

We used a Bayesian approach [8] to reconstruct movies from the evoked BOLD signals (see also Figure S2). We estimated the posterior probability by combining a likelihood function (given by the estimated motion-energy model; see Supplemental Experimental Procedures) and a sampled natural movie prior. The sampled natural movie prior consists of ~18,000,000 s of natural movies sampled at random from the internet. These clips were assigned uniform prior probability (and consequently all other clips were assigned zero prior probability; note also that none of the clips in the prior were used in the experiment). Furthermore, to make decoding tractable, reconstructions were based on 1 s clips (15 frames), using BOLD signals with a delay of 4 s. In effect, this procedure enforces an assumption that the spatiotemporal stimulus that elicited each measured BOLD signal must be one of the movie clips in the sampled prior.

Figure 4 shows typical reconstructions of natural movies obtained using the motion-energy encoding model and the Bayesian decoding approach (see Movie S1 for the corresponding movies). The posterior probability was estimated across the entire sampled natural movie prior separately for each BOLD signal in the test data. The peak of this posterior distribution was the conventional maximum a posteriori (MAP) reconstruction [8] for each BOLD signal (see second row in Figure 4). When the sampled natural movie prior contained clips similar to the viewed clip, the MAP reconstructions were good (e.g., the close-up of a human speaker shown in Figure 4A). However, when the prior contained no clips similar to the viewed clip, the reconstructions are poor (e.g., Figure 4B). This likely reflects both the limited size of the sampled natural movie prior and noise in the fMRI measurements. One way to

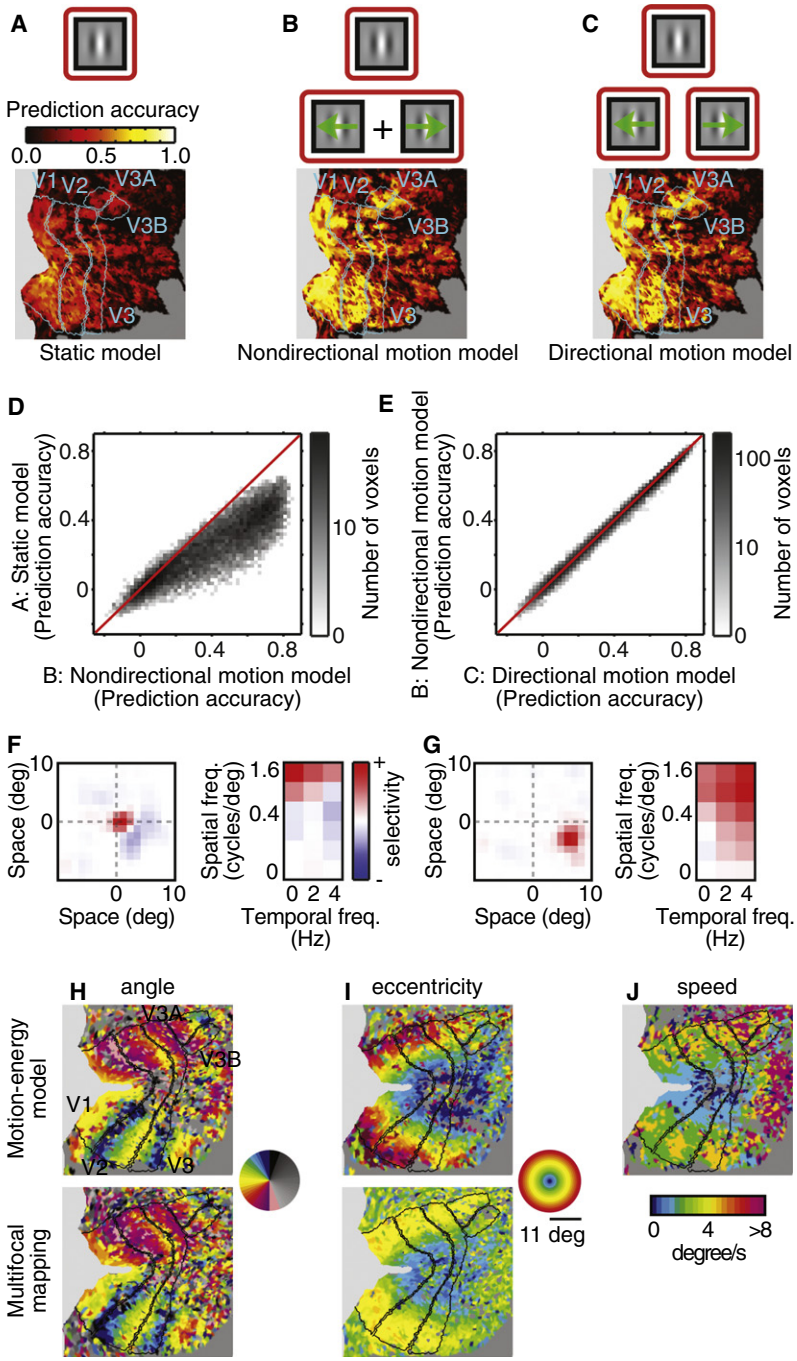


Figure 2. The Directional Motion-Energy Model Captures Motion Information

(A) Top: the static encoding model includes only Gabor filters that are not sensitive to motion. Bottom: prediction accuracy of the static model is shown on a flattened map of the cortical surface of one subject (S1). Prediction accuracy is relatively poor.

(B) The nondirectional motion-energy encoding model includes Gabor filters tuned to a range of temporal frequencies, but motion in opponent directions is pooled. Prediction accuracy of this model is better than the static model.

(C) The directional motion-energy encoding model includes Gabor filters tuned to a range of temporal frequencies and directions. This model provides the most accurate predictions of all models tested.

(D and E) Voxel-wise comparisons of prediction accuracy between the three models. The directional motion-energy model performs significantly better than the other two models, although the difference between the nondirectional and directional motion models is small. See also Figure S1 for subject- and area-wise comparisons.

(F) The spatial receptive field of one voxel (left) and its spatial and temporal frequency selectivity (right). This receptive field is located near the fovea, and it is high-pass for spatial frequency and low-pass for temporal frequency. This voxel thus prefers static or low-speed motion.

(G) Receptive field for a second voxel. This receptive field is located lower periphery, and it is band-pass for spatial frequency and high-pass for temporal frequency. This voxel thus prefers higher-speed motion than the voxel in (F).

(H) Comparison of retinotopic angle maps estimated using the motion-energy encoding model (top) and conventional multifocal mapping (bottom) on a flattened cortical map [47]. The angle maps are similar, even though they were estimated using independent data sets and methods.

(I) Comparison of eccentricity maps estimated as in (H). The maps are similar except in the far periphery, where the multifocal mapping stimulus was coarse.

(J) Optimal speed projected on to a flattened map as in (H). Voxels near the fovea tend to prefer slow-speed motion, whereas those in the periphery tend to prefer high-speed motion. See also Figure S1B for subject-wise comparisons.

achieve more robust reconstructions without enlarging the prior is to interpolate over the sparse samples in the prior. We therefore created an averaged high posterior (AHP) reconstruction by averaging the 100 clips in the sampled natural movie prior that had the highest posterior probability (see also Figure S2; note that the AHP reconstruction can be viewed as a Bayesian version of bagging [31]). The AHP reconstruction captures the spatiotemporal structure within a viewed clip even when it is completely unique (e.g., the spreading of an inkblot from the center of the visual field shown in Figure 4B).

To quantify reconstruction quality, we calculated the correlation between the motion-energy content of the original movies and their reconstructions (see Supplemental Experimental Procedures). A correlation of 1.0 indicates perfect reconstruction of the spatiotemporal energy in the original movies, and a correlation of 0.0 indicates that the movies and their reconstruction are spatiotemporally uncorrelated. The results for both MAP and AHP reconstructions are shown in Figure 4D. In both cases, reconstruction accuracy was significantly higher than chance ($p < 0.0001$, Wilcoxon rank-sum test; see Supplemental Experimental Procedures). Furthermore, AHP reconstructions were significantly better than MAP reconstructions ($p < 0.0001$, Wilcoxon signed-rank test). Although still crude (motion-energy correlation ~ 0.3), these results validate our general approach to reconstruction and demonstrate that the AHP estimate improves reconstruction over the MAP estimate.

Discussion

In this study, we developed an encoding model that predicts BOLD signals in early visual areas with unprecedented accuracy. By using this model in a Bayesian framework, we

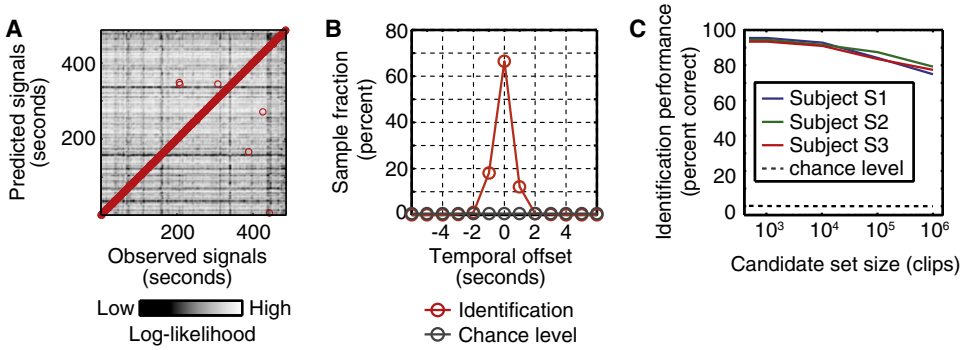


Figure 3. Identification Analysis

(A) Identification accuracy for one subject (S1). The test data in our experiment consisted of 486 volumes (s) of BOLD signals evoked by the test movies. The estimated model yielded 486 volumes of BOLD signals predicted for the same movies. The brightness of the point in the *m*th column and *n*th row represents the log-likelihood (see Supplemental Experimental Procedures) of the BOLD signals evoked at the *m*th second given the BOLD signal predicted at the *n*th second. The highest log-likelihood in each column is designated by a red circle and thus indicates the choice of the identification algorithm.

(B) Temporal offset between the correct timing and the timing identified by the algorithm for the same subject shown in (A). The algorithm was correct to within ± one volume (s) 95% of the time (464 of 486 volumes); chance performance is <1% (3 of 486 volumes; i.e., three volumes centered at the correct timing).

(C) Scaling of identification accuracy with set size. To understand how identification accuracy scales with size of stimulus set, we enlarged the identification stimulus set to include additional stimuli drawn from a natural movie database (which was not actually used in the experiment). For all three subjects, identification accuracy (within ± one volume) was >75% even when the set of potential movies included 1,000,000 clips. This is far above chance (gray dashed line).

provide the first reconstructions of natural movies from human brain activity. This is a critical step toward the creation of brain reading devices that can reconstruct dynamic perceptual experiences. Our solution to this problem rests on two key innovations. The first is a new motion-energy encoding model that is optimized for use with fMRI and that aims to reflect the separate contributions of the underlying neuronal population and hemodynamic coupling (Figure 1). This encoding model recovers fine temporal information from relatively slow BOLD signals. The second is a sampled natural movie prior that is embedded within a Bayesian decoding framework. This approach provides a simple method for reconstructing spatiotemporal stimuli from the sparsely sampled and slow BOLD signals.

Our results provide the first evidence that there is a positive correlation between eccentricity and optimal speed in human early visual areas. This provides a functional explanation for previous behavioral studies indicating that speed sensitivity

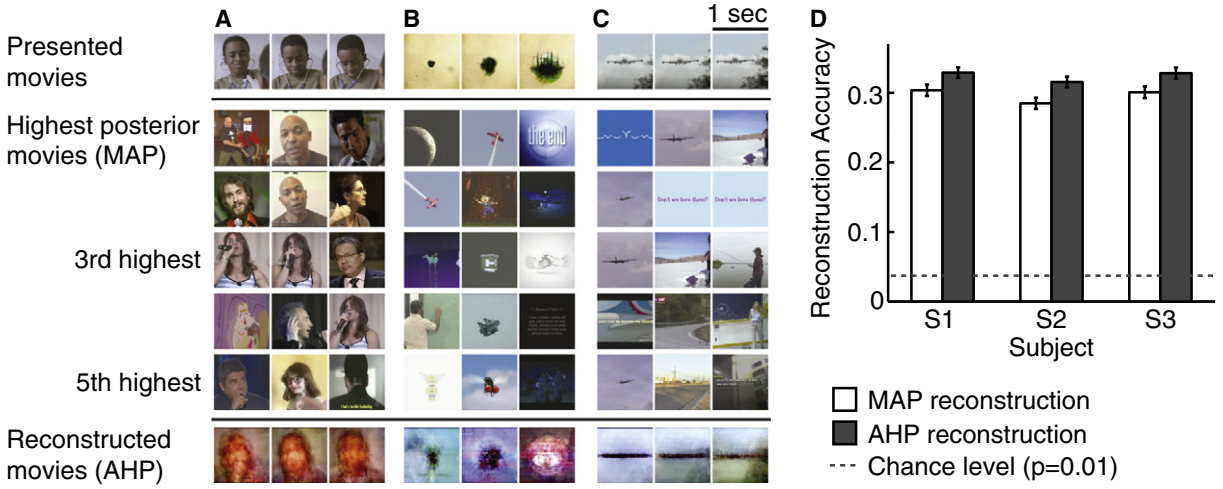


Figure 4. Reconstructions of Natural Movies from BOLD Signals

(A) The first (top) row shows three frames from a natural movie used in the experiment, taken 1 s apart. The second through sixth rows show frames from the five clips with the highest posterior probability. The maximum a posteriori (MAP) reconstruction is shown in the second row. The seventh (bottom) row shows the averaged high posterior (AHP) reconstruction. The MAP provides a good reconstruction of the second and third frames, whereas the AHP provides more robust reconstructions across frames.

(B and C) Additional examples of reconstructions, in the same format as (A).

(D) Reconstruction accuracy (correlation in motion-energy; see Supplemental Experimental Procedures) for all three subjects. Error bars indicate ±1 standard error of the mean across 1 s clips. Both the MAP and AHP reconstructions are significant, though the AHP reconstructions are significantly better than the MAP reconstructions. Dashed lines show chance performance (p = 0.01). See also Figure S2.

depends on eccentricity [26–28]. This systematic variation in optimal speed across the visual field may be an adaptation to the nonuniform distribution of speed signals induced by selective foveation in natural scenes [32]. From the perspective of decoding, this result suggests that we might further optimize reconstruction by including eccentricity-dependent speed tuning in the prior.

We found that a motion-energy model that incorporates directional motion signals was only slightly better than a model that does not include direction. We believe that this likely reflects limitations in the spatial resolution of fMRI recordings. Indeed, a recent study reported that hemodynamic signals were sufficient to visualize a columnar organization of motion direction in macaque area V2 [33]. Future fMRI experiments at higher spatial or temporal resolution [34, 35] might therefore be able to recover clearer directional signals in human visual cortex.

In preliminary work for this study, we explored several encoding models that incorporated color information explicitly. However, we found that color information did not improve the accuracy of predictions or identification beyond what could be achieved with models that include only luminance information. We believe that this reflects the fact that luminance and color borders are often correlated in natural scenes ([36, 37], but see [38]). (Note that when isoluminant, monochromatic stimuli are used, color can be reconstructed from evoked BOLD signals [39].) The correlation between luminance and color information in natural scenes has an interesting side effect: our reconstructions tended to recover color borders (e.g., borders between hair versus face or face versus body), even though the encoding model makes no use of color information. This is a positive aspect of the sampled natural movie prior and provides additional cues to aid in recognition of reconstructed scenes (see also [40]).

We found that the quality of reconstruction could be improved by simply averaging around the maximum of the posterior movies. This suggests that reconstructions might be further improved if the number of samples in the prior is much larger than the one used here. Likelihood estimation (and thus reconstruction) would also improve if additional knowledge about the neural representation of movies was used to construct better encoding models (e.g., [41]).

In a landmark study, Thirion et al. [6] first reconstructed static imaginary patterns from BOLD signals in early visual areas. Other studies have decoded subjective mental states, such as the contents of visual working memory [42], or whether subjects are attending to one or another orientation or direction [3, 43]. The modeling framework presented here provides the first reconstructions of dynamic perceptual experiences from BOLD signals. Therefore, this modeling framework might also permit reconstruction of dynamic mental content such as continuous natural visual imagery. In contrast to earlier studies that reconstruct visual patterns defined by checkerboard contrast [6, 7], our framework could potentially be used to decode involuntary subjective mental states (e.g., dreaming or hallucination), though it would be difficult to determine whether the decoded content was accurate. One recent study showed that BOLD signals elicited by visual imagery are more prominent in ventral-temporal visual areas than in early visual areas [44]. This finding suggests that a hybrid encoding model that combines the structural motion-energy model developed here with a semantic model of the form developed in previous studies [8, 45, 46] could provide even better reconstructions of subjective mental experiences.

Experimental Procedures

Stimuli

Visual stimuli consisted of color natural movies drawn from the Apple QuickTime HD gallery (http://trailers.apple.com/) and YouTube (http://www.youtube.com/; see the list of movies in Supplemental Experimental Procedures). The original high-definition movies were cropped to a square and then spatially downsampled to 512 × 512 pixels. Movies were then clipped to 10–20 s in length, and the stimulus sequence was created by randomly drawing movies from the entire set. Movies were displayed using a VisuaStim LCD goggle system (20° × 20° at 15 Hz). A colored fixation spot (4 pixels or 0.16° square) was presented on top of the movie. The color of the fixation spot changed three times per second to ensure that it was visible regardless of the color of the movie.

MRI Parameters

The experimental protocol was approved by the Committee for the Protection of Human Subjects at University of California, Berkeley. Functional scans were conducted using a 4 Tesla Varian INOVA scanner (Varian, Inc.) with a quadrature transmit/receive surface coil (Midwest RF). Scans were obtained using T2*-weighted gradient-echo EPI: TR = 1 s, TE = 28 ms, flip angle = 56°, voxel size = 2.0 × 2.0 × 2.5 mm^3, FOV = 128 × 128 mm^2. The slice prescription consisted of 18 coronal slices beginning at the posterior pole and covering the posterior portion of occipital cortex.

Data Collection

Functional MRI scans were made from three human subjects, S1 (author S.N., age 30), S2 (author T.N., age 34), and S3 (author A.T.V., age 23). All subjects were healthy and had normal or corrected-to-normal vision. The training data were collected in 12 separate 10 min blocks (7,200 s total). The training movies were shown only once each. The test data were collected in nine separate 10 min blocks (5,400 s total) consisting of 9 min movies repeated ten times each. To minimize effects from potential adaptation and long-term drift in the test data, we divided the 9 min movies into 1 min chunks, and these were randomly permuted across blocks. Each test block was thus constructed by concatenating ten separate 1 min movies. All data were collected across multiple sessions for each subject, and each session contained multiple training and test blocks. The training and test data sets used different movies.

Additional methods can be found in Supplemental Experimental Procedures.

Supplemental Information

Supplemental Information includes two figures, Supplemental Experimental Procedures, and one movie and can be found with this article online at doi:10.1016/j.cub.2011.08.031.

Acknowledgments

We thank B. Inglis for assistance with MRI and K. Kay and K. Hansen for assistance with retinotopic mapping. We also thank M. Oliver, R. Prenger, D. Stansbury, A. Huth, and J. Gao for their assistance with various aspects of this research. This work was supported by the National Institutes of Health and the National Eye Institute.

Received: May 3, 2011
Revised: July 23, 2011
Accepted: August 15, 2011
Published online: September 22, 2011

References

1. Wu, M.C., David, S.V., and Gallant, J.L. (2006). Complete functional characterization of sensory neurons by system identification. Annu. Rev. Neurosci. *29*, 477–505.
2. Naselaris, T., Kay, K.N., Nishimoto, S., and Gallant, J.L. (2011). Encoding and decoding in fMRI. Neuroimage *56*, 400–410.
3. Kamitani, Y., and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. Nat. Neurosci. *8*, 679–685.
4. Haynes, J.D., and Rees, G. (2006). Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. *7*, 523–534.
5. Kay, K.N., and Gallant, J.L. (2009). I can see what you see. Nat. Neurosci. *12*, 245.

6. Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J.B., Lebihan, D., and Dehaene, S. (2006). Inverse retinotopy: inferring the visual content of images from brain activation patterns. Neuroimage *33*, 1104–1116.
7. Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M.A., Morito, Y., Tanabe, H.C., Sadato, N., and Kamitani, Y. (2008). Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron *60*, 915–929.
8. Naselaris, T., Prenger, R.J., Kay, K.N., Oliver, M., and Gallant, J.L. (2009). Bayesian reconstruction of natural images from human brain activity. Neuron *63*, 902–915.
9. Friston, K.J., Jezzard, P., and Turner, R. (1994). Analysis of functional MRI time-series. Hum. Brain Mapp. *1*, 153–171.
10. Adelson, E.H., and Bergen, J.R. (1985). Spatiotemporal energy models for the perception of motion. J. Opt. Soc. Am. A *2*, 284–299.
11. Watson, A.B., and Ahumada, A.J., Jr. (1985). Model of human visual-motion sensing. J. Opt. Soc. Am. A *2*, 322–341.
12. Engel, S.A., Rumelhart, D.E., Wandell, B.A., Lee, A.T., Glover, G.H., Chichilnisky, E.J., and Shadlen, M.N. (1994). fMRI of human visual cortex. Nature *369*, 525.
13. Logothetis, N.K. (2008). What we can do and what we cannot do with fMRI. Nature *453*, 869–878.
14. Kriegeskorte, N., Simmons, W.K., Bellgowan, P.S., and Baker, C.I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. Nat. Neurosci. *12*, 535–540.
15. Li, Y., and Osher, S. (2009). Coordinate descent optimization for l1 minimization with application to compressed sensing; a greedy algorithm. Inverse Probl. Imaging *3*, 487–503.
16. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. J. R. Stat. Soc. B *58*, 267–288.
17. Friston, K.J., Frith, C.D., Turner, R., and Frackowiak, R.S. (1995). Characterizing evoked hemodynamics with fMRI. Neuroimage *2*, 157–165.
18. Boynton, G.M., Engel, S.A., Glover, G.H., and Heeger, D.J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. J. Neurosci. *16*, 4207–4221.
19. Kay, K.N., Naselaris, T., Prenger, R.J., and Gallant, J.L. (2008). Identifying natural images from human brain activity. Nature *452*, 352–355.
20. Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., and Tootell, R.B.H. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science *268*, 889–893.
21. DeYoe, E.A., Carman, G.J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., Miller, D., and Neitz, J. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. Proc. Natl. Acad. Sci. USA *93*, 2382–2386.
22. Wandell, B.A., Dumoulin, S.O., and Brewer, A.A. (2007). Visual field maps in human cortex. Neuron *56*, 366–383.
23. Dumoulin, S.O., and Wandell, B.A. (2008). Population receptive field estimates in human visual cortex. Neuroimage *39*, 647–660.
24. Singh, K.D., Smith, A.T., and Greenlee, M.W. (2000). Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. Neuroimage *12*, 550–564.
25. Henriksson, L., Nurminen, L., Hyvärinen, A., and Vanni, S. (2008). Spatial frequency tuning in human retinotopic visual areas. J. Vis. *8*, 5.1–13.
26. Kelly, D.H. (1984). Retinal inhomogeneity. I. Spatiotemporal contrast sensitivity. J. Opt. Soc. Am. A *1*, 107–113.
27. McKee, S.P., and Nakayama, K. (1984). The detection of motion in the peripheral visual field. Vision Res. *24*, 25–32.
28. Orban, G.A., Van Calenbergh, F., De Bruyn, B., and Maes, H. (1985). Velocity discrimination in central and peripheral visual field. J. Opt. Soc. Am. A *2*, 1836–1847.
29. Orban, G.A., Kennedy, H., and Bullier, J. (1986). Velocity sensitivity and direction selectivity of neurons in areas V1 and V2 of the monkey: influence of eccentricity. J. Neurophysiol. *56*, 462–480.
30. Yu, H.H., Verma, R., Yang, Y., Tibballs, H.A., Lui, L.L., Reser, D.H., and Rosa, M.G. (2010). Spatial and temporal frequency tuning in striate cortex: functional uniformity and specializations related to receptive field eccentricity. Eur. J. Neurosci. *31*, 1043–1062.
31. Domingos, P. (1997). Why does bagging work? A Bayesian account and its implications. In Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, D. Heckerman, H. Mannila, D. Pregibon, and R. Uthurusamy, eds., pp. 155–158.
32. Eckert, M.P., and Buchsbaum, G. (1993). Efficient coding of natural time varying images in the early visual system. Philos. Trans. R. Soc. Lond. B Biol. Sci. *339*, 385–395.
33. Lu, H.D., Chen, G., Tanigawa, H., and Roe, A.W. (2010). A motion direction map in macaque V2. Neuron *68*, 1002–1013.
34. Moeller, S., Yacoub, E., Olman, C.A., Auerbach, E., Strupp, J., Harel, N., and Uğurbil, K. (2010). Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. Magn. Reson. Med. *63*, 1144–1153.
35. Feinberg, D.A., Moeller, S., Smith, S.M., Auerbach, E., Ramanna, S., Glasser, M.F., Miller, K.L., Ugurbil, K., and Yacoub, E. (2010). Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. PLoS ONE *5*, e15710.
36. Fine, I., MacLeod, D.I., and Boynton, G.M. (2003). Surface segmentation based on the luminance and color statistics of natural scenes. J. Opt. Soc. Am. A Opt. Image Sci. Vis. *20*, 1283–1291.
37. Zhou, C., and Mel, B.W. (2008). Cue combination and color edge detection in natural scenes. J. Vis. *8*, 4.1–25.
38. Hansen, T., and Gegenfurtner, K.R. (2009). Independence of color and luminance edges in natural scenes. Vis. Neurosci. *26*, 35–49.
39. Brouwer, G.J., and Heeger, D.J. (2009). Decoding and reconstructing color from responses in human visual cortex. J. Neurosci. *29*, 13992–14003.
40. Oliva, A., and Schyns, P.G. (2000). Diagnostic colors mediate scene recognition. Cognit. Psychol. *41*, 176–210.
41. Bartels, A., Zeki, S., and Logothetis, N.K. (2008). Natural vision reveals regional specialization to local motion and to contrast-invariant, global flow in the human brain. Cereb. Cortex *18*, 705–717.
42. Harrison, S.A., and Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. Nature *458*, 632–635.
43. Kamitani, Y., and Tong, F. (2006). Decoding seen and attended motion directions from activity in the human visual cortex. Curr. Biol. *16*, 1096–1102.
44. Reddy, L., Tsuchiya, N., and Serre, T. (2010). Reading the mind's eye: decoding category information during mental imagery. Neuroimage *50*, 818–825.
45. Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K.M., Malave, V.L., Mason, R.A., and Just, M.A. (2008). Predicting human brain activity associated with the meanings of nouns. Science *320*, 1191–1195.
46. Li, L., Socher, R., and Li, F. (2009). Towards total scene understanding: Classification, annotation and segmentation in an automatic framework. In IEEE Computer Science Conference on Computer Vision and Pattern Recognition, pp. 2036–2043.
47. Hansen, K.A., David, S.V., and Gallant, J.L. (2004). Parametric reverse correlation reveals spatial linearity of retinotopic human V1 BOLD response. Neuroimage *23*, 233–241.

Behavioral/Systems/Cognitive

A Three-Dimensional Spatiotemporal Receptive Field Model Explains Responses of Area MT Neurons to Naturalistic Movies

Shinji Nishimoto[1] **and Jack L. Gallant**[1,2]

[1]Helen Wills Neuroscience Institute and [2]Department of Psychology, University of California, Berkeley, California 94720

Area MT has been an important target for studies of motion processing. However, previous neurophysiological studies of MT have used simple stimuli that do not contain many of the motion signals that occur during natural vision. In this study we sought to determine whether views of area MT neurons developed using simple stimuli can account for MT responses under more naturalistic conditions. We recorded responses from macaque area MT neurons during stimulation with naturalistic movies. We then used a quantitative modeling framework to discover which specific mechanisms best predict neuronal responses under these challenging conditions. We find that the simplest model that accurately predicts responses of MT neurons consists of a bank of V1-like filters, each followed by a compressive nonlinearity, a divisive nonlinearity, and linear pooling. Inspection of the fit models shows that the excitatory receptive fields of MT neurons tend to lie on a single plane within the three-dimensional spatiotemporal frequency domain, and suppressive receptive fields lie off this plane. However, most excitatory receptive fields form a partial ring in the plane and avoid low temporal frequencies. This receptive field organization ensures that most MT neurons are tuned for velocity but do not tend to respond to ambiguous static textures that are aligned with the direction of motion. In sum, MT responses to naturalistic movies are largely consistent with predictions based on simple stimuli. However, models fit using naturalistic stimuli reveal several novel properties of MT receptive fields that had not been shown in prior experiments.

Introduction

Area MT is an important site of motion processing that lies downstream from areas V1 and V2 (Felleman and Van Essen, 1991; Born and Bradley, 2005). Many studies have examined how MT neurons represent motion information, using synthetic stimuli such as bars (Albright, 1984; Okamoto et al., 1999), gratings (Movshon et al., 1985; Pack and Born, 2001; Perrone and Thiele, 2001), dots (Britten et al., 1993), and noise (Livingstone et al., 2001). Several influential models have been proposed to account for these neurophysiological findings (Simoncelli and Heeger, 1998; Rust et al., 2006; Bradley and Goyal, 2008).

The ultimate goal of visual neuroscience is to understand the neural mechanisms mediating normal vision. For this reason, it is generally agreed that models of visual processing should ultimately predict responses observed during natural vision (Rust and Movshon, 2005; Wu et al., 2006; Stanley, 2008). Do the neuronal models of area MT developed from experiments that used synthetic stimuli predict responses under more natural viewing conditions? The answer to this question is not known, because no neurophysiology study has yet reported data that reflect the full range of stimulus–response relationships that can occur during natural vision. Natural moving stimuli occupy a three-dimensional spatiotemporal frequency domain: two dimensions of space and one of time. Previous neurophysiological studies of MT have only focused on a subspace within the three-dimensional frequency domain: a one-dimensional ring (i.e., direction) (Movshon et al., 1985; Pack and Born, 2001; Smith et al., 2005; Rust et al., 2006; Majaj et al., 2007), a two-dimensional slice (Perrone and Thiele, 2001; Priebe et al., 2003), or a cylinder (Okamoto et al., 1999). When a model is constructed based on data in a restricted stimulus subspace, generalizing the model to naturalistic stimuli is an ill-posed problem that will inevitably involve untested assumptions.

Recent studies raise another concern: the way that neurons represent visual information might be different if measured using synthetic (e.g., white noise or grating) versus more naturalistic stimuli. Several groups have addressed this issue in area V1 (David and Gallant, 2005; Felsen et al., 2005; Sharpee et al., 2006). These studies found that while models developed using synthetic stimuli generally explained responses evoked by naturalistic stimuli, receptive fields observed using synthetic stimuli deviated systematically from those observed using naturalistic stimuli. These deviations suggest that neurons possess nonlinear mechanisms that depend on stimulus statistics. Given the stimulus-dependent deviations found in V1, it is possible that some

Received Dec. 30, 2010; revised Aug. 10, 2011; accepted Aug. 13, 2011.

Author contributions: S.N. and J.L.G. designed research; S.N. performed research; S.N. contributed unpublished reagents/analytic tools; S.N. analyzed data; S.N. and J.L.G. wrote the paper.

This work was supported by grants from the National Eye Institute and the National Institute of Mental Health (J.L.G.). James Mazer wrote the neurophysiology software suite, and Stephen David wrote the database software. We thank Kendrick Kay, Thomas Naselaris, An Vu, and Liberty Hamilton for comments on this manuscript. We also thank Michael Oliver, Ryan Prenger, Michael Wu, and Ben Willmore for their help and for fruitful discussions.

The authors declare no competing financial interests.

Correspondence should be addressed to Jack L. Gallant, University of California at Berkeley, 3210 Tolman Hall, #1650, Berkeley, CA 94720. E-mail: gallant@berkeley.edu.

DOI:10.1523/JNEUROSCI.6801-10.2011

 0270-6474/11/3114551-14$15.00/0

properties of MT neurons might differ depending on whether they are measured using synthetic stimuli versus under more naturalistic conditions.

Here we addressed this issue by using naturalistic motion-enhanced movies to characterize receptive properties of macaque area MT neurons. These movies allowed us to probe the full three-dimensional frequency domain within the time constraints of neurophysiological experiments. We used a quantitative modeling approach to characterize responses of single MT neurons to these movies. We compared recovered receptive fields with theoretical predictions and with the results of previous studies that used synthetic stimuli.

Materials and Methods

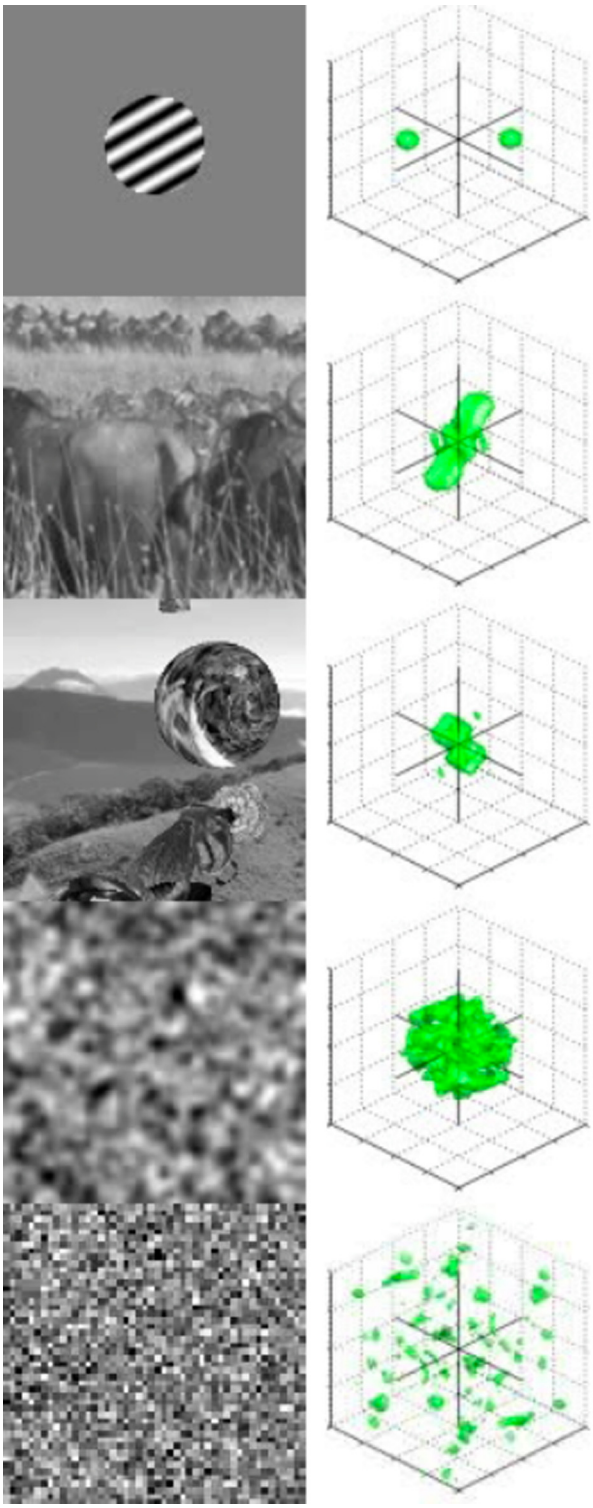
Physiology and behavioral tasks. Extracellular single-unit recordings were made from two adult male macaques (*Macaca mulatta*), prepared for recording as described previously (Mazer and Gallant, 2003). Recordings were made with epoxy-coated tungsten electrodes (FHC). Signals were amplified, bandpass filtered, and sorted (Plexon Instruments) to identify single units. Area MT was located by exterior cranial landmarks, anatomical images from magnetic resonance imaging (MRI), and/or physiological properties. During recordings, subjects performed a fixation task for liquid reward. Eye movements were monitored by an infrared eye tracker at either 250 or 500 Hz (EyeLink II; SR Research). All animal procedures were approved by the Animal Care and Use Committees at the University of California, Berkeley, and met or exceeded all NIH and U.S. Department of Agriculture regulations.

Visual stimuli. The primary stimuli consisted of motion-enhanced natural movies (see Fig. 1; Movie 1) constructed by combining full-screen natural movies (background) with an overlay of textured, moving three-dimensional objects (foreground). The movies were obtained from high-definition natural movie libraries provided by the Cornell Laboratory of Ornithology or the BBC. The moving objects consisted of cubes, spheres, and animal shapes, synthesized using a three-dimensional rendering library (Panda3D by Disney and Carnegie Melon University). The object textures were static natural images obtained from the McGill Calibrated Color Image Database (Olmos and Kingdom, 2004). The objects moved around a virtual three-dimensional space, and their accelerations for spatial position and rotation angles were updated by random walk. Movies were converted to grayscale by taking the average luminance across the three color channels. After the movies were constructed and concatenated into a single sequence, simulated saccadic eye movements were introduced by cutting the movies into sequences of 350 ± 50 ms and shuffling the order of the segments (David et al., 2004). (Note that this scheme provides a synchronous scene cut of both foreground and background.)

To ensure that the motion-enhanced natural movies did not produce biased receptive field estimates, some recordings were made with natural movies that contained no motion enhancement. These movies were constructed exactly as described above, but without the addition of rendered foreground objects.

The display device was a Sony Trinitron CRT monitor with a refresh rate of 83 Hz and a display resolution of 640×480 pixels (36×27 degrees of visual field at the viewing distance of 57 cm). Stimuli were presented while subjects fixated on a small spot ($<0.1°$) for 3–5 s per trial (1° diameter fixation window). After each successful fixation, there was a 200 ms delay (neutral gray background, 60 cd/m^2) followed in turn by the movie. The first 220 ms of the movie shown on each trial overlapped with the final frames of the movies shown on the previous trial. This permitted us to reduce the effects of the initial transient response during receptive field estimation by removing the associated responses before analysis (David et al., 2004).

Recordings were made from 52 single area MT neurons while showing a total of 20,000–40,000 frames of motion-enhanced natural movies (mean number of frames, 27,120). These training data were used to estimate the spectral receptive field for each neuron. Each neuron was also probed with a different motion-enhanced natural movie, 2000 frames in length, and repeated 5–20 times. These validation data were



Movie 1. Comparison of the three-dimensional amplitude spectra for several different stimulus classes. The left column shows several different classes of stimuli that have been used in visual neurophysiology experiments. From the top to the bottom, these are grating sequences, natural movies, motion-enhanced natural movies, pink noise, and white noise. The right column shows the three-dimensional spectral amplitudes of these stimuli. The green blobs delineate the isosurfaces for the spectral amplitudes. There are clear differences in the amplitude spectra of these stimuli. The amplitude spectrum of gratings is sparse, while the spectrum of white noise is dense. Natural movies and motion-enhanced natural movies have very similar 1/*f* amplitude spectra, but because the motion-enhanced natural movies are biased toward higher temporal frequencies, they span the frequency space more efficiently. The amplitude spectrum of pink noise is also 1/*f*, but in other respects the spectrum of pink noise is quite different from that of natural movies.

used to evaluate prediction accuracy. The number of spikes obtained from a single neuron in the training data was 8413 on average, and the number of spikes in the validation data was 6108 on average. To estimate a single response time course from the repeated validation data, we first averaged the responses across repeats and then applied a 12 ms temporal Gaussian filter.

As a control, on a subset of 15 MT neurons we collected additional validation data using 2000 frames of natural movies without motion enhancement. Each movie was repeated 5–20 times. The number of spikes in the natural movie data set was 5266.

Notes on stimulus design. The stimuli used in this study were designed with three important constraints in mind. First, because motion information in natural stimuli can only be defined by three-dimensional changes of luminance patterns, the stimulus set should cover the full three-dimensional frequency domain (two for space, one for time). Second, because neurophysiological recordings from single neurons tend to be data limited, the stimuli should facilitate efficient estimation of receptive field properties. Third, because stimulus statistics might affect receptive field properties, the stimuli should be naturalistic. Note that "natural stimuli" do not belong to a discrete category. Rather, naturalness is a continuum. A very natural stimulus would possess a natural dynamic luminance range, natural color distribution, binocular disparity, and so on. It would also reflect the influence of the observer's eye movements and motion through the environment. A very unnatural stimulus would be white noise or gratings. The stimuli used in any neurophysiological experiment lies somewhere on this continuum.

The motion-enhanced natural movies used here satisfy all three of these constraints. They span the full three-dimensional frequency domain. They contain naturalistic texture and naturally structured motion (i.e., rotation, expansion, contraction, and translations). They tend to reduce the spatial and temporal correlations found in natural movies, thereby making it easier to correct estimated spectral receptive fields for stimulus bias.

Movie 1 shows the three-dimensional spectrum of the drifting gratings, natural stimuli, motion-enhanced movies, $1/f$ noise, and white noise. Drifting gratings are very unnatural and do not sample the three-dimensional frequency domain efficiently. Natural movies have natural three-dimensional spectrum, but they do not contain much motion information and so are likely to be an inefficient stimulus for characterizing area MT neurons. Motion-enhanced movies retain the second-order $1/f$ amplitude spectrum characteristic of natural movies, but contain substantially more motion information. White noise has a very different spectrum from natural movies, and $1/f$ noise has no statistical structure beyond second order.

Model estimation. It is difficult in principle to model the nonlinear relationship between stimulus and response in visual neurons (Wu et al., 2006). Estimation of receptive field properties for these nonlinear neurons requires an optimization method that can search through a complex error surface without becoming stuck in a local minimum. One way to solve this problem is to nonlinearly transform the stimulus into a new space in which the relationship between the transformed stimulus and the response is linear. In this case, linear optimization methods can be used to find the optimal weights that map between the nonlinearly transformed stimulus and measured responses. In this study we used a V1 filter bank (see below, The V1 filter bank and Tests for nonlinearities) to perform the nonlinear transformation, and we used regularized linear regression to find the optimal weights (see below, Regression by boosting). Note that a similar approach (i.e., using linear combinations of nonlinear local spectral measurements) was used in previous studies to characterize receptive fields of neurons in early visual (Nishimoto et al., 2006) and auditory areas (Theunissen et al., 2000).

The V1 filter bank. The bank of V1 filters chosen to represent each MT neuron were selected from a Gabor basis. The Gabor basis consisted of several thousand individual Gabor filters (see below), each defined as follows:

$$G_{i,p}(x,y,t) = \exp\left(-\frac{(x-cx_i)^2+(y-cy_i)^2}{2ws_i^2}-\frac{(t-ct_i)^2}{2wt_i^2}\right) * \sin((x-cx_i)*fx_i+(y-cy_i)*fy_i+(t-ct_i)ft_i+p), \quad (1)$$

where fx_i, fy_i, and ft_i represent the spatial and temporal frequency; cx_i, cy_i, and ct_i give the center of each Gabor filter in each dimension of the space-time domain; ws_i and wt_i give the width of the Gaussian envelope in space and time; and p gives phase.

The process of filtering each movie $I(x,y,t)$ with these Gabor filters was modeled as linear multiplication:

$$L_{i,p}(t) = \sum_x \sum_y \sum_\tau G_{i,p}(x,y,\tau) I(x, y, t-\tau). \quad (2)$$

The V1 simple cell inputs were modeled as follows:

$$S_{i,p}(t) = HR[L_{i,p}(t)], \quad (3)$$

where $HR[*]$ is half-wave rectification, and $p = 0°$, 90°, 180°, and 270°.

The V1 complex cell inputs were modeled as follows:

$$C_i(t) = \sqrt{L_{i,0}^2 + L_{i,90}^2}. \quad (4)$$

Note that $S(t)$ and $C(t)$ are time series. In this study we call these (and their variants, described below) V1 filter outputs, and denote them as $X(t)$. A previous neurophysiological study showed that V1 afferents to area MT are predominantly direction-selective complex cells, not simple cells (Movshon and Newsome, 1996). Our modeling results confirm this finding: most of the response variance is captured by the model complex cells, and the model simple cells have only small effect on prediction performance (data not shown). However, to be sure that we would obtain the most accurate model possible for each neuron, we included both $S(t)$ and $C(t)$ in this study.

The entire bank of V1 filters used here consisted of 5956 basis functions spanning 12 different directions, five different spatial frequencies, and six different velocities. The spatial frequency of the filters was log distributed from zero to six cycles per classical receptive field (cRF). The temporal frequency was log distributed from 0 to 30 Hz. The filters were spatially tiled on to a two-dimensional Cartesian grid. Grid spacing was set separately at each scale to ensure that the grid width was proportional to the spatial width of the Gaussian envelope in Equation 1. Each adjacent pair of filters was separated by 2.2 σ of the Gaussian envelope. The size of Gaussian envelope was set proportional to the spatial frequency such that one cycle of the sine wave was two σ of the envelope. The overall spatial analysis window was set to two times the size of the classical receptive field. Our preliminary analysis showed that predictions were not improved when the spatial frequency of the simple cell filters increased beyond two cycles per cRF (data not shown). Therefore, to reduce the computational burden, these filters were limited to be no higher than two cycles per cRF.

Regression by boosting. Boosting with early stopping procedure (Friedman, 2001; David et al., 2007; Willmore et al., 2010) was used to model the relationship between V1 filter outputs and responses of each MT neuron. The procedure has the effect of shrinking the total sum of absolute weights (compared with the ordinary least squares regression). This suppresses small weights that cannot be estimated accurately with the data available. The procedure produces a robust fit even when the number of model parameters to be estimated is much larger than the number of data samples. Note that only the training data were used for fitting the model weights; the validation data were preserved for estimating model predictions.

The regression model was defined as follows (see Fig. 2):

$$Y(t) = \sum_i \sum_\tau X_i(t-\tau) W_i(\tau), \quad (5)$$

where $X(t)$ represents the V1 filter outputs given some input (i.e., a segment of a movie), $Y(t)$ is the predicted response, and $W(t)$ is a weight matrix containing linear weights between $X(t)$ and $Y(t)$. (Note that the weight matrix contains weights for correlation delays, τ, up to 10 frames or 130 ms.) According to this definition, fitting the model to the responses of each neuron is simply a matter of estimating the optimal weight matrix.

An iterative procedure was used to estimate the weight matrix as follows: (1) Set all the elements of the weight matrix $W(t)$ to 0. (2) Calculate gradient of the square error E between model and neural responses:

$$E = (Y(t) - r(t))^2, \quad (6)$$

$$\frac{\partial E}{\partial W} = \sum_t (Y(t) - r(t)) X_i(t - \tau). \quad (7)$$

(3) Identify the element with the steepest gradient:

$$(i_m, \tau_m) = \text{argmax}\left(\left| \frac{\partial E}{\partial W} \right| \right). \quad (8)$$

(4) Update the element in the weight matrix by a small step size ε:

$$W_{i_m}(\tau_m) \leftarrow W_{i_m}(\tau_m) - \text{sgn}\left(\frac{\partial E}{\partial W_{i_m,\tau_m}} \right) \varepsilon. \quad (9)$$

(5) Loop back to Step 2 until the termination criterion is met.

Early stopping with cross-validation was used to determine when to terminate boosting. On each iteration of the fitting procedure, 80% of the data from the training set were used to fit the model, and the remaining 20% of the training data were used to evaluate predictions. The boosting procedure was terminated when prediction errors on this training subset began to increase. To estimate parameters optimally this entire procedure was repeated five times, each time reserving a different 20% of the training data as a prediction subset. The final weight estimates were obtained by averaging across these five repetitions. Note that the validation data were never used in any aspect of the fitting procedure, but were only used to estimate prediction accuracy of the final model.

Tests for nonlinearities. To identify additional nonlinearities that might be critical for the model, we developed a switching framework that allowed us to compare several different nonlinear mechanisms directly (see Fig. 2). Each stage could be switched in or out of the circuit, and we exhaustively explored all possible models by parametrically varying which elements were included or excluded from the model. (The V1 filters were present in all cases and were never switched out of the circuit.) The switching framework included three kinds of nonlinearity: (1) luminance and contrast normalization placed before the V1 filters, (2) a static nonlinearity, and (3) divisive normalization.

The luminance and contrast normalization were implemented as a stimulus preprocessing stage interposed between the movie and the V1 filters:

$$I'(x,y,t) = \frac{I(x,y,t) - \text{Lum}(t)}{\text{Con}(t)}, \quad (10)$$

where the $I'(x,y,t)$ is the normalized luminance. The time course of luminance, $\text{Lum}(t)$ was defined as follows:

$$\text{Lum}(t) = \sum_x \sum_y I(x,y,t). \quad (11)$$

$\text{Con}(t)$, the time course of contrast, was defined as follows:

$$\text{Con}(t) = \sqrt{\sum_x \sum_y (I(x,y,t) - \text{Lum}(t))^2} \quad (12)$$

The static output nonlinearity was implemented as a half-wave rectification followed by a power function:

$$X_i'(t) = [X_i(t)]^\alpha \quad (13)$$

where Xi represents the output of the linearized Gabor filters, and Xi' represents the nonlinearly transformed output of the filter bank. Three values for α were used: half-wave rectification given by $\alpha = 1.0$, a compressive nonlinearity given by $\alpha = 0.5$, and an expansive nonlinearity given by $\alpha = 2.0$. Note that contrast responses for V1 neurons are often described using the Naka–Rushton equation (Albrecht and Hamilton, 1982). The Naka–Rushton equation can be linear, compressive, or expansive, depending on the range of stimulus contrast. Which form the contrast response of area MT neurons will take under naturalistic conditions is an open question that can be addressed using our modeling framework.

Divisive normalization was implemented as follows:

$$X_i'(t) = \frac{X_i(t)}{\sum_n X_n^{\text{norm}}(t) + \beta} \quad (14)$$

where $\sum_n X_n^{\text{norm}}(t)$ represents pooled responses of the V1 filters. The $X_n^{\text{norm}}(t)$ were prenormalized so that each of the nth filters had unit SD over the time course of output. (The second term in the denominator, β, is the semisaturation constant for normalization.)

Note that divisive normalization is not selective for any particular range of spatial positions or spatial or temporal frequencies. The suppressive effect is global, both spatially and spectrally. In contrast, the suppressive spectral receptive field (see Figs. 4, 5, blue blobs) is spatially and spectrally localized.

Relationship to other models. The MT neuron model developed here offers both a generalization and a simplification of models proposed previously. Our model is similar in many respects to those proposed in previous neurophysiological studies (Perrone and Thiele, 2001; Rust et al., 2006). However, those studies only probed receptive field properties within a one- or two-dimensional subspace of the full three-dimensional frequency domain [i.e., a two-dimensional slice in the study by Perrone and Thiele (2001) and a one-dimensional ring in the study by Rust et al. (2006)]. Because our model describes receptive field properties within the full three-dimensional spectral domain, it is more general than those proposed previously.

Our model produces receptive fields that are in many respects consistent with those proposed in other studies (Simoncelli and Heeger, 1998; Rust et al., 2006), even though the model requires fewer nonlinear mechanisms than those proposed previously. Simoncelli and Heeger (1998) proposed that rectification and normalization occur within area MT. Rust et al. (2006) proposed a directionally dependent normalization. In our preliminary modeling work (data not shown), we explored models with a second output nonlinearity located within MT (Simoncelli and Heeger, 1998; Rust et al., 2006). However, we found that the second output nonlinearity had no significant effect on predictions, so we discarded it to simplify the model. We also did not include a divisive normalization component within MT, because that component would require strong assumptions regarding the specific form of MT receptive fields (Simoncelli and Heeger, 1998).

A recent study suggested that there are local nonlinear interactions between the receptive subfields of area MT neurons (Majaj et al., 2007). We did not include such interactions in our modeling effort for two reasons. First, we were most interested in the organization of receptive field profiles within the three-dimensional frequency domain, and any interaction effects would not materially affect these estimates. Second, including nonlinear interaction terms between the constituent V1 filters would dramatically increase the number of parameters of the model and so make estimation much more difficult. (Note that the MT model used here contains ~6000 regression channels. Including all possible two-way interactions would require regressions of $\sim 6000^2$, or ~36,000,000 channels).

Our model is also limited in temporal respects. The movies used in this experiment were shown at a frame rate of 83 Hz. Therefore, we did not attempt to model responses at a time scale finer than 12 ms. For this reason, the model does not address subframe nonlinear responses (e.g., transients and bursts).

Optimal velocity plane. To compare estimated spectral receptive fields across the sample of area MT neurons, we computed two indices for each neuron: an on-plane index and a horizontal–vertical ratio index. Both these measures required estimating the optimal velocity plane, that is, the plane within the three-dimensional frequency domain that best fits the spectral amplitude distribution of the excitatory receptive field. The optimal velocity plane crosses the zero point and can be defined by its azimuth (direction) and elevation (speed). Note that

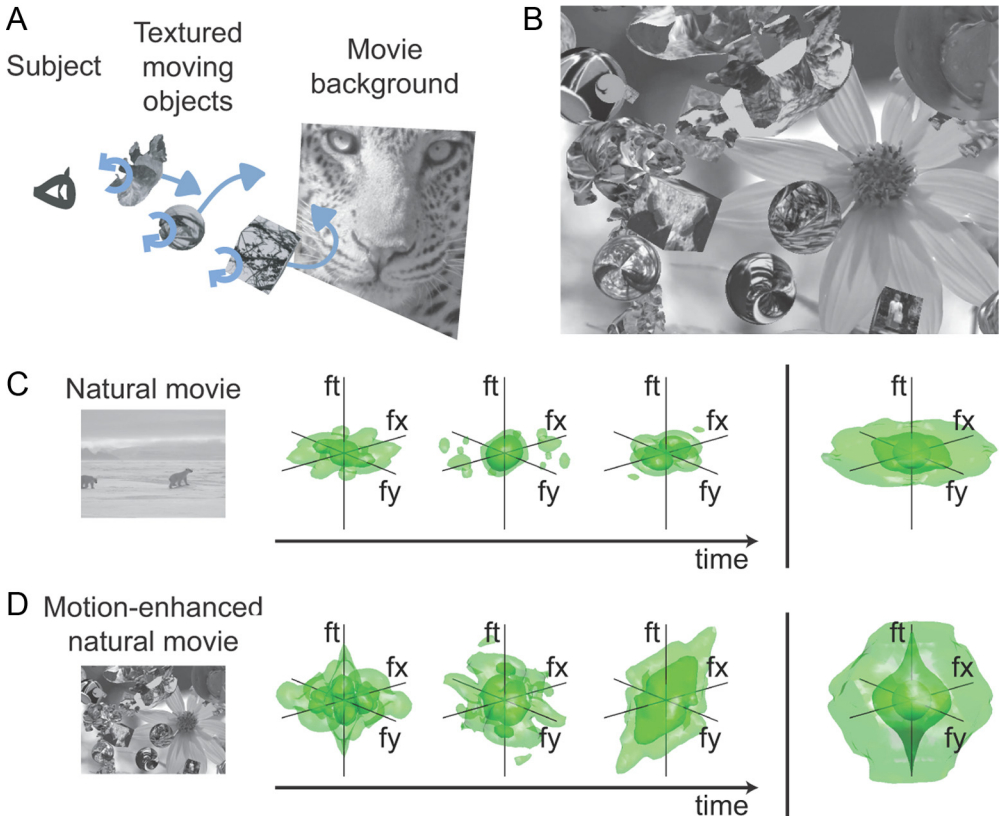


Figure 1. Spatial and spectral structure of motion-enhanced natural movies. ***A***, Schematic diagram of motion-enhanced natural movies. The movies are constructed by combining two distinct components: the background is a natural movie and the foreground contains several textured objects that move along random trajectories. The entire display is updated approximately three times per second to simulate the visual stimulation that would occur due to natural saccadic eye movements. The addition of these foreground objects increases high temporal frequency energy and decorrelates the stimulus, thereby increasing efficiency of receptive field estimation. ***B***, One typical frame of a motion-enhanced natural movie. ***C***, Three-dimensional frequency spectrum of natural movies. The leftmost column shows one frame from a natural movie. The middle columns show three snapshots of the three-dimensional amplitude spectrum of three specific frames from the movie, plotted in the three-dimensional spatiotemporal frequency domain. The rightmost column shows the average spectrum for a long movie. Three isospectral surfaces are plotted (1, 4, and 16% of the maximum) to facilitate visualization. The amplitude spectrum follows a $1/f$ distribution in both space and time. ***D***, Three-dimensional amplitude spectrum for motion-enhanced natural movies. The format is the same as in ***C***. The amplitude spectrum of motion-enhanced natural movies has relatively more energy at high temporal frequencies than is found in natural movies but is otherwise similar to the amplitude spectrum found in natural movies.

the spectrum of any image that translates in a fixed direction and at constant speed will lie on a plane (Watson and Ahumada, 1985; Simoncelli and Heeger, 1998).

To find the azimuth and elevation of the optimal plane for each neuron, we introduced two constraints. The maximum coverage constraint identified the plane that had the maximal coverage of excitatory components on and near the plane. This was found by summing V1 filter weights whose temporal frequency was within ± 1 octave or ± 5 Hz from the plane (whichever was largest). The symmetry constraint identified the plane at the optimal direction. This was found by summing the V1 filter weights separately on the two sides of the azimuth of the plane, subtracting these quantities and taking the negative. [The symmetry constraint was important for neurons such as the one shown in Fig. 4*B*, where the maximal coverage constraint alone would not produce a unique optimal velocity plane (see also Fig. 10 and Discussion).] We manually balanced the importance of the two constraints to produce the most stable estimates of the optimal plane across the neurons in our sample.

Null hypotheses test for on-plane ratio. To determine statistical significance of the on-plane ratio index, we used Monte Carlo simulation to obtain the null distribution. First, 150 model area MT neurons were constructed by randomly assigning weights from a normal distribution with mean 0 and SD 1 (arbitrary units). Second, these model neurons were used to filter the motion-enhanced natural movies that had been used as stimuli in the experiment. Poisson noise was added to the model responses at this stage. Third, spectral receptive fields were estimated for each of the model neurons, using the same techniques described above. Finally, the on-plane ratio index was calculated for each model neuron. These ratios constituted the null comparison distribution. A Wilcoxon rank-sum test was used to assess statistical significance.

Results

We recorded from 52 area MT neurons in two animals while they performed a simple fixation task. During recording, each MT neuron was stimulated with 22,000 to 42,000 frames of motion-enhanced natural movies (Fig. 1; Movie 1). To ensure that the motion-enhanced natural movies did not bias estimated receptive field models, additional recordings were made from a subset of 15 MT neurons using 2000 frames of natural movies without motion enhancement. The data acquired from each neuron were split into two parts: the first part was used to fit receptive field models, and the second was used to evaluate model predictions. A modeling framework based on nonlinear system identification (David et al., 2004; Nishimoto et al., 2006; Wu et al., 2006; Willmore et al., 2010) was used to fit several quantitative computational receptive field models to the data recorded from each MT neuron. To facilitate interpretation and comparison of tuning properties, all receptive fields were visualized in the three-dimensional spatiotemporal frequency domain. Here we refer to these as spectral receptive fields.

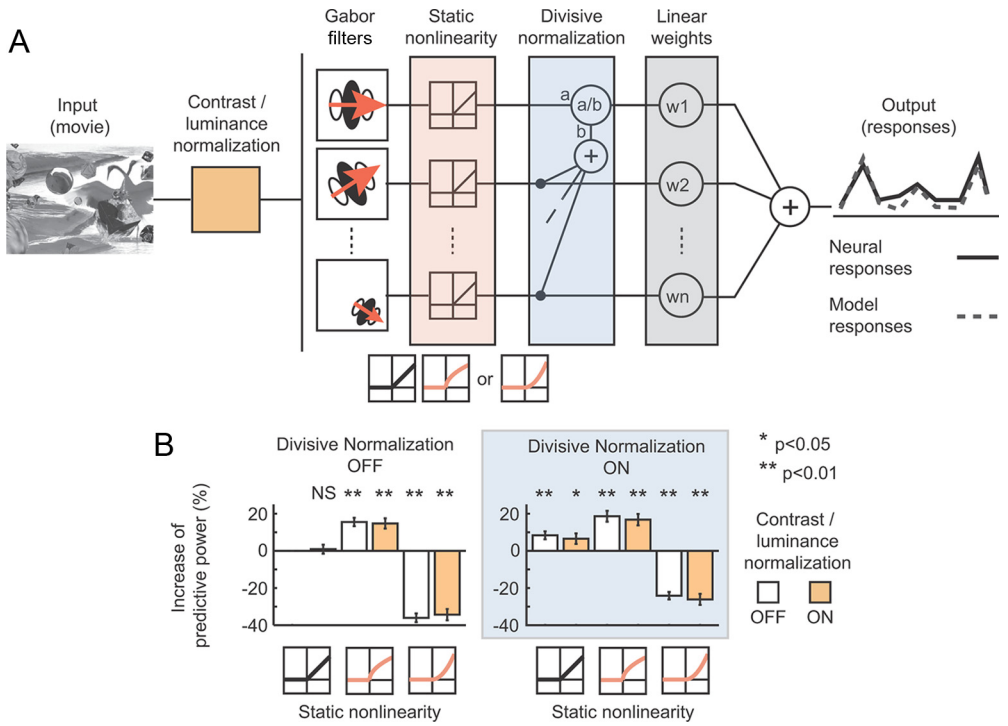


Figure 2. Analysis of MT neurons using the switched model. ***A***, The switched model framework used to describe each MT neuron. The complete model consists of several linear and nonlinear filtering stages. Incoming images are first transformed by a nonlinear contrast and luminance normalization stage (orange square). These signals are fed into a bank of simple and complex type V1 filters (schematic Gabor filters). The output of each V1 filter is rectified with a static linear, compressive, or expansive nonlinearity with half-wave rectification (pink rectangle). These signals are fed in turn into a divisive normalization stage (blue rectangle). Finally, the results are summed linearly according to weights estimated by regularized linear regression (gray rectangle). ***B***, To determine which nonlinear filtering components improve model predictions, we systematically switched each of the nonlinear processing stages shown in ***A*** in and out of the model and compared the predictions of each model. A total of 12 different models [i.e., 2 (with or without luminance and contrast normalization) × 3 (three types of static nonlinearity) × 2 (with or without divisive normalization) different switching conditions] were examined, and results were averaged across all 52 MT neurons in the sample. Each bar represents the average increase in predictive power relative to the simplest model examined. (The simplest model contains linear half-wave rectification, no luminance normalization, and no divisive normalization.) Results from models with or without the luminance and contrast normalization stage are represented as open or filled bars, respectively (see legend, right). The shapes of the three static nonlinearities tested here are shown below each bar. (The three static nonlinearities were linear, compressive, and expansive, all with half-wave rectification.) The left and right panels compare results from models without or with the divisive normalization stage, respectively. Note that the leftmost bar shows the comparison to itself (the simplest model) and thus is guaranteed to be zero. Error bars show bootstrap estimates of the SE ($n = 52$). Significance of prediction power relative to the simplest model is shown above each bar (*$p < 0.05$; **$p < 0.01$; Wilcoxon signed-rank test with Bonferroni correction; NS, not significant).

Linear and nonlinear mechanisms that predict natural visual responses in area MT

The most common framework for modeling single neurons in area MT is to describe each neuron in terms of its inputs from previous stages of visual processing (Simoncelli and Heeger, 1998; Rust et al., 2006; Bradley and Goyal, 2008). The simplest plausible model consists of two stages: a spatiotemporal Gabor filter that represents a pool of area V1 simple and complex neurons (Adelson and Bergen, 1985; Jones and Palmer, 1987), and a linear pooling mechanism that selectively integrates information from a specific subset of putative V1 inputs. This two-stage model provides a simple framework for describing MT neurons, but a complete functional model that can account for responses to naturalistic stimuli will likely require additional nonlinear mechanisms like those that have been reported at more peripheral stages of processing (Kaplan et al., 1987; Heeger, 1992a,b; Carandini et al., 1997; Mante et al., 2005; Bonin et al., 2006). To identify these critical nonlinearities efficiently, we developed a switched model framework that encompassed several different nonlinear mechanisms identified previously in area MT or at more peripheral stages of processing: luminance and contrast normalization (Kaplan et al., 1987; Bonin et al., 2006), a static nonlinearity (Heeger, 1992a), and divisive normalization (Heeger, 1992b; Carandini et al., 1997).

The prediction accuracy of each of the nonlinear models is summarized in Figure 2*B* (see Materials and Methods for details). Each bar in the histograms shows how well one specific model predicts responses in the validation data set, relative to the simplest model that includes only the V1 filtering stage with static rectification and no additional nonlinearities. Contrast normalization does not have a significant effect on predictions compared with the simplest model ($p > 0.10$; Wilcoxon signed-rank test with Bonferroni correction). However, the static output nonlinearity does have a significant effect. The compressive static output nonlinearity consistently increases predictions ($p < 0.01$), while the expansive nonlinearity always decreases predictions ($p < 0.01$). Divisive normalization also improves predictions significantly ($p < 0.01$). The model that contains both a compressive nonlinearity and divisive normalization has significantly more predictive power than a model that contains only a compressive nonlinearity ($p < 0.01$). Note, however, that the effects of the compressive and expansive nonlinearities do not depend on whether divisive normalization is present or not. Based on these results, we conclude that the simplest model of MT neurons that gives accurate predictions of responses to motion-enhanced natural movies consists of a bank of V1 filters, each followed by a compressive nonlinearity, a divisive nonlinearity, and a linear

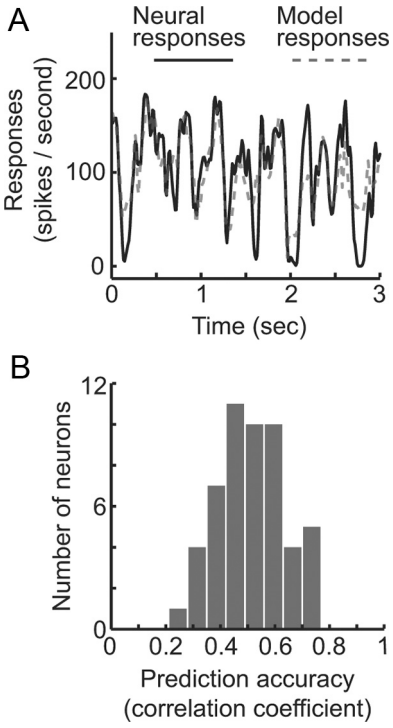


Figure 3. Prediction accuracy of the neuronal model used in this study. ***A***, Model predictions for one MT neuron. The horizontal axis indicates time, and the vertical axis the response rate. The solid curve gives the mean response observed over time (10 repetitions), and the dotted curve indicates predicted responses. For this neuron, the correlation between predicted and observed responses is $r = 0.72$. ***B***, Summary of prediction performance across the entire sample of 52 neurons. The average correlation is $r = 0.52$. The proportion of variance explained by the model is 35%, which is approximately comparable to the predictions obtained in previous studies of areas V1 and V2 from our laboratory (V1, 40%; V2, 30%) (David and Gallant 2005; Willmore et al., 2010).

pooling stage whose weights are determined uniquely for each neuron.

Figure 3*A* illustrates the predictions of the compressive and divisive V1 filter model fit to responses of one MT neuron. The correlation between model predictions and observed responses for this neuron is quite good ($r = 0.72$) considering the inherent variability of neural responses. Figure 3*B* summarizes prediction performance of the compressive/divisive V1 filter model across the entire sample of 52 MT neurons. The average correlation between model predictions and observed responses is $r = 0.52$, and the model accounts for 35% of the explainable response variance (David and Gallant, 2005). This is somewhat lower than the response variance that can be explained in LGN neurons (~60%) (Mante et al., 2008), but comparable to the response variance explained in V1 (~40%) (David and Gallant, 2005; Willmore et al., 2010) and in V2 (~30%) (Willmore et al., 2010). The prediction accuracy achieved here is remarkable given that predictions were estimated for time-varying responses, evoked by movies that were not used to fit the model.

Spectral receptive fields of area MT neurons

In a seminal theoretical paper, Simoncelli and Heeger (1998) hypothesized that the receptive fields of MT neurons should lie on a single plane within the three-dimensional frequency domain. Their reasoning was based on the fact that the three-dimensional power spectrum of an image translating at a fixed velocity will lie on a plane whose azimuth and elevation reflect the speed and direction of image motion (Watson and Ahumada, 1985; Simoncelli and Heeger, 1998; Bradley and Goyal, 2008). Thus, any neuron that has a planar receptive field in the three-dimensional frequency domain will be optimally tuned for one specific image velocity. Due to the spatiotemporal bandpass nature of the V1 inputs to MT, Simoncelli and Heeger (1998) predicted that spectral receptive fields of MT neurons would form a ring in the three-dimensional frequency domain. In the same paper, Simoncelli and Heeger (1998) also postulated that MT neurons might possess suppressive receptive fields that lie off the optimal excitatory velocity plane. These suppressive components would tend to sharpen velocity tuning.

The velocity plane tuning model proposed by Simoncelli and Heeger (1998) is consistent with many previous neurophysiological studies in area MT (Movshon et al., 1985; Rodman and Albright, 1987; Snowden et al., 1991; Britten et al., 1993; Perrone and Thiele, 2001), and with human psycophysical studies (Schrater and Simoncelli, 1998; Schrater et al., 2000). However, previous neurophysiological studies examined only a one- or a two-dimensional subspace within the full three-dimensional frequency domain (Okamoto et al., 1999; Perrone and Thiele, 2001; Priebe et al., 2003). Therefore, none of them provided direct evidence of tuning along the three-dimensional velocity plane (see also Discussion, Relationship to previous reports of speed-tuned neurons). Furthermore, no previous study has investigated three-dimensional suppressive tuning in area MT.

In this section, we present data that resolve both of these long-standing issues. To directly examine excitatory and suppressive tuning, we visualize the receptive field of each neuron in the full three-dimensional frequency domain. Spectral receptive fields were obtained by first multiplying the three-dimensional amplitude spectrum of each V1 filter by its fitted weight (Fig. 2*A*) and then summing across filters and correlation delays. Excitatory and suppressive receptive fields were obtained by summing spectra for either positive or negative weights separately.

The receptive fields of some of the MT neurons in our sample are consistent with the predictions of Simoncelli and Heeger (1998). One such neuron is shown in Figure 4*A*. The first three columns show the spectral receptive field of this neuron, viewed from three different angles. (For clarity, the plot has been rotated so that the preferred direction of motion is aligned with the *x*-axis.) The three rows show the excitatory components (top, positive fit weights), the suppressive components (middle, negative fit weights), and the combined receptive field (bottom). The transparent red and blue surfaces in each panel delineate the excitatory and suppressive isospectral surfaces. These surfaces were obtained by thresholding the aggregated amplitude spectrum of the Gabor filters at 25, 50, and 75% of the spectral peak. For this neuron, the excitatory receptive field forms a ring that lies on a single velocity plane in the frequency domain, and the suppressive receptive field lies off the excitatory plane.

The receptive fields of some of the other MT neurons in our sample are not rings, but rather encompass a narrow range of spatial and temporal frequencies. One such neuron is shown in Figure 4*B* (format same as Fig. 4*A*). The excitatory receptive field of this neuron is confined to a single point in the three-dimensional frequency domain, and there is little evidence of any substantial suppressive receptive field.

The neurons shown in Figure 4 represent the most extreme examples in our sample. In fact, most of the MT neurons lie between these two extremes. Two examples that are more typical of the sample as a whole are shown in Figure 5 (format same as Fig. 4). The excitatory receptive fields of these neurons lie predominantly on a single velocity plane, and they are elongated

along the frequency axis perpendicular to the optimal direction. However, they form only a partial ring in the plane. Thus, these neurons are insensitive to frequencies near the zero temporal frequency axis (compare Figs. 4*A*, 5). As far as we know, this pattern of tuning in area MT has not been described previously.

Excitatory spectral receptive fields lie on a plane

Simoncelli and Heeger (1998) predicted that the excitatory receptive fields of MT neurons should tend to lie on a single plane in the three-dimensional frequency domain. To address this issue quantitatively, we created an index that describes the proportion of the spectral receptive field of each MT neuron that lies on versus off of the optimal velocity plane. If the excitatory receptive field of an MT neuron lies on or near the optimal plane, then this on-plane ratio index will be near 1, and if it lies off of this plane, the index will be near 0. Figure 6*A* summarizes the on-plane ratio for the excitatory receptive fields of all 52 area MT neurons in our sample. The average ratio is 0.59.

Because our definition of the optimal velocity plane relies on maximizing the coverage of excitatory receptive fields on and near this plane (see Materials and Methods), by definition the on-plane ratio will be biased toward positive values. Therefore, to determine which index values were significantly greater than chance, we ran a Monte Carlo simulation to estimate the null distribution (see Materials and Methods). We created 150 model compressive/divisive neurons, each model seeded with random weights. We then estimated the on-plane ratio for each of these random model neurons. The average on-plane ratio for this null model is 0.36 (Fig. 6*A*, dashed line). This value is significantly lower than the value we observed across the real sample of neurons ($p < 0.01$, Wilcoxon rank-sum test). These data confirm that the excitatory receptive fields of MT neurons tend to lie on a single plane in the three-dimensional frequency domain, consistent with the predictions of Simoncelli and Heeger (1998).

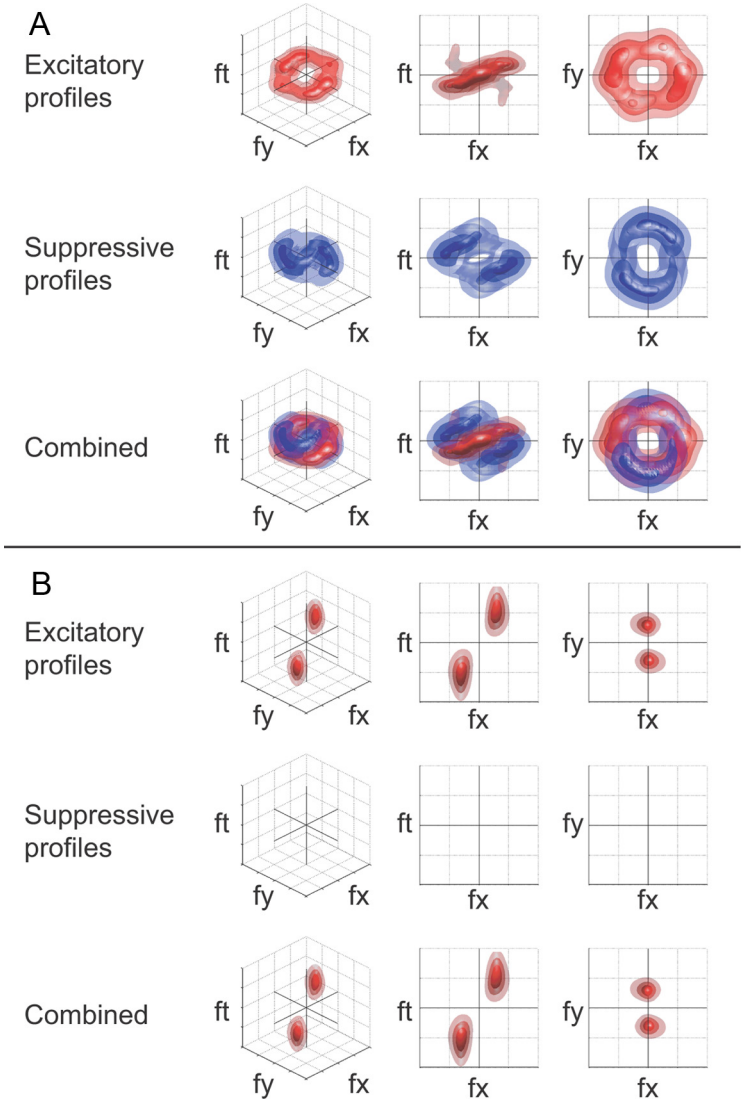


Figure 4. Estimated excitatory spectral receptive fields for two MT neurons. ***A***, Excitatory spectral receptive field for an MT neuron that is consistent with the predictions of Simoncelli and Heeger (1998). The three columns represent different views of the three-dimensional frequency domain. The red shells in the top row indicate excitatory isospectral contours (25, 50, and 75% of the spectral peak), the blue shells in the middle row indicate suppressive isospectral contours, and the bottom row shows both excitatory and suppressive shells in the same plots. Ticks for each axis show five cycles per receptive field for spatial frequency and 25 Hz for temporal frequency. To facilitate visualization, the spectral receptive field has been rotated so that the preferred direction of motion is aligned with the *x*-axis in the frequency domain. The excitatory receptive field for this neuron forms a ring in the three-dimensional frequency domain, and the suppressive receptive field encompasses a wide band of off-plane frequencies. Thus, this neuron is tuned for a single velocity. ***B***, Spectral receptive field for an MT neuron that encompass a narrow range of spatial and temporal frequencies. The format is the same as in ***A***.

Suppressive spectral receptive fields lie off the optimal excitatory plane

Simoncelli and Heeger (1998) also speculated that the suppressive receptive fields of MT neurons tend to lie off the optimal excitatory plane. To address this issue we simply applied the on-plane ratio to the suppressive (rather than the excitatory) spectral receptive field of each neuron. If the suppressive receptive field of an MT neuron tends to avoid the optimal velocity plane, then this ratio will be near 0, and if it lies on the optimal plane, then the ratio will be near 1. Figure 6*B* summarizes the suppressive on-plane ratio obtained across our sample. The average ratio is 0.18, which is significantly smaller than the value for the null model described above ($p < 0.01$, Wilcoxon rank-sum test). These data confirm that the suppressive receptive fields of MT neurons tend to avoid the optimal excitatory plane, as proposed by Simoncelli and Heeger (1998).

Excitatory spectral receptive fields of most MT neurons form a partial ring in the plane

Inspection of the spectral receptive fields estimated for individual MT neurons suggests that the population might differ along one simple dimension: the degree to which their excitatory receptive fields fill the optimal velocity plane within the three dimensional frequency domain. To characterize this, we created a separate

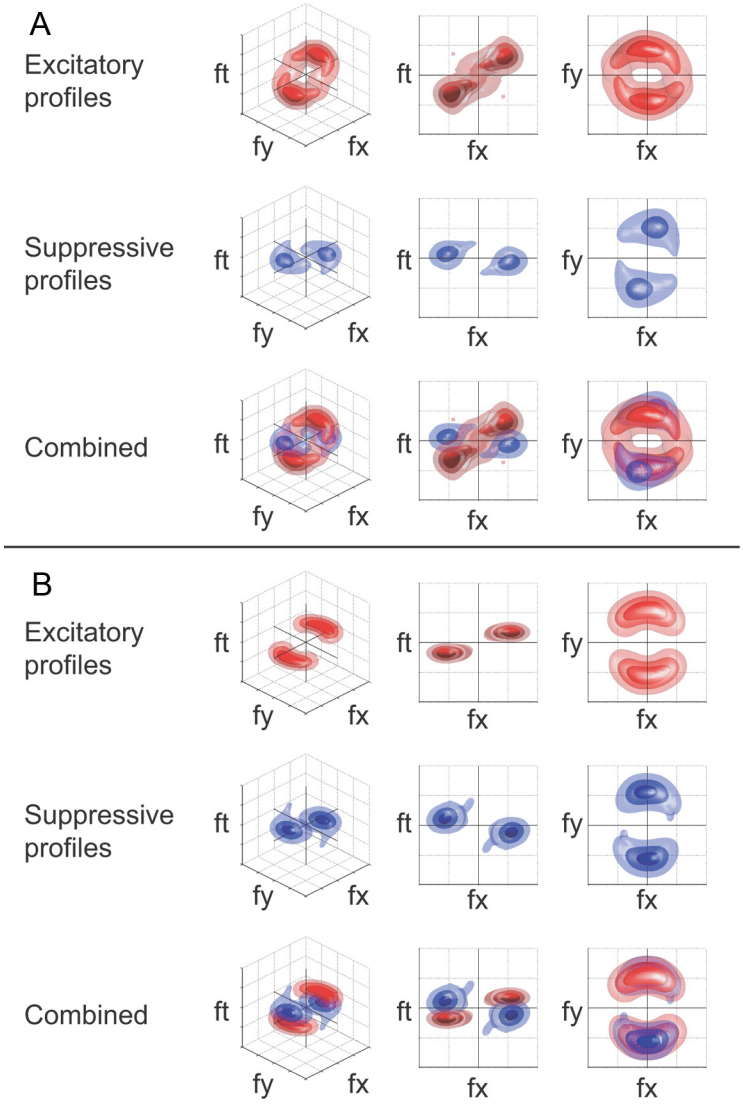


Figure 5. Estimated spectral receptive fields for two MT neurons typical of the sample as a whole. The format is the same as in Figure 4. ***A***, An MT neuron whose excitatory receptive field forms a partial ring in the optimal velocity plane. This neuron is relatively insensitive to temporal frequencies near zero. Thus, this neuron is tuned for a specific velocity but is not very sensitive to static texture oriented along the optimal direction of motion. ***B***, A second MT neuron whose excitatory receptive field forms a partial ring in the three-dimensional frequency domain. This neuron is tuned for a specific velocity, but will not respond to static texture oriented along the optimal direction of motion.

index that describes how well the excitatory spectral receptive field of each neuron fills the optimal velocity plane. First, we divided the optimal plane into four quadrants, two centered around the vertical axis (along the optimal direction) and two centered around the horizontal axis (the axis embedded in the zero temporal frequency plane). Then we integrated the excitatory spectral receptive field amplitudes in the vertical versus horizontal quadrants and took the ratio of these values. According to this horizontal–vertical index, a ratio of 1 indicates that an MT neuron forms a perfect ring in the optimal velocity plane, while a ratio of 0 indicates that the neuron is tuned to a single spatial and temporal frequency.

Figure 7*A* summarizes the horizontal–vertical ratios estimated for all 52 area MT neurons in our sample. The distribution is clearly continuous. At one end of the distribution lie MT neurons that have spectral receptive fields that form a ring in the optimal frequency plane, consistent with predictions of Simoncelli and Heeger (1998). At the opposite end of the distribution lie neurons that have spectral receptive fields confined to a unique spatial and temporal frequency. However, the receptive fields of the large majority of MT neurons lie between these extremes, forming a partial ring in the optimal velocity plane and avoiding the region near zero temporal frequency. Thus, the vast majority of MT neurons are relatively insensitive to static patterns aligned with the optimal direction of motion as opposed to what would be predicted according to the proposal of Simoncelli and Heeger (1998). [Note that there was no significant correlation between prediction performance and the horizontal–vertical ratio ($p > 0.10$).]

The horizontal–vertical ratio is correlated with simulated pattern selectivity

Many previous studies have used plaid stimuli to assess motion selectivity of area MT neurons (Movshon et al., 1985; Pack and Born, 2001; Smith et al., 2005; Rust et al., 2006; Majaj et al., 2007). A plaid consists of two superimposed gratings that move in different directions. MT neurons vary substantially in their responses to plaids. Some MT neurons are selective to the direction of the component gratings, whereas others are selective for the aggregate direction. These are called component-selective and pattern-selective neurons, respectively. Recent studies have reported that MT neurons do not form discrete component- and pattern-selective classes, but rather that selectivity is distributed continuously between these two extremes (Smith et al., 2005; Rust et al., 2006).

We performed a simulation to determine how the continuum of tuning in the optimal plane that we report here is related to the continuum of component and pattern selectivity found in previous studies (Smith et al., 2005; Rust et al., 2006). We first simulated responses of all MT neurons in our sample to both single grating and plaids. Then we used the pattern index developed in previous studies (Smith et al., 2005) to characterize plaid selectivity. The pattern index quantifies directional selectivity for plaid stimuli: it is positive if a neuron is selective for the aggregate direction of a plaid, and it is negative if a neuron is selective for the direction of the component gratings. Thus, the results of this simulation can be interpreted as a prediction about the plaid selectivity that we would expect to obtain for each of the neurons in our sample had we probed them with plaids.

Figure 7*B* summarizes the pattern index distribution predicted for all 52 area MT neurons in our sample. The distribution forms a clear continuum that captures the major distributions reported in previous studies (e.g., Rust et al., 2006, their Fig. 2*b*). The pattern index for most neurons ranges from −6 to 2, and highly pattern-selective neurons are relatively rare. To determine how the pattern index used in plaid studies is related to the horizontal–vertical ratio developed here, we compared these two in-

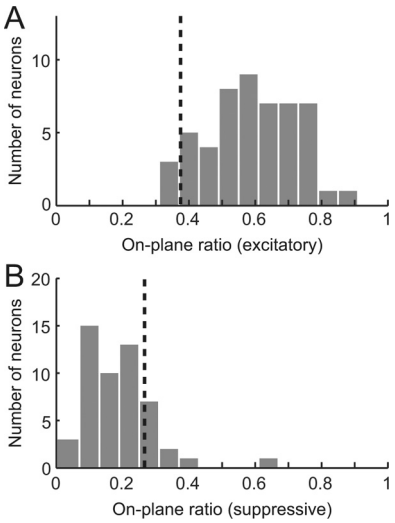


Figure 6. Spectral tuning to frequencies on versus off of the optimal plane within the three-dimensional frequency domain. ***A***, For each neuron, we estimated the plane that captures the largest amount of positive (excitatory) weight within the three-dimensional frequency domain. We then calculated the on-plane ratio, the ratio of positive weights near the plane over the total positive weights. Across the sample of 52 neurons, this ratio is significantly larger than chance ($p < 0.01$, Wilcoxon rank-sum test), indicating that the excitatory spectral receptive fields of MT neurons tend to lie along the optimal velocity plane. ***B***, On-plane ratios as in ***A***, but calculated using the negative (suppressive) weights for each neuron, and the same optimal plane as described in ***A***. Across the sample, the ratio is significantly smaller than chance ($p < 0.01$, Wilcoxon rank-sum test), indicating that the suppressive spectral receptive fields of MT neurons tend to lie off of the optimal velocity plane.

dices for each MT neuron in our sample (Fig. 7*C*). The correlation between the two indices is significant ($r = 0.46$, $p < 0.01$; *t* test for correlation coefficients). Thus, responses to plaids can be partly described in terms of the horizontal–vertical tuning ratio. There was no significant correlation between prediction performance and the pattern index ($p > 0.10$).

Control to identify any bias arising from the use of motion-enhanced natural movies

One potential concern with our results is that the stimuli used in our experiments were motion-enhanced natural movies whose spatial and temporal frequency spectra differ somewhat from those of natural movies (for details, see Materials and Methods; Fig. 1). To ensure that our use of motion-enhanced natural movies did not bias receptive field estimates we recorded from a subset of 15 area MT neurons using both motion-enhanced natural movies and simple natural movies as stimuli. We then compared predictions obtained when neuronal models were fit using motion-enhanced natural movies and tested using a separate set of motion-enhanced natural movies versus when those same models were tested using simple natural movies without motion enhancement. If simple natural movies evoke nonlinear responses that cannot be described by receptive fields estimated using motion-enhanced natural movies, then models estimated using motion-enhanced movies will fail to predict responses to movies without motion enhancement (David et al., 2004).

Receptive field models of area MT neurons estimated using motion-enhanced natural movies predicted responses to novel simple natural movies just as well as they predicted responses to novel motion-enhanced natural movies (average $r = 0.533$ for natural movies, average $r = 0.537$ for motion-enhanced movies; $p > 0.10$, Wilcoxon signed-rank test). This important control demonstrates that models estimated using motion-enhanced

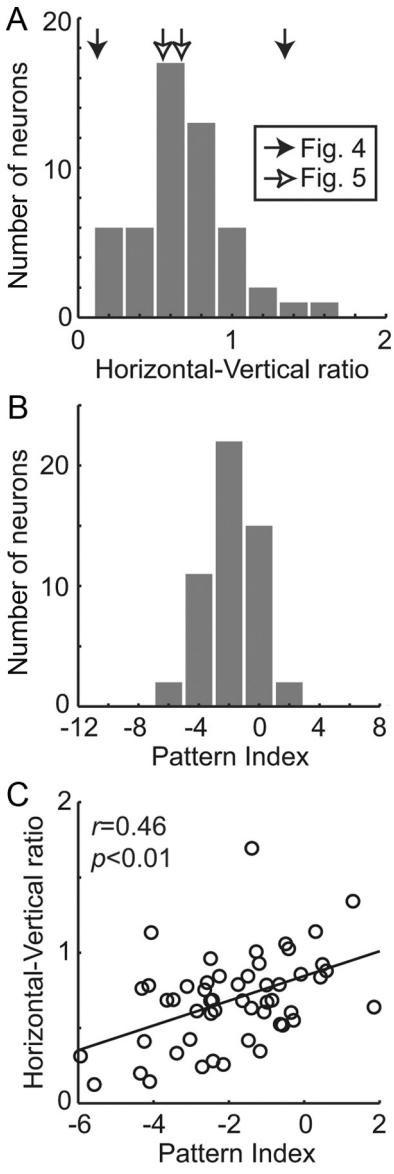


Figure 7. MT neurons differ in their sensitivity to low temporal frequencies. ***A***, Distribution of the horizontal–vertical ratio across the entire sample of 52 area MT neurons. The four neurons shown in Figures 4 and 5 are indicated by arrows. Only a small fraction of MT neurons have profiles consistent with the Simoncelli and Heeger (1998) model (ratio of ~1; Fig. 4*A*) or the unique energy model (ratio of ~0; Fig. 4*B*). The majority of MT neurons have profiles that are midway between these two extremes (Fig. 5). These neurons are insensitive to temporal frequencies near zero, so they do not respond to static texture patterns aligned with the optimal direction of motion. ***B***, Distribution of the pattern index derived from simulated responses to plaid and grating stimuli across the entire sample of MT neurons. The distribution generally agrees with previous studies (Rust et al. 2006, their Fig. 2*b*). ***C***, Joint scatter plot of horizontal–vertical ratio and pattern index. There is a significant correlation ($p < 0.01$) between these two indices.

natural movies accurately describe and predict responses to natural movies. Based on this result, the predictions reported throughout this manuscript reflect the average prediction for both natural and motion-enhanced and natural movies (when the latter were available).

Control to ensure that the model estimation procedure can recover spectral receptive fields of any shape

The model-fitting algorithms used here are closely related to those used in previous papers from our laboratory (David and

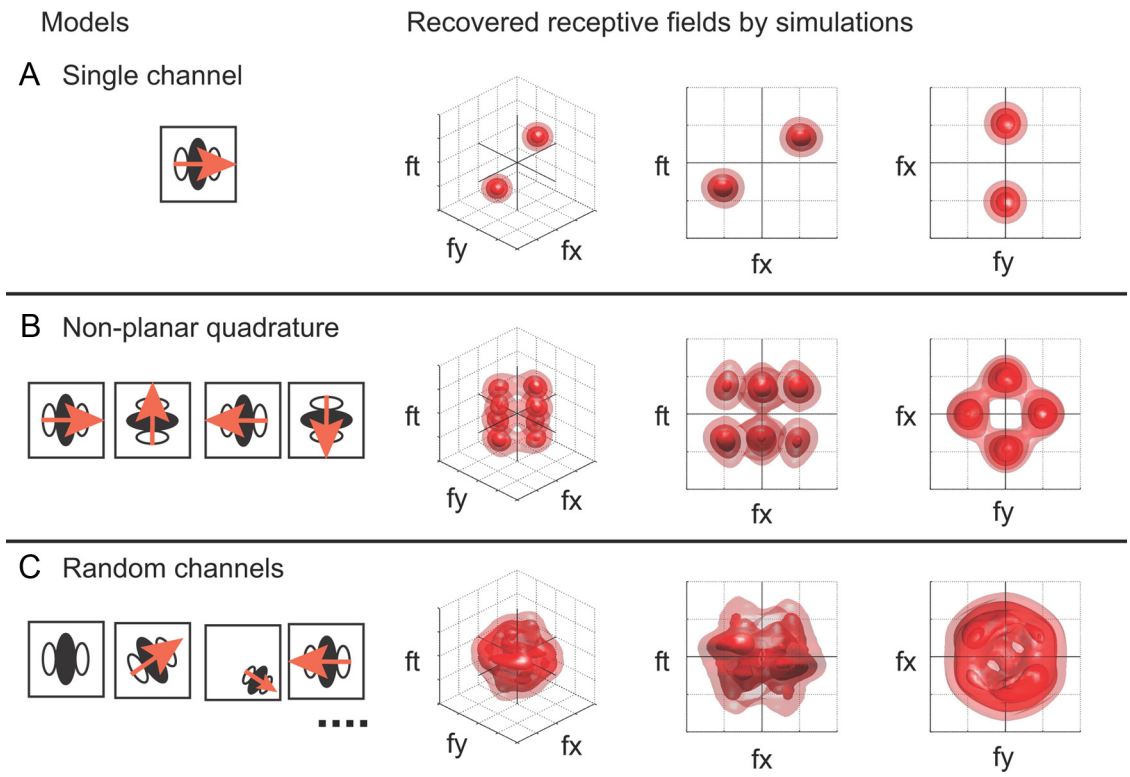


Figure 8. Demonstration that our receptive field estimation procedures can recover spectral receptive fields of various shapes. ***A***, A simulated MT neuron that receives input from a set of V1 neurons that are all tuned to a narrow range of spatial and temporal frequencies. These input neurons are all consistent with the compressive/divisive Gabor model (but only excitatory weights were included here). The left column shows a schematic description of the simulated MT neuron. The right column shows the spectral receptive field for the simulated neuron estimated using the same procedures used to characterize the real MT neurons in this study. The format is the same as in Figure 4. Our procedure correctly recovers the spectral receptive field. ***B***, A simulated MT neuron that receives input from four sets of V1 neurons each tuned to a different direction, but where all inputs are tuned for the same spatial and temporal frequency. Our procedure correctly recovers the spectral receptive field. ***C***, A simulated MT neuron that receives input from many V1 neurons, each tuned to a random direction and spatial and temporal frequency. Our procedure correctly recovers the spectral receptive field.

Gallant, 2005; Willmore et al., 2010) and from other groups (David et al., 2007). These powerful algorithms are rather complicated, and some readers might be concerned that the fitting procedures might have biased the results so as to produce the spectral receptive fields reported here. For example, given that translational motion in natural movies always will produce a planar three-dimensional frequency spectrum, would it ever be possible to recover nonplanar receptive fields? To address this concern, we used the same receptive field estimation procedures described previously in this paper to estimate receptive fields of three simulated neurons whose receptive field organization was quite different from that observed in our sample of real MT neurons.

Figure 8 shows spectral receptive fields for three simulated MT neurons (the format is the same as in Fig. 4; note that the model neurons did not have suppressive receptive fields). The first simulated MT neuron receives input from a set of V1 neurons that are all tuned to a narrow range of spatial and temporal frequencies. Our procedure correctly recovers the spectral receptive field for this neuron. The second simulated MT neuron receives input from four sets of V1 neurons each tuned to a different direction, but where all inputs are tuned for the same spatial and temporal frequency. This is an interesting test case because this simulated MT neuron does not have a planar spectral tuning profile. However, our procedure still recovers the correct spectral receptive field. Finally, the third simulated MT neuron receives input from many V1 neurons, each tuned to a random direction and spatial and temporal frequency. This is another interesting test case because this model MT neuron does not have a planar spectral tuning profile, and it does not avoid low temporal frequencies. Once again our procedure correctly recovers the spectral receptive field. These results demonstrate that the receptive field estimation procedure used in this study can characterize arbitrary spectral receptive fields, regardless of their organization within the three-dimensional frequency domain.

Discussion

Area MT has been the target of intensive neurophysiological investigation over the last 25 years (for review, see Born and Bradley, 2005). However, most previous studies of MT have used simple, parameterized stimuli that spanned only a subspace within the full three-dimensional frequency domain, and it is unclear how the mechanisms revealed under those conditions will generalize to natural vision. We investigated this issue by recording responses of MT neurons evoked by naturalistic movies. We found that the simplest model of MT neurons that accurately predicts responses to these movies consists of a bank of Gabor filters, each followed by either a half-wave rectification or motion-energy computation, a compressive nonlinearity, a divisive nonlinearity, and a linear pooling stage whose weights are determined uniquely for each neuron. This result confirms that concepts of motion coding in MT developed using synthetic stimuli (Simoncelli and Heeger, 1998) are generally valid under more naturalistic conditions.

Our study provides the first reconstructions of spectral receptive fields of MT neurons within the full three-dimensional fre-

quency domain, and it demonstrates that these neurons have planar receptive fields within this domain. The excitatory receptive fields of a few of these neurons form a ring in the optimal velocity plane, consistent with predictions in Simoncelli and Heeger (1998). Another small group of MT neurons have excitatory receptive fields tuned for one unique spatial and temporal frequency. However, the receptive fields of most MT neurons form a partial ring in the optimal velocity plane, avoiding very low temporal frequencies. In sum, the entire population of MT neurons can be characterized along three dimensions: the orientation and the elevation of the optimal velocity plane, and the extent to which the excitatory receptive field forms a ring in the optimal plane (Fig. 9).

Simoncelli and Heeger (1998) also predicted that the receptive fields of some MT neurons might form a partial ring in the optimal velocity plane. However, they predicted that this partial ring would be elongated toward the origin (i.e., zero spatial and temporal frequency). In contrast, we find that these partial ring receptive fields are elongate horizontally along the optimal direction of motion (Fig. 5). This has important functional implications: elongation toward the origin does not preserve velocity tuning, while horizontal elongation along the optimal direction of motion does preserve velocity tuning (see also below and Fig. 10).

We used a computational simulation to show that the MT receptive fields estimated here explain general aspects of plaid responses reported previously (Movshon et al., 1985; Pack and Born, 2001; Smith et al., 2005; Rust et al., 2006; Majaj et al., 2007). However, while several previous studies have reported that a small minority of MT neurons are extremely pattern selective (Smith et al., 2005; Rust et al., 2006), our simulations did not identify any such neurons. One possibility is that extreme pattern selectivity depends on a tuned normalization mechanism (Rust et al., 2006) that was not included explicitly in our model. However, our model does include a compressive nonlinearity on each channel, and this might perform a function similar to the tuned normalization of Rust et al. (2006). Furthermore, Rust et al. (2006) reported no significant relationship between tuned normalization and the pattern index, suggesting that the mechanism does not play a major role in explaining the pattern index quantitatively. Another possibility is that extreme pattern selectivity is only observed when MT neurons are probed with simple plaids,

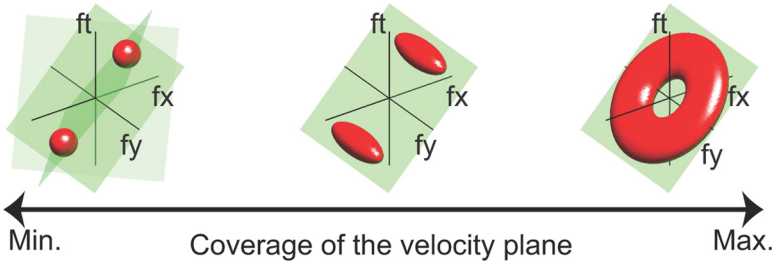


Figure 9. Area MT neurons vary in the degree to which their excitatory spectral receptive fields form a ring within the optimal velocity plane. On one extreme lie MT neurons whose spectral receptive fields are tuned for one unique spatial and temporal frequency. These neurons are not tuned for speed or velocity because there are infinite combinations of speed and direction that are consistent with the receptive field. On the other extreme lie neurons whose spectral receptive fields form a ring on the optimal velocity plane. These neurons are tuned for velocity as originally proposed by Simoncelli and Heeger (1998). However, they also respond to static texture that is aligned with the optimal direction of motion. The majority of MT neurons lie between these extremes. These neurons have excitatory spectral receptive fields that are elongated parallel to the optimal velocity plane, forming a partial ring in the plane and avoiding low temporal frequencies. These neurons are also tuned for velocity, but they do not respond to static texture.

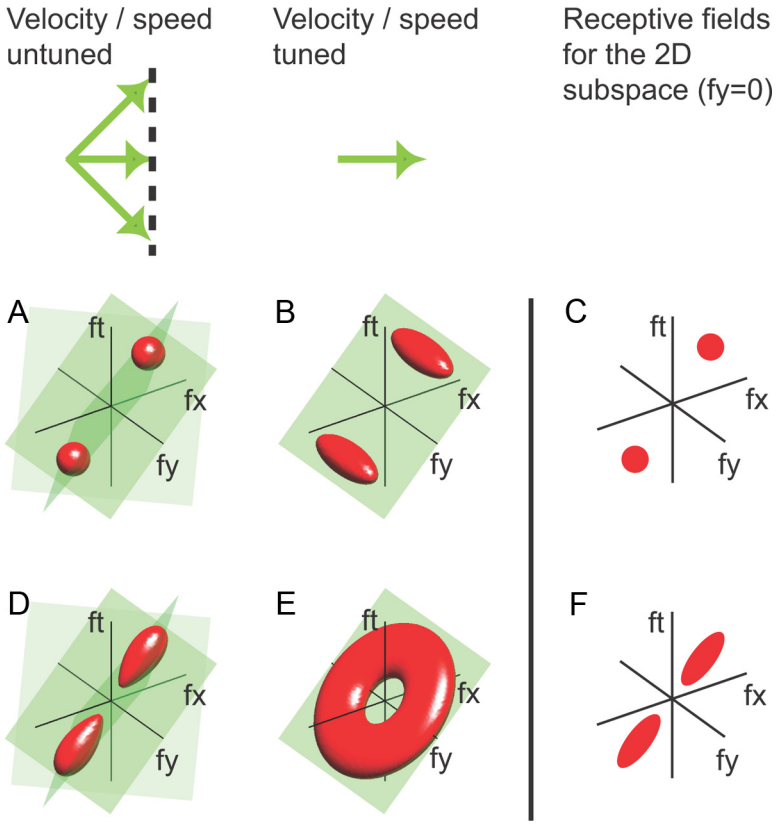


Figure 10. Spectral receptive fields estimated within a two-dimensional frequency subspace cannot unambiguously reveal tuning for speed and velocity. ***A***, Spectral receptive field of a hypothetical neuron that is tuned for a unique spatial and temporal frequency. The format is the same as in Figure 4, except that the only excitatory receptive fields are shown. This neuron is not tuned for speed and velocity because there are many velocity planes (green) that pass through the receptive field. ***B***, Spectral receptive field of a hypothetical neuron that is similar to those typically observed in area MT. This neuron is tuned for both speed and velocity, because one specific velocity plane maximizes overlap with the receptive field. ***C***, Two-dimensional slice ($fy = 0$) (Perrone and Thiele 2001; Priebe et al. 2003) through the spectral receptive field of the hypothetical neurons shown in ***A*** and ***B***. A researcher who only had access to the receptive fields within this subspace might conclude that neither of the neurons shown in ***A*** and ***B*** are tuned for speed, although the neuron shown in ***B*** is tuned for both speed and velocity. ***D***, Hypothetical neuron whose spectral receptive field is elongated toward the origin. This neuron is also not tuned for speed or velocity, because there are many velocity planes that pass through the receptive field. ***E***, Spectral receptive field of a hypothetical neuron that forms a ring as proposed by Simoncelli and Heeger (1998). The neuron is tuned for velocity, because only one velocity plane maximizes overlap with the receptive field. ***F***, Two-dimensional slice through the spectral receptive field of the hypothetical neurons shown in ***D*** and ***E***. A researcher who had access to only the receptive fields within this subspace might conclude that both of the neurons shown in ***D*** and ***E*** are tuned for speed, although the neuron shown in ***D*** is not tuned for speed or velocity. These examples illustrate that it is impossible to definitively determine whether a neuron is tuned for speed or velocity by examining only a two-dimensional subspace within the full three-dimensional frequency domain.

and that responses change when these neurons are probed using more naturalistic stimuli. Analogous stimulus-dependent effects have been reported in studies of area V1 (David et al., 2004; Felsen et al., 2005; Sharpee et al., 2006).

We found that luminance and contrast normalization does not appear to improve model predictions beyond what can be achieved using a simpler model without these mechanisms (Fig. 2). However, although our stimuli span the range of luminance commonly used in neurophysiology experiments (8 to 130 cd/m^2), it is conceivable that these luminance and contrast normalization might be important under more natural conditions containing a wider luminance range (Lewen et al., 2001).

Relationship to previous reports of speed-tuned neurons

Several neurophysiological studies have used drifting gratings to measure speed tuning in area MT (Perrone and Thiele, 2001; Priebe et al., 2003). These studies optimized the direction of grating drift for each neuron individually while systematically varying spatial and temporal frequency. Thus, each study probed a two-dimensional slice of the full three-dimensional frequency domain. However, speed (and velocity) tuning cannot be established unequivocally with stimuli that are confined to a two-dimensional slice. To see why this is so, consider the four hypothetical neurons shown in Figure 10. Figure 10*A* shows the spectral receptive field for a hypothetical neuron tuned for a unique combination of spatial and temporal frequencies. This neuron is not tuned for speed or velocity because many different velocity planes (i.e., many different combinations of speeds and directions; shown in green) pass through the receptive field (Movshon et al., 1985; Simoncelli and Heeger, 1998; Bradley and Goyal, 2008). Figure 10*B* shows the spectral receptive field of a hypothetical neuron similar to those typically found in area MT. This neuron is tuned for speed and velocity because only one velocity plane maximizes overlap with the receptive field. Figure 10*C* shows a two-dimensional slice through the spectral receptive field of the hypothetical neurons shown in *A* and *B*. In the three-dimensional space, the two-dimensional subspace examined in previous studies (Perrone and Thiele, 2001; Priebe et al., 2003) forms the slice defined by $f_y = 0$. A researcher who had access only to receptive fields within this subspace might conclude that neither of the neurons shown in Figure 10, *A* and *B*, are tuned for speed, though the neuron shown in *B* is tuned for speed (and velocity).

Figure 10*D* shows a hypothetical neuron whose spectral receptive field is elongated toward the origin, as predicted by Simoncelli and Heeger (1998). This neuron is not tuned for speed or velocity because many different velocity planes pass through the receptive field. Figure 10*E* shows the spectral receptive field of a hypothetical neuron that forms a ring, as predicted by Simoncelli and Heeger (1998). This neuron is tuned for speed and velocity because only one velocity plane maximizes overlap with the receptive field. Figure 10*F* shows a two-dimensional slice through the spectral receptive fields of the hypothetical neurons shown in Figure 10, *D* and *E*. A researcher who only had access to receptive fields within this subspace might conclude that both of the neurons shown in Figure 10, *D* and *E*, are tuned for speed, though the neuron shown in *D* is not tuned for speed (or velocity). These examples demonstrate that speed and velocity tuning cannot be established by measuring the receptive field within a two-dimensional subspace. This paper represents the first attempt to examine three-dimensional spectral selectivity, which allows direct assessment of speed and velocity tuning.

Functional implications of partial ring structures in the frequency domain

An MT neuron that forms a ring in the optimal three-dimensional velocity plane (Simoncelli and Heeger, 1998) will be tuned for one particular velocity, depending on the slant and tilt of the optimal plane. However, such a neuron will also respond to static stimuli oriented parallel to the optimal direction of motion. Our results show that most area MT neurons avoid this problem. These neurons form a partial ring in the optimal velocity plane, systematically avoiding the region around zero temporal frequency. They respond to moving patterns at the optimal velocity, but they do not tend to respond to static patterns that are oriented parallel to the optimal direction. This scheme provides a representation of image motion that is less ambiguous than that proposed by the Simoncelli and Heeger (1998) model. Consistent with this, Albright (1984) reported that although some MT neurons do respond to a static bar oriented parallel to the optimal direction, these responses are much less vigorous than those elicited by moving stimuli. Our study provides a clear explanation for this phenomenon, and confirms that area MT is optimized to process moving patterns.

It is currently unclear how area MT neurons develop their very precise receptive fields and why most of them are insensitive to low temporal frequencies. Each MT neuron receives input from a specific population of direction-selective V1 neurons (Movshon and Newsome, 1996). Direction-selective neurons in V1 are almost exclusively bandpass for temporal frequency (Hawken et al., 1996) and so do not respond to static stimuli. The fact that most MT neurons do not respond to zero temporal frequency could therefore merely reflect a bias that already exists in the V1 neurons that project to MT. This bias could be genetic, or it could be caused by the learning algorithm that governs the development of receptive fields in MT. If this feature is the product of a learning rule, this finding could provide a new constraint on how corticocortical circuits are optimized to represent information during natural vision.

References

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

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

Albright TD (1984) Direction and orientation selectivity of neurons in visual area MT of the macaque. J Neurophysiol 52:1106–1130.

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

Born RT, Bradley DC (2005) Structure and function of visual area MT. Annu Rev Neurosci 28:157–189.

Bradley DC, Goyal MS (2008) Velocity computation in the primate visual system. Nat Rev Neurosci 9:686–695.

Britten KH, Shadlen MN, Newsome WT, Movshon JA (1993) Responses of neurons in macaque MT to stochastic motion signals. Vis Neurosci 10:1157–1169.

Carandini M, Heeger DJ, Movshon JA (1997) Linearity and normalization in simple cells of the macaque primary visual cortex. J Neurosci 17:8621–8644.

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

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

David SV, Mesgarani N, Shamma SA (2007) Estimating sparse spectro-temporal receptive fields with natural stimuli. Network 18:191–212.

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1–47.

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

Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 29:1189–1232.
Hawken MJ, Shapley RM, Grosof DH (1996) Temporal-frequency selectivity in monkey visual cortex. Vis Neurosci 13:477–492.
Heeger DJ (1992a) Half-squaring in responses of cat striate cells. Vis Neurosci 9:427–443.
Heeger DJ (1992b) Normalization of cell responses in cat striate cortex. Vis Neurosci 9:181–197.
Jones JP, Palmer LA (1987) An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. J Neurophysiol 58:1233–1258.
Kaplan E, Purpura K, Shapley RM (1987) Contrast affects the transmission of visual information through the mammalian lateral geniculate nucleus. J Physiol 391:267–288.
Lewen GD, Bialek W, de Ruyter van Steveninck RR (2001) Neural coding of naturalistic motion stimuli. Network 12:317–329.
Livingstone MS, Pack CC, Born RT (2001) Two-dimensional substructure of MT receptive fields. Neuron 30:781–793.
Majaj NJ, Carandini M, Movshon JA (2007) Motion integration by neurons in macaque MT is local, not global. J Neurosci 27:366–370.
Mante V, Frazor RA, Bonin V, Geisler WS, Carandini M (2005) Independence of luminance and contrast in natural scenes and in the early visual system. Nat Neurosci 8:1690–1697.
Mante V, Bonin V, Carandini M (2008) Functional mechanisms shaping lateral geniculate responses to artificial and natural stimuli. Neuron 58:625–638.
Mazer JA, Gallant JL (2003) Goal related activity in area V4 during free viewing visual search: evidence for a ventral stream salience map. Neuron 40:1241–1250.
Movshon JA, Newsome WT (1996) Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. J Neurosci 16:7733–7741.
Movshon JA, Adelson EH, Gizzi MS, Newsome WT (1985) The analysis of moving visual patterns. In: Pattern recognition mechanisms (Chagas C, Gattass R, Gross C, eds), p 117. New York: Springer.
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.
Okamoto H, Kawakami S, Saito H, Hida E, Odajima K, Tamanoi D, Ohno H (1999) MT neurons in the macaque exhibited two types of bimodal direction tuning as predicted by a model for visual motion detection. Vision Res 39:3465–3479.
Olmos A, Kingdom FAA (2004) A biologically inspired algorithm for the recovery of shading and reflectance images. Perception 33:1463–1473.
Pack CC, Born RT (2001) Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. Nature 409:1040–1042.
Perrone JA, Thiele A (2001) Speed skills: measuring the visual speed analyzing properties of primate MT neurons. Nat Neurosci 4:526–532.
Priebe NJ, Cassanello CR, Lisberger SG (2003) The neural representation of speed in macaque area MT/V5. J Neurosci 23:5650–5661.
Rodman HR, Albright TD (1987) Coding of visual stimulus velocity in area MT of the macaque. Vision Res 27:2035–2048.
Rust NC, Movshon JA (2005) In praise of artifice. Nat Neurosci 8: 1647–1650.
Rust NC, Mante V, Simoncelli EP, Movshon JA (2006) How MT cells analyze the motion of visual patterns. Nat Neurosci 9:1421–1431.
Schrater PR, Simoncelli EP (1998) Local velocity representation: evidence from motion adaptation. Vision Res 38:3899–3912.
Schrater PR, Knill DC, Simoncelli EP (2000) Mechanisms of visual motion detection. Nat Neurosci 3:64–68.
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.
Simoncelli EP, Heeger DJ (1998) A model of neuronal responses in visual area MT. Vision Res 38:743–761.
Smith MA, Majaj NJ, Movshon JA (2005) Dynamics of motion signaling by neurons in macaque area MT. Nat Neurosci 8:220–228.
Snowden RJ, Treue S, Erickson RG, Andersen RA (1991) The response of area MT and V1 neurons to transparent motion. J Neurosci 11:2768–2785.
Stanley GB (2008) Au naturel. Neuron 58:467–469.
Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of non-linear auditory neurons obtained using natural sounds. J Neurosci 20:2315–2331.
Watson AB, Ahumada AJ Jr (1985) Model of human visual-motion sensing. J Opt Soc Am A 2:322–341.
Willmore BD, Prenger RJ, Gallant JL (2010) Neural representation of natural images in visual area V2. J Neurosci 30:2102–2114.
Wu MC, David SV, Gallant JL (2006) Complete functional characterization of sensory neurons by system identification. Annu Rev Neurosci 29:477–505.