Reconstruction of Natural Scenes from Ensemble Responses in the Lateral Geniculate Nucleus

Garrett B. Stanley, Fei F. Li, and Yang Dan

Department of Molecular and Cell Biology, Division of Neurobiology, University of California, Berkeley, California 94720

A major challenge in studying sensory processing is to understand the meaning of the neural messages encoded in the spiking activity of neurons. From the recorded responses in a sensory circuit, what information can we extract about the outside world? Here we used a linear decoding technique to reconstruct spatiotemporal visual inputs from ensemble responses in the lateral geniculate nucleus (LGN) of the cat. From the activity of 177 cells, we have reconstructed natural scenes with recognizable moving objects. The quality of reconstruction depends on the number of cells. For each point in space, the quality of reconstruction begins to saturate at six to eight pairs of on and off cells, approaching the estimated coverage factor in the LGN of the cat. Thus, complex visual inputs can be reconstructed with a simple decoding algorithm, and these analyses provide a basis for understanding ensemble coding in the early visual pathway.

Key words: LGN; reconstruction; natural scenes; ensemble responses; cat; visual system

The foundation of our current knowledge of sensory processing was established by characterizing neuronal responses to various sensory stimuli (Adrian, 1926; Hartline, 1938; Barlow, 1953; Kuffler, 1953; Hubel and Wiesel, 1962). In this paradigm, sensory neurons are studied by measuring their receptive fields and tuning properties, which can in turn be used to predict the responses of neurons to arbitrary sensory inputs (Brodie et al., 1978; Dan et al., 1996). A critical test of our understanding of sensory coding, however, is to take an opposite approach: to reconstruct sensory inputs from recorded neuronal responses. The decoding approach can provide an objective assessment of what and how much information is available in the neuronal responses. Although the function of the brain is not necessarily to reconstruct sensory inputs faithfully, these studies may lead to new insights into the functions of neuronal circuits in sensory processing (Rieke et al., 1997).

The decoding approach has been used to study several sensory systems (Bialek et al., 1991; Theunissen and Miller, 1991; Rieke et al., 1993, 1997; Roddey and Jacobs, 1996; Warland et al., 1997; Dan et al., 1998). Most of these studies aimed to reconstruct temporal signals from the response of a single neuron (Bialek et al., 1991; Rieke et al., 1993, 1995; Roddey and Jacobs, 1996) or a small number of neurons (Warland et al., 1997). An important challenge in understanding the mammalian visual system is to reconstruct more complex, spatiotemporal inputs from the responses of a large number of neurons. Here we have developed an input reconstruction technique to decode information from ensemble activity in the lateral geniculate nucleus (LGN). Rather than analyzing a single neuron at a time, this technique takes into consideration the relationship between neurons within the population. This is crucial for understanding how sensory information is coded, in a distributed manner, in the activity of large neuronal circuits.

As a step toward understanding visual coding in the natural environment, we used natural scenes as visual stimuli in the current study. Although simple artificial stimuli are very useful in characterizing response properties of sensory neurons, the task of the brain is primarily to process information in the natural environment. Natural scenes are known to have characteristic statistical properties (see, for example, Field, 1987; Dong and Atick, 1995); the importance of using such stimuli for studying the visual system has been well demonstrated (Creutzfeldt and Nothdurft, 1978; Olshausen and Field, 1996; Bell and Sejnowski, 1997; Rieke et al., 1997; Gallant et al., 1998). Some studies have further suggested that the nervous system may be specifically adapted for efficient processing of natural stimuli (Barlow, 1961; Laughlin, 1981; Atick, 1992; Rieke et al., 1995, 1997; Dan et al., 1996). Thus, it is important to investigate how natural signals are coded in the activity of visual circuits.

In this study, we reconstructed spatiotemporal natural scenes (movies) from recorded responses in the LGN. The reconstruction algorithm takes into consideration not only the response properties of the neurons, but also the statistics of natural scenes (Bialek and Rieke, 1992). From the responses of 177 cells, we were able to reconstruct time-varying natural scenes with recognizable moving objects. Between 3 and 16 Hz, the signal-to-error ratio of the linear reconstruction reaches the theoretical limit set by noise in the neuronal responses. As expected, the quality of reconstruction depends on the number of cells. For each pixel in the visual scene, the quality begins to saturate at six to eight pairs of on and off cells, approaching the estimated coverage factor in the LGN of the cat. Thus, we have provided a first demonstration that spatiotemporal natural scenes can be reconstructed from the ensemble responses of visual neurons. The results from these studies also provide an explicit test of the linear model of LGN coding and an assessment of the number of neural channels required for coding spatiotemporal natural scenes.

MATERIALS AND METHODS

Physiological preparation

Adult cats ranging in weight from 2 to 3 kg were used in all the experiments. The animals were initially anesthetized with isofluorane (3%, with oxygen) followed by sodium pentothal (10 mg/kg, i.v., supplemented as needed). A local anesthetic (lidocaine) was injected before all incisions. Anesthesia was maintained for the duration of the experiment with sodium pentothal at a dosage of 6 mg/hr.

A tracheostomy was performed for artificial ventilation. The cat was moved to a Horsley–Clarke stereotaxic frame. A craniotomy (~0.5 cm) was performed over the LGN, and the underlying dura was removed. The hole was filled with 3% agar in physiological saline to improve the stability of the recordings.

Pupils were dilated with a topical application of 1% atropine sulfate, and the nictitating membranes were retracted with 10% phenylephrine. The animal was paralyzed with Norcuron (0.2 mg/kg/hr, i.v.) and artificially ventilated. Ventilation was adjusted so that the end-expiratory CO_2 was ~3.5%. Core body temperature was monitored and maintained at 37°C. The electrocardiogram and electroencephalogram were also monitored continuously. Eyes were refracted, fitted with appropriate contact lenses, and focused on a tangent screen. Eye positions were stabilized mechanically by gluing the sclerae to metal posts attached to the stereotaxic apparatus. All experiments were performed as approved by the Animal Care and Use Committee, University of California at Berkeley.

Electrophysiological recording and visual stimulation

Neighboring geniculate cells were recorded with a multielectrode array (Eckhorn and Thomas, 1993). The array allows seven fiber electrodes to be positioned independently with a vertical accuracy of 1 μm. We used a glass guide tube to restrict the lateral scattering of the electrodes in the array. The inner diameter at the tip of the guide tube was <400 μm. All recordings were made in layer A or A1 of the LGN.

Recorded signals were amplified, filtered, and passed to a personal computer (PC) running Datawave (Broomfield, CO) Discovery software. The system accepts inputs from up to eight single electrodes. Up to eight different waveforms can be discriminated on a single electrode, but two or three is a more realistic limit. The waveforms were saved on disk. Spike isolation was based on cluster analysis of waveforms, and the presence of a refractory period, which is reflected in the shape of autocorrelations.

Visual stimuli were created with a PC containing an AT-Vista graphics card (Truevision, Indianapolis, IN) at a frame rate of 128 Hz. The movies were digitized segments of black and white video recordings. They covered a wide range of natural scenes (street, indoor, woods). The movie signals (spatiotemporal natural scenes) were updated every four frames, resulting in an effective frequency of 32 Hz. Because natural scenes have a high degree of temporal correlation, most of the power in the stimuli is captured with this sampling rate (see Fig. 3a). Each frame contained 64 × 64 pixels with a spatial resolution of 0.2°/pixel. Each movie was 16 sec long. To generate enough data to obtain reliable estimates of the reverse filters, we showed eight different movies, each repeated eight times. The root-mean-square contrast of the movies (the square root of the mean of the squared difference between the intensity of each pixel at each frame and the mean intensity of the whole movie) was 30.4%. The cells used for reconstruction were screened based on the reliability of their responses over multiple repeats of the same movie. White-noise stimuli were also generated by the same system. Spatially, the white-noise stimuli were made up of 16 × 16 pixels. The pixel sizes were adjusted to map receptive fields with a reasonable level of detail (0.2–0.4°, at ~10° eccentricity). For every frame of the stimulus, the pixels were either black or white (100% contrast), according to a binary pseudorandom m sequence (Sutter, 1987; Reid et al., 1997). The complete m sequence consisted of 32,767 frames, updated at 128 Hz.

Data analyses

Linear input reconstruction. The spike trains of the neurons were binned according to the frame rate of the stimulus (32 Hz for movies, 128 Hz for white noise) and converted to firing rate signals. We use the following notation: $[r_1 r_2 \ldots r_{n_r}]$ are the responses of the neurons and $[s_1 s_2 \ldots s_{n_s}]$ are the stimuli at different pixels, scaled between −1 and 1. The term r_i is a time-dependent signal representing the firing rate of the ith neuron, and s_j is a time-dependent signal representing the luminance at the jth pixel. To reconstruct spatiotemporal visual inputs from the responses of mul-

tiple neurons, we derived linear reverse filters that minimize the mean-square error of the reconstruction. The multi-input, multi-output reverse filter matrix **h** can be expressed as:

$$
\begin{bmatrix} \mathbf{h}_{11} & \cdots & \mathbf{h}_{1n_s} \\ \vdots & \ddots & \vdots \\ \mathbf{h}_{n_r 1} & \cdots & \mathbf{h}_{n_r n_s} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_{r_1 r_1} & \cdots & \mathbf{P}_{r_1 r_{n_r}} \\ \vdots & \ddots & \vdots \\ \mathbf{P}_{r_{n_r} r_1} & \cdots & \mathbf{P}_{r_{n_r} r_{n_r}} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{p}_{r_1 s_1} & \cdots & \mathbf{p}_{r_1 s_{n_s}} \\ \vdots & \ddots & \vdots \\ \mathbf{p}_{r_{n_r} s_1} & \cdots & \mathbf{p}_{r_{n_r} s_{n_s}} \end{bmatrix} \quad (1)
$$

Note that each term in Equation 1 represents a submatrix. The column vector $\mathbf{h}_{ij} = [h_{ij}[-(L-1)] \ldots h_{ij}[L-1]]^T$ is the time-domain representation of the linear reverse filter from the ith neuron to the jth pixel, where L is the filter length (set to 50 frames in this study). $\mathbf{P}_{r_m r_n}$ is a Toeplitz matrix that represents the cross-covariance between the response of the mth neuron and the response of the nth neuron. The element at the kth row and lth column of $\mathbf{P}_{r_m r_n}$ represents the cross-covariance at a delay of $k - l$ frames. This term ensures that the correlation between different cells is taken into consideration in the computation of reverse filters, a crucial feature for studying ensemble decoding. $\mathbf{p}_{r_i s_j} = [p_{r_i s_j}[-(L-1)] \ldots p_{r_i s_j}[L-1]]^T$ is a column vector that represents the cross-covariance between the response of the ith neuron and the stimulus at the jth pixel. Both $\mathbf{P}_{rr}$ and $\mathbf{p}_{rs}$ were estimated by averaging over the modeling data set, and **h** was then computed directly using Equation 1. Note that this computation is performed entirely in the time domain. For natural scenes, the modeling data set consisted of 63 movie clips (each clip was 16 sec long, so the total duration was 1008 sec). For white noise, the modeling data set consisted of 240 sec of data. Finally, the stimulus signal at the jth pixel, s_j, was reconstructed by convolving the reverse filters with the responses of the corresponding cells and summing over all the cells used for the reconstruction:

$$
\hat{s}_j[t] = \sum_{i=1}^{n_r} \sum_{u=-(L-1)}^{L-1} h_{ij}[u] r_i[t-u] \quad (2)
$$

where $\hat{s}_j[t]$ is the reconstructed stimulus at the jth pixel. The responses used in this convolution (16 sec for natural scenes, 15 sec for white noise) were not used for estimating **h**.

Signal-to-error ratio of the reconstruction. Signal-to-error ratio (SER) was computed in the frequency domain. The error of the reconstruction is defined as the difference between the reconstructed and the actual inputs. The SER is defined as the ratio between the power spectral density of the actual input and that of the error as a function of temporal frequency. The total SER shown in Figure 4, *c* and *e,* is defined as the ratio between the total power of the actual input (integrated between 0.125 and 16 Hz) and the total power of the error. To compute the control SER (see Fig. 3b, *dashed line*), we used the same cells as those used for the real SER, but shuffled the neuronal responses to different movies, so that the visual stimuli and the responses were mismatched. The reconstruction was generated from this randomized data set in which the causal relationship between the stimuli and the responses was eliminated. The control SER was then computed from this reconstruction, which provides a baseline against which the significance of the real SER can be judged.

Theoretical signal-to-error ratio. Assuming a perfect linear relationship between the stimuli and the responses, the theoretical power spectrum of the error can be expressed in the following matrix form:

$$
[\Phi_{ee}(\omega)] = [\Phi_{ss}(\omega)] - [\Phi_{ss}(\omega)][K(\omega)]^*[\Phi_{rr}(\omega)]^{-1}[K(\omega)][\Phi_{ss}(\omega)] \quad (3)
$$

where ω is the temporal frequency, and [ ] denotes matrix. The entry at the ith row and the jth column of the matrix $[\Phi_{ee}(\omega)]$ represents the cross-spectrum between the error at the ith pixel and that at the jth pixel. $[\Phi_{ss}(\omega)]$ represents the cross-spectra between the actual stimuli at different pixels. $[K(\omega)]$ represents the temporal Fourier transform of the linear receptive fields (mapped with white noise) of the neurons used in the reconstruction. $[\Phi_{rr}(\omega)] = [K(\omega)][\Phi_{ss}(\omega)][K(\omega)]^* + [\Phi_{nn}(\omega)]$ represents the theoretical cross-spectra between the responses of different cells. $[\Phi_{nn}(\omega)]$ represents the cross-spectra of the noise in the responses of different cells, and * denotes the complex conjugate transpose. Noise in the response is defined as the difference between the firing rate of each individual repeat and the average PSTH from multiple repeats of the same movie. Because of the rectification of the actual responses, each pair of on and off cells used in the real reconstruction was approximated as a single linear filter in the theoretical computation. For each pixel of the movie, the theoretical error spectrum was therefore computed from

half of the cells used in the real reconstruction. Finally, the theoretical SER was computed as the ratio between the power spectrum of the actual stimulus and the theoretical spectrum of the error, $\Phi_{ee}(\omega)$, as defined above.

Estimation of coverage factor. The coverage factor is defined as the average number of cells whose receptive fields cover a single point in the visual space. Previous studies have shown that the coverage factor for X cells in the cat retina is 7–10 (Peichl and Wässle, 1979). This coverage factor represents the coverage by the centers of the receptive fields. In our analysis, some cells were also used to reconstruct signals in the surround of their receptive fields. In the great majority of cases, the pixel being reconstructed was within an area twice that of the receptive field center. Therefore, we scaled the estimate of Peichl and Wässle (1979) by a factor of 2, resulting in a coverage factor of 14–20 for the retina. In the cat, the number of X cells in the retina is ~75,000 (Wässle and Boycott, 1991), and the number of geniculate X cells representing each eye is ~120,000 (Peters and Payne, 1993). We estimated the coverage factor for LGN X cells by multiplying the coverage factor in the retina (14–20) by the ratio between geniculate and retinal X cells (~1.5). The coverage factor for LGN X cells is therefore ~20–30 cells.

RESULTS

Multiple cells in the LGN of anesthetized cats were recorded simultaneously with multielectrodes (Eckhorn and Thomas, 1993). The receptive fields of these cells were mapped with spatiotemporal white-noise stimuli and the reverse correlation method (Sutter, 1987; Reid et al., 1997). Only X cells were selected for further studies because they are presumably involved in processing the spatial details of visual scenes (Wässle and Boycott, 1991) and they have relatively linear response properties (So and Shapley, 1981). We recorded the responses of the cells to multiple repeats of eight short movies, and these data were used for subsequent analyses. The geniculate cells were well driven by the movie stimuli, as indicated by their mean firing rates, which were higher during movie presentation (11.7 spikes/sec; $n = 57$ cells) than in the absence of visual stimuli (6.8 spikes/sec; $n = 41$ cells).

A multi-input, multi-output linear decoding technique was implemented to reconstruct the spatiotemporal visual inputs. Figure 1a shows the receptive fields of eight neurons recorded simultaneously. Outlined in white are the four pixels at which the movie signals were reconstructed. As a first step in the reconstruction, a set of linear reverse filters from the responses of all the neurons to these pixels was computed (see Data Analyses). These linear reverse filters (Fig. 1b) are optimal in the sense that they minimize the mean-square error of the reconstructed luminance signals. They depend on not only the response property of each cell, but also the correlation between cells, as well as the statistics of natural scenes (Warland et al., 1997; Bialek and Rieke, 1992). The visual signal at each pixel was reconstructed by convolving the spike train of each cell (Fig. 1c) with the corresponding reverse filter and summing the results from all eight cells. Figure 1d shows the actual (black) and the reconstructed (magenta) luminance signals as functions of time. The low frequency, slow varying features of the stimuli were well captured by the reconstruction. Consistent with the known temporal properties of X cells, which respond poorly to stimuli at high frequencies, some of the quick transients were not well reconstructed. For these four pixels, the mean correlation coefficient between the reconstructed and the actual signals was 0.60 ± 0.04.

In addition to reconstructing the temporal features, our goal was also to capture the spatial patterns of natural scenes. To reconstruct movie scenes in an area large enough to contain recognizable objects, we pooled the responses of 177 cells (89 on, 88 off) recorded in 11 experiments using the same visual stimuli.
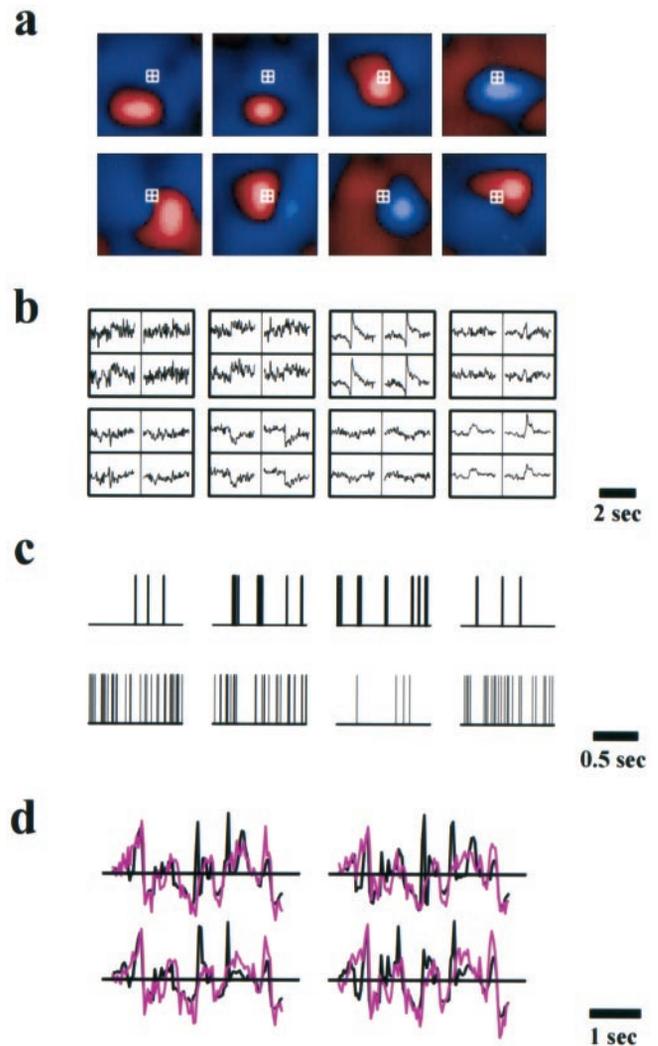
Figure 1. The procedure for reconstructing visual stimuli from the responses of multiple neurons. *a*, Receptive fields of eight neurons recorded simultaneously with multielectrodes. These receptive fields were mapped with white-noise stimuli and the reverse correlation method (Sutter, 1987; Reid et al., 1997). *Red,* On responses. *Blue,* Off responses. The brightest colors correspond to the strongest responses. The area shown is 3.6 × 3.6°. The responses of these cells were used to reconstruct visual inputs at the four pixels (0.2°/pixel) outlined with the *white squares*. *b,* Linear filters for input reconstruction. The eight blocks correspond to the eight cells shown in *a*. Shown in each block are the four filters from that cell to the four pixels outlined in *a*. They represent the linear estimates of the input signals at these pixels immediately preceding and following a spike of that cell. Each filter is 3.1-sec-long, with 1.55 sec before and 1.55 sec after the spike. *c,* Spike trains of the eight neurons in response to movie stimuli. *d,* The actual (*black*) and the reconstructed (*magenta*) movie signals at the four pixels outlined in *a*. Unlike white noise, natural visual signals exhibit more low-frequency, slow variations than high-frequency, fast variations. Such temporal features are well captured by the reconstruction.

Figure 2a shows the receptive fields of these cells, distributed over an area of 6.4 × 6.4°. For each pixel of the movie, we used the responses of 7–20 cells (average 14, with approximately equal numbers of on and off cells) whose receptive fields, including both center and surround, covered that pixel. The reconstruction was carried out in the same manner as illustrated in Figure 1. The results at all pixels were then combined to obtain a spatiotemporal signal. Figure 2b shows three examples of the actual and the reconstructed images in consecutive movie frames. Note that
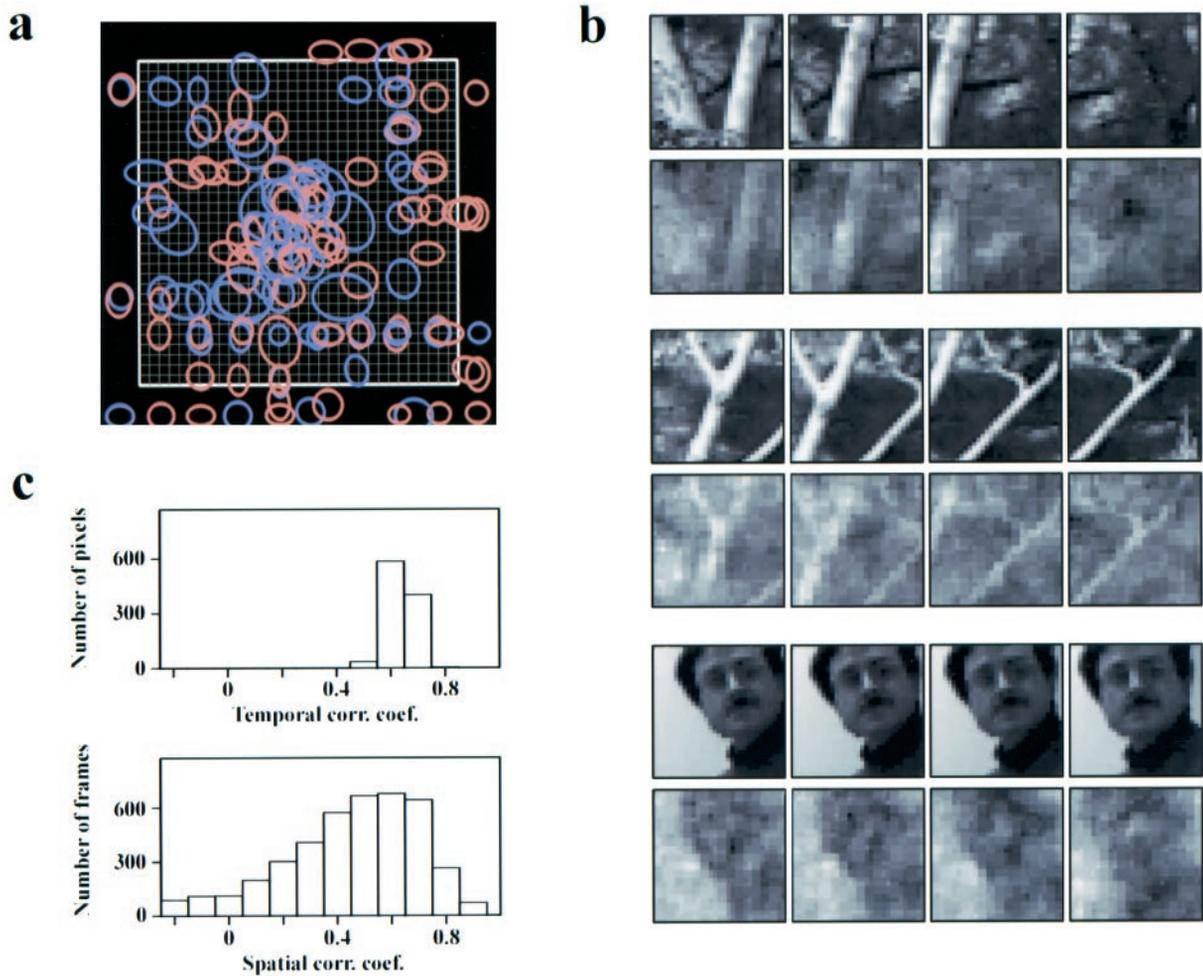
Figure 2. Reconstruction of natural scenes from the responses of a population of neurons. *a,* Receptive fields of 177 cells used in the reconstruction. Each receptive field was fitted with a two-dimensional Gaussian function. Each ellipse represents the contour at one SD from the center of the Gaussian fit. Note that the actual receptive fields (including surround) are considerably larger than these ellipses. *Red,* On center. *Blue,* Off center. An area of 32 × 32 pixels (0.2°/pixel) where movie signals were reconstructed is outlined in *white.* The *grid* inside the *white square* delineates the pixels. *b,* Comparison between the actual and the reconstructed images in an area of 6.4 × 6.4° (*a, white square*). Each panel shows four consecutive frames (interframe interval, 31.1 msec) of the actual (*top*) and the reconstructed (*bottom*) movies. *Top panel,* Scenes in the woods, with two trunks of trees as the most prominent objects. *Middle panel,* Scenes in the woods, with smaller tree branches. *Bottom panel,* A face at slightly different displacements on the screen. *c,* Quantitative comparison between the reconstructed and the actual movie signals. *Top,* Histogram of temporal correlation coefficients between the actual and the reconstructed signals (both as functions of time) at each pixel. The histogram was generated from 1024 (32 × 32) pixels in the *white square.* *Bottom,* Histogram of spatial correlation coefficients between the actual and the reconstructed signals (both as functions of spatial position) at each frame. The histogram was generated from 4096 frames (512 frames per movie; 8 movies).

moving objects (tree trunks, branches, and a human face) are discernible in the reconstructed movies. To evaluate the reconstruction quantitatively, we computed the correlation coefficients between the reconstructed and the actual signals along two dimensions: as functions of time at each pixel and as functions of spatial position at each frame (Fig. 2c). Both the spatial and temporal correlation coefficients peaked at 0.6–0.7. The spatial correlation coefficients are more dispersed than the temporal correlation coefficients, which may be caused by the fact that different pixels were reconstructed from different sets of cells. Such inhomogeneity may introduce an additional source of variability along the spatial dimension.

To further evaluate the linear decoding technique used in the current study, we performed spectral analyses of the reconstruction. First, we computed the temporal power spectra of the actual (Fig. 3a, *thick line*) and the reconstructed (*thin line*) inputs. They closely resemble each other, both exhibiting an ~1/f^2 profile that

is characteristic of natural scenes (Dong and Atick, 1995). Second, we computed the power spectrum of the error, which is the difference between the actual and the reconstructed signals. The SER of the reconstruction was plotted as a function of temporal frequency (Fig. 3b, *solid line*). To evaluate the significance of the SER, we computed a control SER (*dashed line*), which is the SER of the reconstruction generated from randomly matched visual stimuli and neuronal responses. Between 0.125 and 16 Hz, the real SER (*solid line*) is significantly higher than the control SER, indicating that meaningful visual information is extracted at all of these frequencies. Finally, previous studies have shown that geniculate X cells can be modeled as spatiotemporal linear filters (Derrington and Fuchs, 1979; Dan et al., 1996) with additive noise (Sestokas and Lehmkuhle, 1988). Using this model, we estimated the theoretical SER of the reconstruction (*dotted line*) based on its analytical relationship with the noise in the responses, assuming perfect linear encoding (see Data Analyses).

Figure 3. Evaluation of reconstruction using spectral analyses. Because natural scenes were presented at 32 Hz, these analyses were performed at up to 16 Hz. *a*, Temporal power spectra of the actual and the reconstructed inputs. Both were averaged from 192 pixels near the center of the screen. For each pixel, the input was reconstructed from the same cells used in Figure 2. *b*, Comparison between the SER of the reconstruction and the theoretical SER estimated based on the noise in the neuronal responses. Above 3 Hz these two curves are not significantly different. The control SER represents the SER of the reconstruction if there is no causal relationship between the visual stimuli and the neuronal responses. It provides a baseline against which the significance of the real SER can be judged. All three curves were averaged from the same 192 pixels used in *a*. *Vertical lines* represent SEs. For the control SER, the error bars are smaller than the thickness of the line.
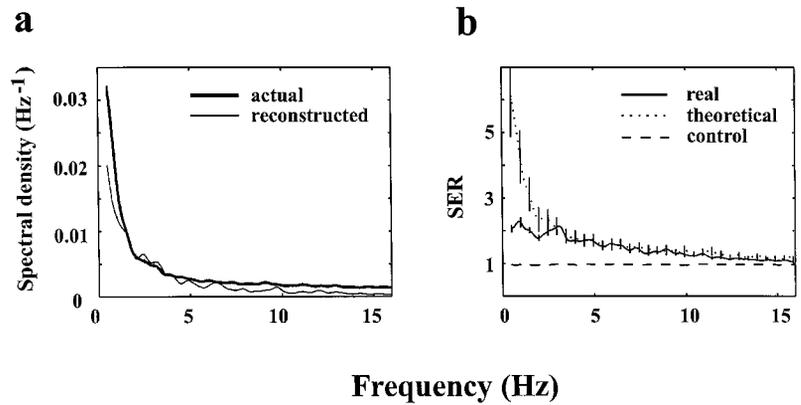
Between 3 and 16 Hz, the SER of the reconstruction (*solid line*) closely matches the theoretical SER, consistent with the notion that geniculate cells function as linear spatiotemporal filters. The difference between these curves below 3 Hz may be due to nonlinearities, such as those caused by light adaptation and contrast gain control (Shapley and Enroth-Cugell, 1985) or nonstationarities of the neuronal response properties.

An important issue is how the density of cells affects the quality of reconstruction. We compared the spatial distribution of receptive fields with the quality of reconstruction. Figure 4*a* shows a map of the receptive fields of cells used in Figure 2 and the correlation coefficient between the actual and the reconstructed stimuli at each pixel. There is a close correspondence between the density of cells and the quality of reconstruction at each area. To investigate this relationship more quantitatively, we systematically varied the number of cells used for reconstructing a single pixel, from 2 to 20 cells (1–10 on/off pairs). Both the total SER of the reconstruction and the correlation coefficient between the reconstructed and the actual signals improve with increasing numbers of cells, but they begin to saturate at 12–16 cells (Fig. 4*b,c*). Natural scenes are known to have a high degree of spatiotemporal correlation (Field, 1987; Dong and Atick, 1995). Such correlation reduces the amount of information in the visual input, which may explain the saturation at a relatively small number of cells. To test whether the saturation is specific to natural scenes, we reconstructed white-noise signals with no spatiotemporal correlation and analyzed the dependence of the reconstruction on the number of cells. Both the correlation coefficient and the SER for white noise are much lower than those for natural scenes (Fig. 4*d,e*). This difference in reconstruction quality is consistent with the notion that white-noise stimuli contain more information than natural scenes. More importantly, the quality of reconstruction does not exhibit saturation below 20 cells. This indicates that the saturation in this range is not a general property of visual coding in the LGN. Rather, it is related to the statistics of natural scenes.

DISCUSSION
The current study is motivated by the following fundamental question: when neurons fire action potentials, what do they tell the brain about the visual world? As a first step to address this question, we reconstructed spatiotemporal natural scenes from ensemble responses in the LGN. Responses of visual neurons to natural images have been studied in the past. Creutzfeldt and Nothdurft (1978) studied the spatial patterns of the responses by moving natural images over the receptive fields of single neurons and recording the responses at corresponding positions. They generated "neural transforms" of static natural scenes that re-

vealed important features of the neurons in coding visual signals. In contrast to this "forward" approach to studying neural coding, we have taken a "reverse" approach, which is to decode information from the neural responses. Given the known properties of X cells in the LGN, significant information can in principle be extracted from their responses with a linear technique. Here we have presented the first direct demonstration that spatiotemporal natural scenes can be reconstructed from experimentally recorded spike trains. The results from the linear technique also provide a benchmark for future decoding studies with nonlinear techniques.

In this study, we extracted information from the responses of a population of neurons. The reconstruction filters not only reflect the response properties of individual neurons, but also take into consideration the correlation between neighboring cells. This is crucial for decoding information from ensemble responses. Not surprisingly, an important factor affecting the quality of reconstruction is the density of cells (Fig. 4). In Figure 2, visual signals within an area of $6.4 \times 6.4°$ (1024 pixels) were reconstructed from the responses of 177 cells, corresponding to an average tiling of 9×9 on/off pairs over a 32×32 array of pixels. As shown in Figure 4*a*, some areas of the scenes were covered with lower densities of cells, resulting in lower correlation coefficients. A better coverage of these areas could potentially improve the reconstruction. For natural scenes, the quality of reconstruction begins to saturate at 12–16 cells, which appears to be related to the spatiotemporal correlation in the inputs. A previous study using a similar technique has shown that the saturation of the reconstruction quality occurs at approximately three pairs of on/off retinal ganglion cells in the salamander (Warland et al., 1997), which is significantly lower than the number that we have observed. This discrepancy may be caused by the difference in the visual inputs. In the earlier study the input was full-field white noise with no spatial variation, whereas in our study the natural scenes contain considerable spatial variation. The more complex natural input ensemble presumably contains more information that is carried by a larger number of cells. On the other hand, compared to spatiotemporal white noise, natural scenes contain more spatiotemporal correlation and therefore less information. This results in the difference in the saturation for white noise and for natural scenes (Fig. 4). Based on previous anatomical studies in the retina and the LGN of the cat (Peichl and Wässle, 1979; Wässle and Boycott, 1991; Peters and Payne, 1993), we estimated that every point in visual space is covered by the receptive fields of 20–30 geniculate X cells (see Data Analyses). The density of cells required for optimal reconstruction of natural scenes ap-

Natural Scenes

White Noise
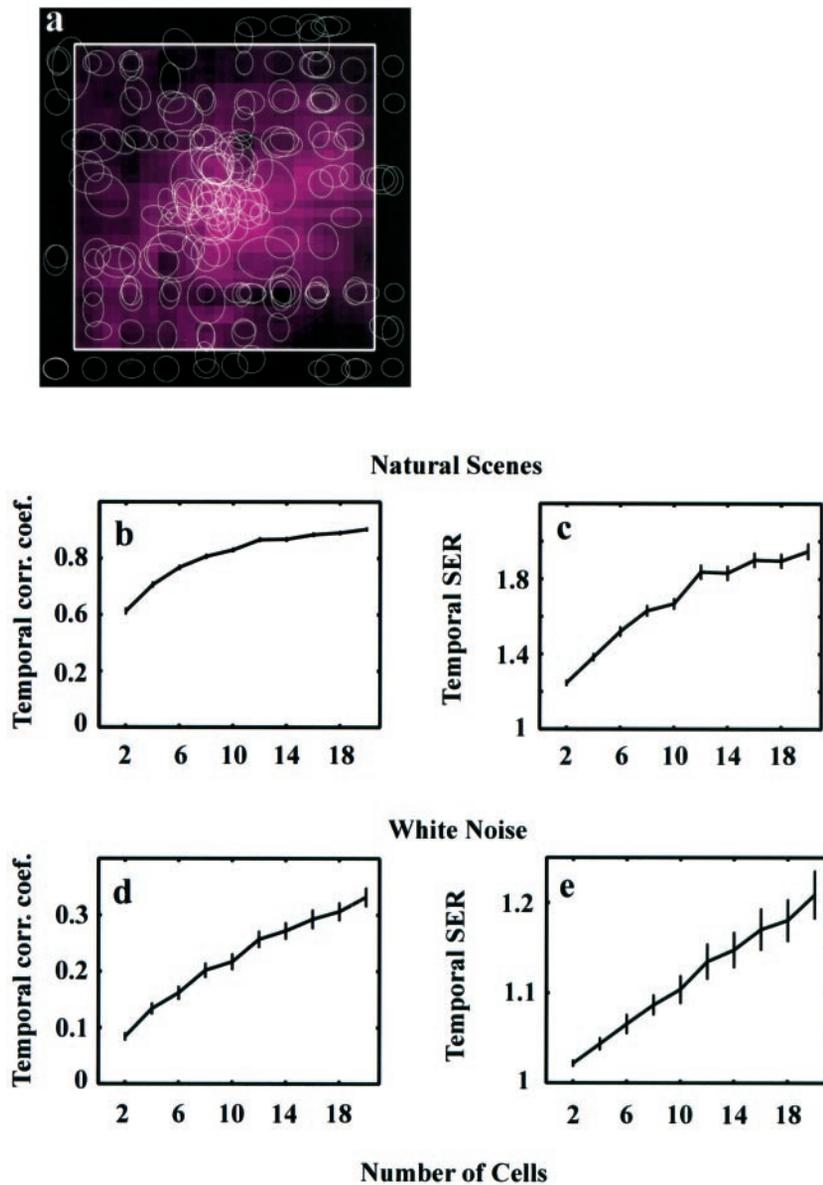
Number of Cells

Figure 4. Dependence of the quality of reconstruction on the number of cells. *a*, Spatial distribution of cell density and reconstruction quality from the results shown in Figure 2. *Ellipses* represent centers of receptive fields. On and off cells are represented by the same color. Correlation coefficient between the actual and the reconstructed stimuli (minimum, 0.52; maximum, 0.79) is indicated by the brightness at each pixel. Note that areas covered by higher densities of cells have higher correlation coefficients between the reconstructed and the actual inputs. *b*, The average temporal correlation coefficient between the actual and the reconstructed natural scenes versus the number of cells used for reconstruction. *c*, The temporal total SER of the reconstruction (natural scenes) of each pixel versus the number of cells used for that pixel. Total SER was defined as the ratio between the total power of the actual input (integrated between 0.125 and 16 Hz) and the total power of the error. In this analysis, we always used equal numbers of on and off cells. Each point represents the mean from multiple (160–192) pixels near the center of the screen. The *vertical lines* represent SEs. For both *b* and *c*, we included data from four of the eight different movie clips whose statistics closely matched that for natural scenes described in previous studies (Field, 1987; Dong and Atick, 1995). *d,e,* The same as *b* and *c*, respectively, except the stimulus is spatiotemporal white noise. Here, the white noise was presented at 128 Hz. For calculating the correlation coefficient, both the actual and reconstructed white-noise signals were averaged every four frames to match the sampling rate of natural scenes.

proaches this coverage factor, supporting the notion that the early visual pathway is well adapted for information processing in natural environments.

Several factors contribute to the error in the reconstruction, including noise, nonlinearity, and nonstationarity in the neuronal responses. Assuming linear encoding, we derived the theoretical error of the reconstruction based on noise in the neuronal responses (Fig. 3*b*). The quality of the actual reconstruction reached this theoretical limit between 3 and 16 Hz, suggesting that noise in the responses is the major source of reconstruction error over this frequency range. We would like to emphasize, however, that the definition of noise is directly related to the assumed mechanism of encoding. Here, our conclusion is based on the assumption of rate coding, under which noise in the responses is defined as the difference between the firing rate of each repeat and the average firing rate from multiple repeats of the same movie. The disagreement between the actual and the theoretical reconstruction errors below 3 Hz is likely to reflect deviation of the neuronal responses from the linear model in this frequency range. To further confirm this hypothesis, we predicted the neuronal re-

sponses based on the linear model. The prediction was computed by convolving the visual stimuli and the linear spatiotemporal receptive fields of the cells, followed by a rectification (Dan et al., 1996). The predicted and the actual firing rates were then Fourier transformed and compared in the frequency domain. Significant difference was observed only below 3 Hz (data not shown), supporting the notion that at these low frequencies, the neuronal responses significantly deviate from the linear model. Such deviation is not surprising because certain nonlinearities, such as light adaptation and contrast gain control, are known to occur at slower time scales (Shapley and Enroth-Cugell, 1985). Also, some of the geniculate cells used in our study may be in the bursting firing mode (Sherman and Koch, 1986; Mukherjee and Kaplan, 1995). Cells in the bursting mode are known to exhibit nonlinear responses, which could contribute to the errors in the reconstruction. In addition, there may be nonstationarities in the responses unaccounted for by the model, such as that caused by slow drifts in the general responsiveness of the visual circuit. Future studies incorporating these mechanisms may further improve the reconstruction.

Our current decoding method assumes that all information is coded in the firing rates of neurons. It is optimal only in the sense that it minimizes the mean-square error under the linear constraint. Although this technique proved to be effective in the present study, more complex, nonlinear decoding techniques (de Ruyter van Steveninck and Bialek, 1988; Churchland and Sejnowski, 1992; Abbott, 1994; Warland et al., 1997; Zhang et al., 1998) may further improve the reconstruction from ensemble thalamic responses. Furthermore, recent studies have shown that neighboring geniculate cells exhibit precisely correlated spiking (Alonso et al., 1996). With white-noise stimuli, up to 20% more information can be extracted if correlated spikes are considered separately (Dan et al., 1998). Such correlated spiking may also contribute to coding of natural scenes. Ultimately, the success of any reconstruction algorithm is related to the underlying model of the neural code. The decoding approach therefore provides a critical measure of our understanding of sensory processing in the nervous system.

REFERENCES

Abbott LF (1994) Decoding neuronal firing and modelling neural networks. Q Rev Biophys 27:291–331.

Adrian ED (1926) The impulses produced by sensory nerve endings. Part I. J Physiol (Lond) 61:49–72.

Alonso JM, Usrey WM, Reid RC (1996) Precisely correlated firing in cells of the lateral geniculate nucleus. Nature 383:815–819.

Atick JJ (1992) Could information theory provide an ecological theory of sensory processing? Network: Comput Neural Syst 3:213–251.

Barlow HB (1953) Summation and inhibition in the frog's retina. J Physiol (Lond) 119:69–88.

Barlow HB (1961) Possible principles underlying the transformation of sensory messages. In: Sensory communications (Rosenblith W, ed), pp 217–234. Cambridge: MIT.

Bialek W, Rieke F (1992) Reliability and information transmission in spiking neurons. Trends Neurosci 15:428–434.

Bialek W, Rieke F, de Ruyter Van Steveninck RR, Warland D (1991) Reading a neural code. Science 252:1854–1857.

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

Brodie SE, Knight BW, Ratliff F (1978) The response of the limulus retina to moving stimuli: a prediction by Fourier synthesis. J Gen Physiol 72:129–165.

Churchland PS, Sejnowski TJ (1992) The computational brain. Cambridge, MA: MIT.

Creutzfeldt OD, Nothdurft HC (1978) Representation of complex visual stimuli in the brain. Naturwissenschaften 65:307–318.

Dan Y, Atick JJ, Reid RC (1996) Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. J Neurosci 16:3351–3362.

Dan Y, Alonso J-M, Usrey WM, Reid RC (1998) Coding of visual information by precisely correlated spikes in the LGN. Nat Neurosci 1:501–507.

de Ruyter van Steveninck R, Bialek W (1988) Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. Proc R Soc Lond B Biol Sci 234:379–414.

Derrington AM, Fuchs AF (1979) Spatial and temporal properties of X and Y cells in the cat lateral geniculate nucleus. J Physiol (Lond) 293:347–364.

Dong DW, Atick JJ (1995) Statistics of natural time-varying images. Network: Comput Neural Syst 6:345–358.

Eckhorn R, Thomas U (1993) A new method for the insertion of multiple microprobes into neural and muscular tissue, including fiber electrodes, fine wires, needles and microsensors. J Neurosci Methods 49:175–179.

Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Am A 4:2379–2394.

Gallant JL, Connor CE, Van Essen DC (1998) Neural activity in areas V1, V2, and V4 during free viewing of natural scenes compared to controlled viewing. NeuroReport 9:1673–1678.

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

Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol (Lond) 160:106–154.

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

Laughlin SB (1981) A simple coding procedure enhances a neuron's information capacity. Z Naturforsch 36c:910–912.

Mukherjee P, Kaplan E (1995) Dynamics of neurons in cat lateral geniculate nucleus: in vivo electrophysiology and computational modeling. J Neurophysiol 74:1222–1243.

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

Peichl L, Wässle H (1979) Size, scatter and coverage of ganglion cell receptive field centers in the cat retina. J Physiol (Lond) 291:117–141.

Peters A, Payne BR (1993) Numerical relationships between geniculocortical afferents and pyramidal cell modules in cat primary visual cortex. Cereb Cortex 3:69–78.

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

Rieke F, Warland D, Bialek W (1993) Coding efficiency and information rates in sensory neurons. Europhys Lett 22:151–156.

Rieke F, Bodnar DA, Bialek W (1995) Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. Proc R Soc Lond B Biol Sci 262:259–265.

Rieke F, Warland D, de Ruyter Van Steveninck RR, Bialek W (1997) Spikes: exploring the neural code. Cambridge, MA: MIT.

Roddey JC, Jacobs GA (1996) Information theoretic analysis of dynamical encoding by filiform mechanoreceptors in the cricket cercal system. J Neurophysiol 75:1365–1376.

Sestokas AK, Lehmkuhle S (1988) Response variability of X- and Y-cells in the dorsal lateral geniculate nucleus of the cat. J Neurophysiol 59:317–325.

Shapley R, Enroth-Cugell C (1985) Visual adaptation and retinal gain controls. Prog Retin Res 3:263–346.

Sherman SM, Koch C (1986) The control of retinogeniculate transmission in the mammalian lateral geniculate nucleus. Exp Brain Res 63:1–20.

So YT, Shapley R (1981) Spatial tuning of cells in and around lateral geniculate nucleus of the cat: X and Y relay cells and perigeniculate interneurons. J Neurophysiol 45:107–120.

Sutter EE (1987) A practical non-stochastic approach to nonlinear time-domain analysis. Adv Meth Physiol Systems Model 1:303–315.

Theunissen F, Miller JP (1991) Representation of sensory information in the cricket cercal sensory system. II. Information theoretic calculation of system accuracy and optimal tuning-curve widths of four primary interneurons. J Neurophysiol 66:1690–1703.

Warland DK, Reinagel P, Meister M (1997) Decoding visual information from a population of retinal ganglion cells. J Neurophysiol 78:2336–2350.

Wässle H, Boycott B (1991) Functional architecture of the mammalian retina. Physiol Rev 71:447–480.

Zhang K, Ginzburg I, McNaughton BL, Sejnowski TJ (1998) Interpreting neuronal population activity by reconstruction: unified framework with application to hippocampal place cells. J Neurophysiol 79:1017–1044.