On the Importance of Static Nonlinearity in Estimating Spatiotemporal Neural Filters With Natural Stimuli

Tatyana O. Sharpee,[1,2] Kenneth D. Miller,[1,3] and Michael P. Stryker[1]

[1]*Sloan-Swartz Center for Theoretical Neurobiology and Department of Physiology, University of California, San Francisco;* [2]*Crick-Jacobs Center for Theoretical and Computational Biology, Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, California; and* [3]*Center for Theoretical Neuroscience and Department of Neuroscience, College of Physicians and Surgeons, Columbia University, New York, New York*

Sharpee TO, Miller KD, Stryker MP. On the importance of static nonlinearity in estimating spatiotemporal neural filters with natural stimuli. *J Neurophysiol* 99: 2496–2509, 2008. First published March 19, 2008; doi:10.1152/jn.01397.2007. Understanding neural responses with natural stimuli has increasingly become an essential part of characterizing neural coding. Neural responses are commonly characterized by a linear–nonlinear (LN) model, in which the output of a linear filter applied to the stimulus is transformed by a static nonlinearity to determine neural response. To estimate the linear filter in the LN model, studies of responses to natural stimuli commonly use methods that are unbiased only for a linear model (in which there is no static nonlinearity): spike-triggered averages with correction for stimulus power spectrum, with or without regularization. Although these methods work well for artificial stimuli, such as Gaussian white noise, we show here that they estimate neural filters of LN models from responses to natural stimuli much more poorly. We studied simple cells in cat primary visual cortex. We demonstrate that the filters computed by directly taking the nonlinearity into account have better predictive power and depend less on the stimulus than those computed under the linear model. With noise stimuli, filters computed using the linear and LN models were similar, as predicted theoretically. With natural stimuli, filters of the two models can differ profoundly. Noise and natural stimulus filters differed significantly in spatial properties, but these differences were exaggerated when filters were computed using the linear rather than the LN model. Although regularization of filters computed under the linear model improved their predictive power, it also led to systematic distortions of their spatial frequency profiles, especially at low spatial and temporal frequencies.

INTRODUCTION

Over the course of nearly 50 years neurons in the primary visual cortex have been studied with a variety of stimuli, ranging from the classic studies using edges, bars, and moving gratings, to more recent studies with random inputs and stimuli derived from the natural environment, as reviewed by Felsen and Dan (2005) and Rust and Movshon (2005). It is becoming increasingly important to develop unified models for neural responses to stimuli with a wide range of statistical properties, ideally extending to the fully natural case.

To accomplish this, it is important to be able to derive a neural response model from responses to natural stimuli. One cannot simply derive a response model from responses to a simpler ensemble, such as noise, and then extrapolate to the natural case. Typically there are significant differences between response models derived for a given cell from responses to different stimulus ensembles. In particular, neural response models derived from one stimulus ensemble often provide better predictions for neural responses to novel stimuli of the same type than to novel stimuli with different statistical properties (David et al. 2004; Sharpee et al. 2006; Woolley et al. 2006). The two general reasons for this are that neural responses are nonlinear and that they are adaptive. Even if neural responses were stationary and nonadaptive, it could be difficult to build a single model that adequately describes responses to different stimulus ensembles that evoke different regimes of neural responses. In addition, neurons appear capable of adapting to many statistical properties of the stimulus ensemble, meaning that response properties can change over time of exposure to a single ensemble (e.g., Lesica et al. 2007; Sharpee et al. 2006; Webster et al. 2002, 2006).

A common model for characterizing neural responses to a given ensemble is the linear–nonlinear (LN) model. In this model, a neuron's response is determined in two steps. First, the stimulus is linearly weighted with the neuron's spatiotemporal filter and this linearly weighted stimulus is summed to produce a number, the filter output. The firing rate is then given by an arbitrary static nonlinear function (such as a sigmoidal function) of the filter output, which we can call the neuron's input–output function. Here we compare two commonly used methods for estimating the spatiotemporal filter when fitting the LN model to a neuron's response, which we refer to as the linear and LN methods. The linear methods give unbiased answers (meaning answers that are correct in the limit of infinite data) for an arbitrary stimulus ensemble only for a neuron whose responses are determined by a linear model, that is, an LN model with a linear input–output function (cf. Fig. 1). For an LN model with a nonlinear input–output function, the linear methods give an unbiased estimate of the filter only if the stimulus ensemble is uncorrelated or correlated Gaussian noise (Agüera y Arcas et al. 2003; Bialek and de Ruyter van Steveninck 2005; Bussgang 1952; Chichilnisky 2001; de Boer and Kuyper 1968; Paninski 2003a; Ringach et al. 1997; Schwartz et al. 2006; Sharpee et al. 2004). LN methods give unbiased answers for arbitrary stimulus ensembles for neurons whose responses are determined by an arbitrary LN model.

linear model

stimulus   RF

X

spike probability

stimulus filtered by RF

linear-nonlinear model

stimulus   RF

X

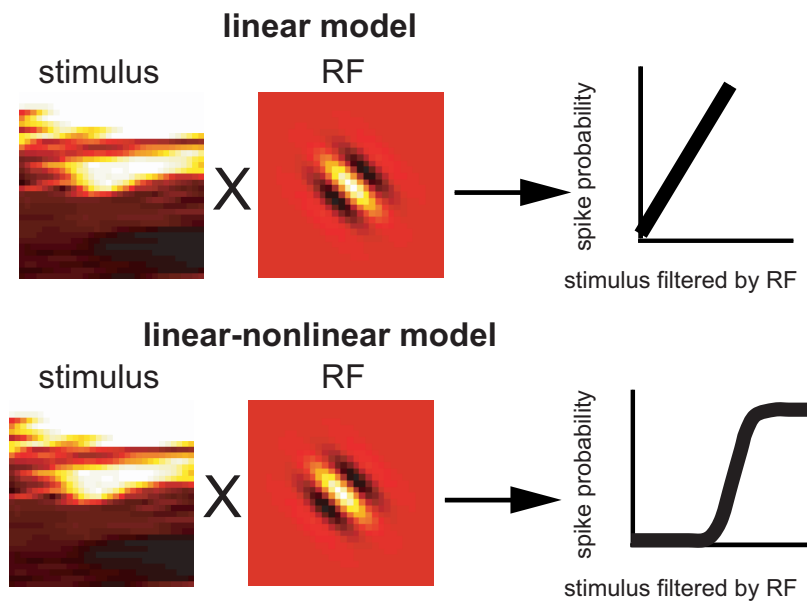spike probability

stimulus filtered by RF

FIG. 1. Schematic illustration of how the linear and linear–nonlinear (LN) models describe neural response. In the linear model, spike probability is proportional to the product of the stimulus with a receptive field (RF) filter. In the LN model, spike probability is an arbitrary nonlinear function of such a product. The first stages of the 2 models, convolution of RF filter with the stimulus, can be identical, but they differ in the linearity or nonlinearity of the transformation relating the convolution of RF filter with the stimulus to the firing rate. With Gaussian inputs, filters of the LN model can be correctly estimated assuming the simpler, linear model. However, for natural stimuli, which are non-Gaussian (Ruderman and Bialek 1994; Simoncelli and Olshausen 2001), the filter of the linear model will generally differ from that of the LN model.

Natural stimuli are strongly correlated and non-Gaussian (Field 1987; Ruderman and Bialek 1994; Simoncelli and Olshausen 2001). Thus the static nonlinearity causes bias in the spatiotemporal filters estimated from natural stimuli by the linear methods, but not by the LN methods. The more general conditions under which each method gives biased or unbiased answers are well known (Paninski 2003a; Sharpee et al. 2004).

The linear methods are often used to estimate filters for responses to natural stimuli, in spite of the bias they might have with such stimuli. Presumably this is due to the computational simplicity of the linear methods compared with the LN methods, and the belief that the bias is not too great a problem. Errors in estimation have at least two origins: systematic bias of the method used, which is independent of data size, and sampling error, meaning error in the estimation due to the finite amount of data. For realistic data sizes, it might be that the sampling error overwhelms the bias, so that the bias would be an insignificant source of error. Thus it is important to compare, with realistic amounts of data, the performance of the linear and the LN methods of spatiotemporal filter estimation, as we do here.

Here we find, using data recorded from simple cells in cat primary visual cortex (V1), that the spatiotemporal filter computed for natural stimuli in the LN model is, in general, significantly different from that computed in the linear model. We show that, although spatiotemporal filters computed in the LN model change with stimulus statistics, these changes are exaggerated when the spatiotemporal filters are computed in the purely linear model. Here, we have compared neural filters derived from responses to two stimulus sets: white noise and natural stimuli. Each stimulus set had the same mean luminance and contrast, but the two stimulus sets had different power spectra and higher-order statistical correlations. We find that neural filters computed in the LN model provide a consistently better description of the responses of simple cells in the primary visual cortex than those computed in the purely linear model. This was true for predicting responses to novel stimuli both within the same stimulus set and across different stimulus sets.

METHODS

Data collection and stimulus ensembles

All experimental recordings were conducted under a protocol approved by the University of California, San Francisco on Animal Research with procedures previously described (Emondi et al. 2004; Sharpee et al. 2004, 2006). Briefly, spike trains were recorded using tetrode electrodes from the primary visual cortex of anesthetized adult cats and manually sorted off-line. After manually estimating the size and position of the receptive fields, neurons were probed with full-field moving periodic patterns (gratings). In later off-line analysis, cells were selected as simple if, under stimulation by a moving sinusoidal grating with optimal parameters, the ratio of their response modulation (F1, i.e., amplitude of the Fourier transform of the response at the temporal frequency of the grating) to the mean response (F0) was >1 (Skottun et al. 1991). The rest of the protocol typically consisted of presenting an interlaced sequence of three different noise input ensembles of identical statistical properties (only seed values for the random number generator differed among the three input ensembles) and three different natural input ensembles. Visual stimulus ensembles of white noise and natural scenes were each 546 s long. The interval between presentations varied in duration as necessary to provide adequate animal care. All natural input ensembles were recorded in a wooded environment with a handheld digital video camera in similar conditions on the same day (Sharpee et al. 2006). The noise ensembles were white overall, but each particular frame was limited to spatial frequencies within a certain band. There were eight spatial frequency bands total. The intent of such design was to increase the number of elicited spikes. The mean luminance and contrast of the noise ensembles were adjusted to match those of the natural ensembles. Contrast was defined as the SD of luminance values relative to the mean. Both noise and natural inputs were shown at 128×128 pixel resolution, with angular resolution of about $0.12°$/pixel.

Spatiotemporal filters of the linear and LN models

Figure 1 compares the structure of the linear and LN model; each has a linear filter that is convolved with the stimulus (illustrated are purely spatial filters, although spatiotemporal filters will be analyzed for real neurons). The difference between models is that in the LN model the result of this convolution between stimulus and the linear filter is transformed by a nonlinear function into spike probability,

whereas in the linear model only the slope of the transformation can be adjusted.

To calculate spatiotemporal filters of real neurons, stimuli were downsampled from 128×128 pixels (angular resolution 0.12°/pixel) to 32×32 pixels (angular resolution 0.48°/pixel). All spatiotemporal receptive field analysis was carried out at this resolution. To find the center of the receptive field (RF), we computed the spike-triggered average (STA) stimulus for noise and natural ensembles. To make analysis computationally feasible and to minimize effects due to undersampling [we strove to have the number of spikes greater than the dimensionality of the RFs (Paninski 2003a; Sharpee et al. 2003, 2004)], a patch of 16×16 pixels (angular resolution of 0.48°/pixel) was selected around the maximum in the STA for each ensemble. All of the filter computations were confined to this patch, which was identical for noise and natural stimuli on a cell-by-cell basis. In all cases, subsequent analysis of the filter verified that it was fully contained within the selected patch.

In the case of white noise stimuli, neural spatiotemporal filters were computed as STAs (Chichilnisky 2001; de Boer and Kuyper 1968; Rieke et al. 1997) and as maximally informative dimensions (MIDs) (Sharpee et al. 2004, 2006). The STA represents the stimulus dimension along which the mean of the stimuli associated with spikes differs from the mean of all stimuli. The MID represents the stimulus dimension that carries the maximal amount of information about the arrival times of individual spikes. These two analyses should give the same answer in the case of white Gaussian noise or more generally of uncorrelated stimuli [by which we mean that each pixel's luminance at each time is drawn independently from a single luminance distribution; this is equivalent to the ensemble with mean luminance set to zero being spherically symmetric (Chichilnisky 2001)], provided only one stimulus feature is relevant for determining spike probability. This is because in this case the spatiotemporal filter of the linear model coincides with that of the LN model with one relevant dimension (Agüera y Arcas et al. 2003; Bialek and de Ruyter van Steveninck 2005; Bussgang 1952; Chichilnisky 2001; de Boer and Kuyper 1968; Paninski 2003a; Ringach et al. 1997; Sharpee et al. 2004). Because more than one stimulus dimension may be relevant for spikes of real neurons (Agüera y Arcas and Fairhall 2003; Bialek and de Ruyter van Steveninck 2005; Brenner et al. 2000a; de Ruyter van Steveninck and Bialek 1988; Fairhall et al. 2006; Felsen et al. 2005b; Rust et al. 2005; Slee et al. 2005; Touryan et al. 2002, 2005), the relevant dimensions may combine differently to form the dimension along which the mean changes most and that which is most informative.

In our case, the noise stimulus ensemble was composed of eight bandlimited Gaussian ensembles that together produced white noise (within each spatial frequency band, signals were Gaussian and white). Therefore there were small non-Gaussian effects and deviations from spherical symmetry, although these deviations were unlikely to affect the argument for the validity of the STA. For convenience, and in anticipation of its role for a natural ensemble, we refer to the STA as the spatiotemporal filter of the linear model for our noise inputs. The MID gives the spatiotemporal filter of the LN model regardless of the statistical properties of the stimulus ensemble (Paninski 2003a; Sharpee 2007; Sharpee et al. 2003, 2004). As we subsequently show, the differences between STA and MID filters computed for noise inputs were miniscule or absent.

In the case of natural stimuli, to compute spatiotemporal filters of the linear or LN model it is necessary to account for stimulus correlations, both pairwise and higher order. In the LN model, this is automatically done by the MID method. In the linear model, pairwise stimulus correlations can be accounted for by multiplying the STA by the inverse of the stimulus covariance matrix (David et al. 2004; Felsen et al. 2005b; Machens et al. 2004; Rieke et al. 1997; Ringach et al. 2002; Sharpee et al. 2004; Smyth et al. 2003; Theunissen et al. 2000, 2001; Touryan et al. 2005; Woolley et al. 2006). We will follow the common convention and refer to the resulting filter as the decorrelated STA (dSTA).

The dSTA, and thus the linear model, in principle is the correct filter for the LN model with an arbitrary nonlinearity if the stimulus ensemble is a multidimensional Gaussian, but it is biased for correlated non-Gaussian ensembles (Agüera y Arcas et al. 2003; Ahrens et al. 2008; Bialek and de Ruyter van Steveninck 2005; Bussgang 1952; Chichilnisky 2001; Christianson et al. 2008; de Boer and Kuyper 1968; Paninski 2003a; Ringach et al. 1997; Schwartz et al. 2006; Sharpee 2007; Sharpee et al. 2004). By a multidimensional Gaussian, we mean that the probability of a stimulus $\mathbf{s}$ (a vector of pixel luminances minus the mean luminance) is given by $P(\mathbf{s}) \propto \exp(-\frac{1}{2}\mathbf{s}^T\mathbf{C}^{-1}\mathbf{s})$, where $\mathbf{C}$ is the matrix of pixel–pixel covariance. For a multidimensional Gaussian, any one-dimensional slice through the distribution also has a Gaussian distribution, including, in particular, the distribution of luminances for any single pixel. Thus measuring the degree to which such a one-dimensional slice or the distribution across all pixels deviates from Gaussian is one measure of the degree to which the overall distribution deviates from Gaussian. We use this fact in the DISCUSSION where we use the kurtosis of the distribution of luminances at individual pixels across time as one measure of the deviation of a natural scenes distribution from a Gaussian distribution.

Because the procedure of inverting the stimulus covariance matrix tends to amplify sampling noise at the relatively underrepresented spatial and temporal frequencies, dSTA may not work well in practice. To correct this, several regularization strategies have been proposed (David et al. 2004; Machens et al. 2004; Smyth et al. 2003; Theunissen et al. 2000, 2001; Touryan et al. 2005; Woolley et al. 2006). To produce the regularized decorrelated STA filter (RdSTA) we first compute a pseudoinverse of the covariance matrix. Data were separated into three parts (1/8, 1/8, 3/4). One of the two smaller data sets was set aside for later use in evaluating the predictive power of the model. The largest data set was used to compute the STA. To form the pseudoinverse for a given cutoff value λ, we diagonalized the covariance matrix, finding its eigenvectors and the corresponding eigenvalues, and then computed the pseudoinverse based on all eigenvectors that had eigenvalues larger than λ (Press et al. 1992; Woolley et al. 2006). The candidate spatiotemporal filter of the linear model was then obtained by applying the pseudoinverse to the STA, and its performance was evaluated on the remaining small data set (1/8 of the whole data set) by the amount of mutual information it provided (Adelman et al. 2003; Agüera y Arcas et al. 2003; Sharpee et al. 2004). By trying all of the possible cutoff values λ, we selected that value for the cutoff that resulted in the best prediction. The corresponding filter was the RdSTA. Simulations on model cells (Sharpee 2007) showed that increasing the validation set size to 25% of the available data resulted in similar relative performance of the different methods (dSTA, RdSTA, MID) as we report here for cortical cells. Using percentage of explained variance instead of the mutual information also did not change performance of these methods on model neurons.

We considered both the dSTA and RdSTA as two alternative methods for estimating spatiotemporal filters of the linear model in the case of natural stimuli. Spatiotemporal filters of the LN model probed with natural stimuli were computed as MIDs (Sharpee et al. 2004, 2006). We note that computing STA, dSTA, and MID filters required setting aside only one validation data set, whereas computing the RdSTA required, in some calculations, setting aside two validation data sets: one of the data sets was used in selecting the optimal cutoff on eigenvalues of the covariance matrix that contributed to its pseudoinverse and the other to later evaluate the predictive power of the linear model based on the RdSTA filters. Computing the regularization without separate data sets for cutoff selection and validation would artifactually enhance the apparent predictive power.

Analysis of receptive field properties

To be able to perform statistical analysis of the properties of neural spatiotemporal filters, we computed 8 jackknife estimates for each of the methods and stimulus ensembles (Efron 1998). To obtain a jackknife estimate, the data were split in the same manner as described earlier for the validation purposes: a jackknife estimate was made using 7/8 of the data and the predictive power of the estimate was assessed using the remaining 1/8 of the data. The 8 jackknife estimates were obtained by dividing the data into 8 segments and using each 1/8-long segment as the omitted/validation segment for one estimate. As described earlier for the case of RdSTA filter estimates, each of the 8 jackknife estimates of RdSTA filter was obtained by further separating the training part of the data set (7/8 of the data) into 3/4 and 1/8, of which the latter small data subset was used to select the optimal cutoff value. The remaining 1/8 of the data were used to estimate the percentage of information or variance explained by the given jackknife estimate

(Fig. 7). Comparisons between filter estimates (Figs. 2–6) that did not involve computing predictive power on a novel data set were based on the RdSTA filters computed using 7/8 of the overall data set to find the STA and 1/8 to find the optimal cutoff λ.

To find the preferred orientation and spatial frequency we performed a one-dimensional Fourier transform in time, selecting a 2-Hz temporal frequency to match the temporal frequency used in stimulation with moving gratings. Then, the two-dimensional (2D) spatial Fourier transform was computed and the position of the peak was used as an initial guess for the preferred spatial frequency and orientation of the receptive field. Starting with this initial guess, the fitting of Gabor functions (Ringach 2002) to the spatiotemporal filter was always successful. The parameters of the best-fitting Gabor function were used as the estimate of the preferred orientation and spatial frequency. Such a procedure was repeated for each of the jackknife estimates, so that the mean and SD were obtained [according to jackknife *Eq. 11.5* in Efron
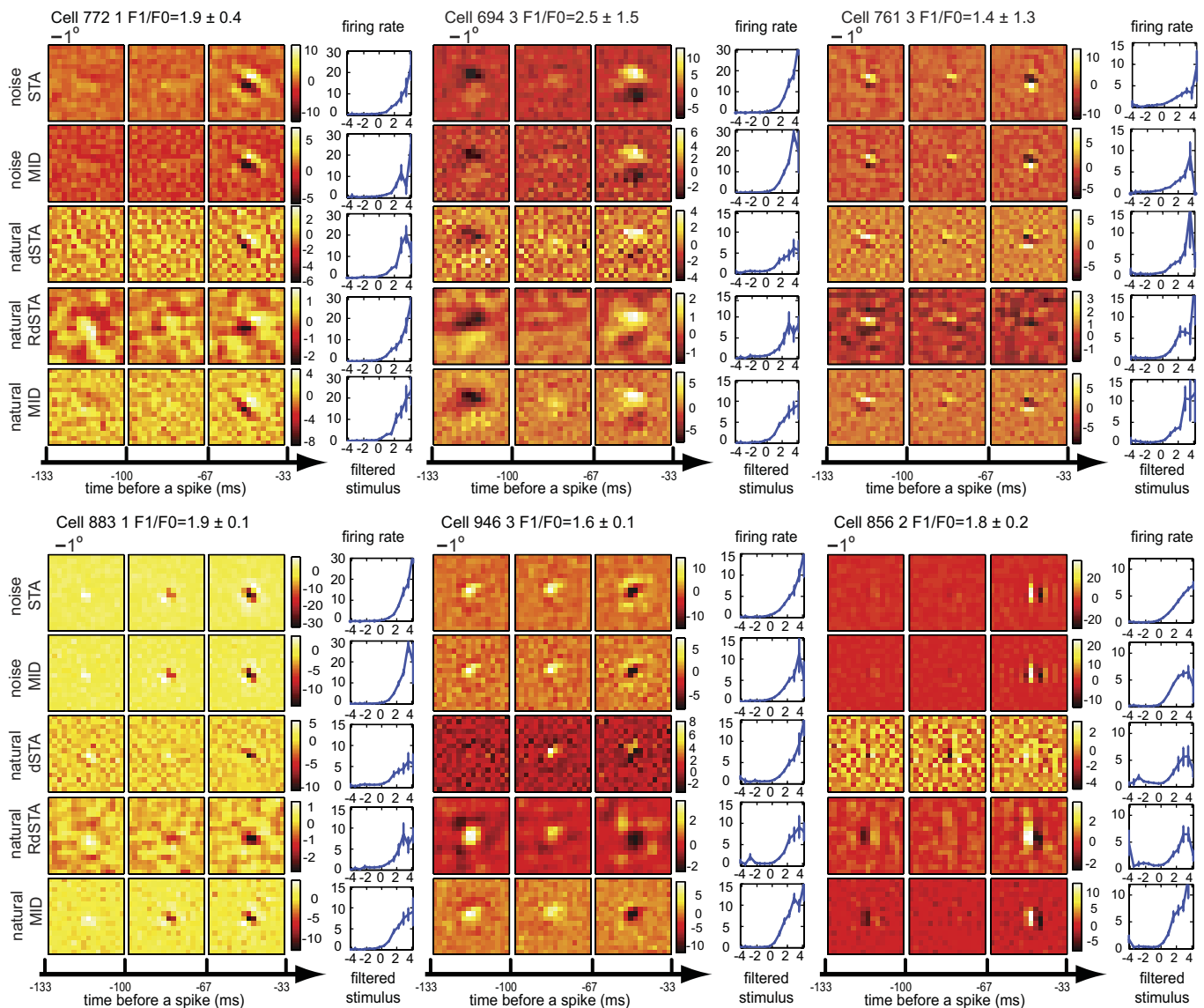

FIG. 2. Spatiotemporal filters computed from the linear and the LN models for noise and natural stimuli. Six example simple cells are shown. Filters were computed with the following methods and stimuli, from *top* to *bottom*: STA and MID for noise ensemble; dSTA, RdSTA, and MID for natural ensemble. Spatiotemporal filters have 3 time frames covering time interval (−133 to −33 ms) before a spike. The color scale shows the filter in units of its average noise level; scale bars: 1°. To the *right* of each spatiotemporal filter, we show the corresponding nonlinear gain function − the average firing rate (Hz; *y*-axis) for a given value of the filtered stimulus (*x*-axis; filtered stimulus values are shown relative to their mean and normalized by their SDs). STA, spike-triggered average; dSTA, decorrelated spike-triggered average; RdSTA, regularized decorrelated spike-triggered average; MID, most informative dimension.

(1998)] for both preferred orientation and spatial frequency values, as shown in Figs. 3 and 4. Comparisons between parameters of the neural filters derived from different stimuli and models were evaluated using linear fits that took into account error bars in both *x*- and *y*-axes, using Mathematica software (Wolfram Research). Point-by-point comparison between spatiotemporal filters from different models and stimulus ensembles (cf. Fig. 6) was also done based on eight jackknife estimates.

Spatial frequency profiles shown in Fig. 5 were obtained by taking the Fourier transform in time and, with zero-padding to 32 × 32, in space. Linear interpolation between pixels of the 2D transform was used to derive one-dimensional profiles along the preferred orientation of each cell. Before averaging across cells, the frequency profiles of individual cells were normalized to unit length (sum of squares equal to 1) across all spatial and temporal frequencies.

Generating and evaluating predictions

We used mutual information between the stimulus convolved with a particular filter and the firing rate as a measure of the filter's predictive power (Adelman et al. 2003; Agüera y Arcas et al. 2003; Fairhall et al. 2006; Sharpee et al. 2004, 2006). Specifically, mutual information accounted for by a spatiotemporal filter is computed as

$$I(\mathbf{L}) = \int dx P_L(x|\text{spike}) \log_2 \left[\frac{P_\mathbf{L}(x|\text{spike})}{P_\mathbf{L}(x)} \right] \qquad (1)$$

Here, $P_L(x)$ is the probability distribution of the projections x onto the filter $\mathbf{L}$ of all of the presented stimuli. Similarly, $P_\mathbf{L}(x \mid \text{spike})$ is the probability distribution of projections onto the filter $\mathbf{L}$ of all stimuli that lead to a spike. Information along any
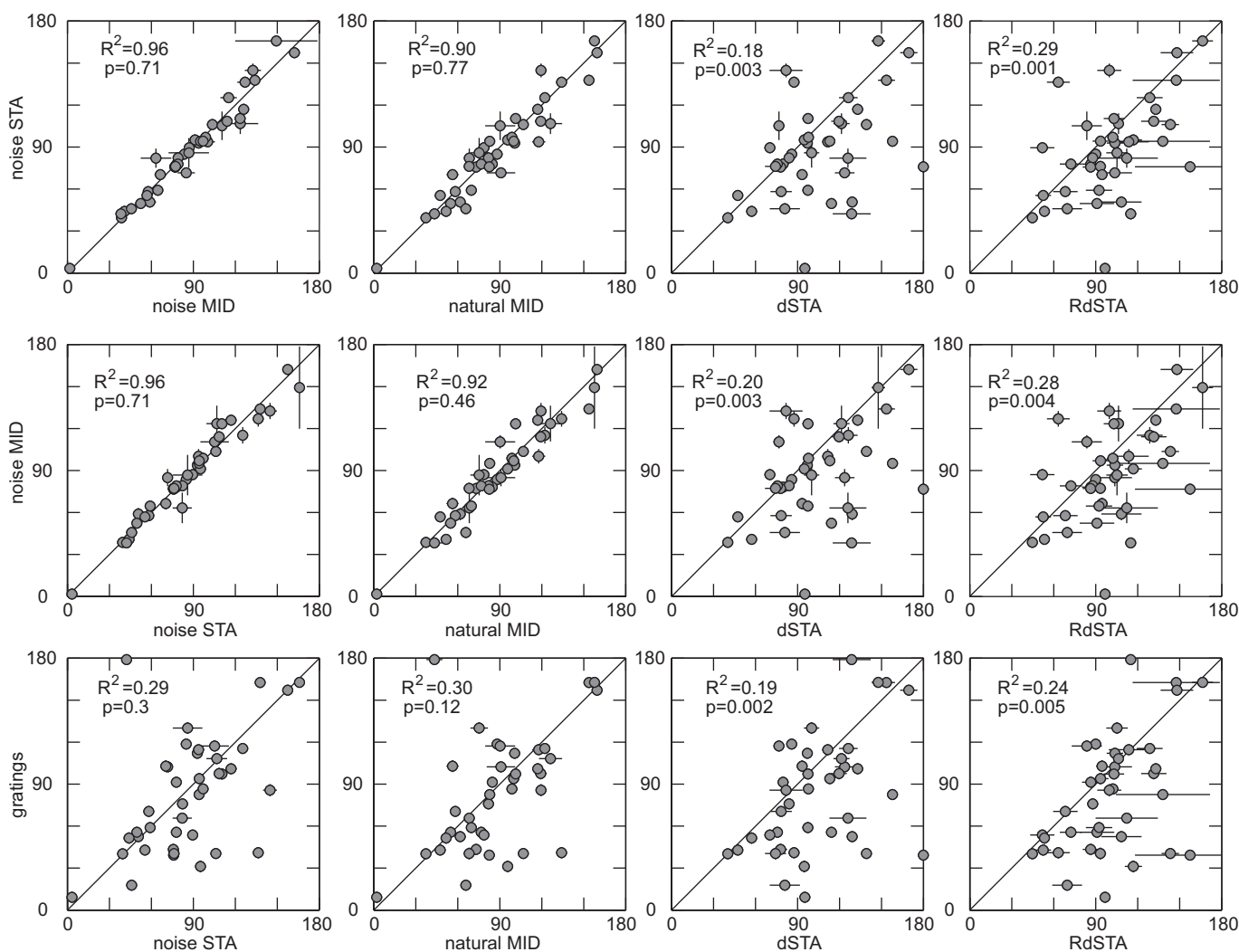

FIG. 3. Absence of change in preferred orientation derived from filters of the linear and LN model, with natural or noise stimuli. Preferred orientation values were computed by fitting Gabor functions to spatial filters obtained as 2-Hz component of spatiotemporal filters (see METHODS). *x*- and *y*-axes: orientation values are in degrees. The 2-sided length of an error bar shows 1/2 SE. Solid line has a slope of one. Panels in the *top row* provide comparison between preferred orientation of the noise STA (*y*-axis) and that of the remaining 4 filter estimates (*x*-axis). In panels of the *middle row*, preferred orientation from the noise MID filter (*y*-axis) is compared with those from the other 4 filter estimates (*x*-axis). Comparison with preferred orientation values from moving square gratings are shown in the *bottom row*. From *left* to *right*, the *x*-axis is: the noise MID (*top row*) or noise STA (*middle* and *bottom rows*); the natural MID; the dSTA for natural stimuli; the RdSTA for natural stimuli. Comparison between the preferred orientations of the dSTA (RdSTA) derived from natural stimuli and the preferred orientations of the MID filter also derived from natural stimuli (data not shown) were similar to the corresponding graphs based on the noise MID filter from the *middle row*. P values are from paired Wilcoxon test on the equality of *x*- and *y*-values.
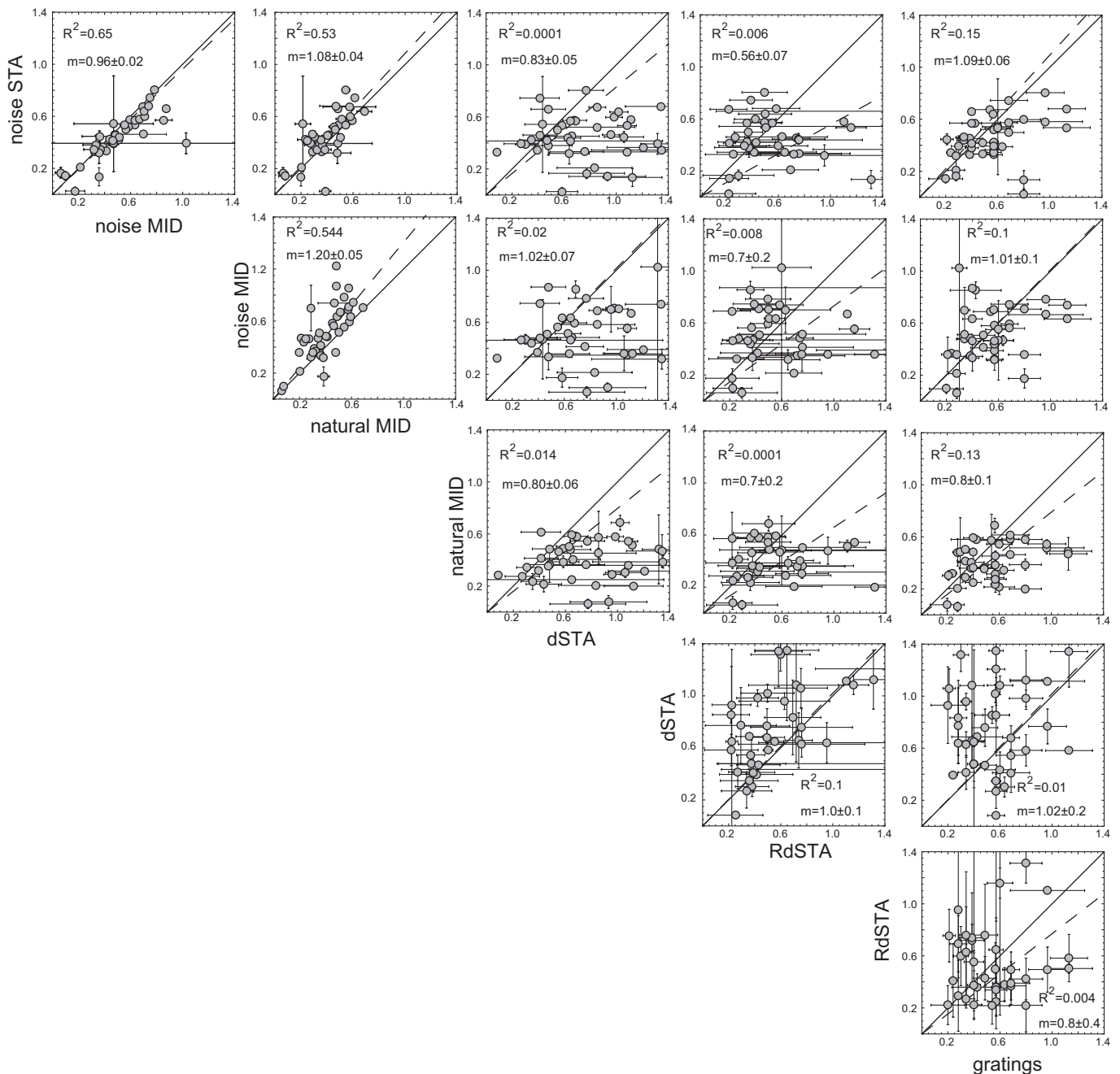
FIG. 4. Comparison of preferred spatial frequency derived from noise and natural spatiotemporal filters and gratings responses. Plots are arranged as a top half of a 6 × 6 matrix, because there are 5 filter estimates in total and one measurement from gratings. Each row has the same data set on the *x*-axis and each column has the same data set on the *y*-axis. From *top* to *bottom*, data sets on the *y*-axis represent spatial frequency values derived from the STA for noise stimuli, the MID for noise stimuli, the MID for natural stimuli, the dSTA for natural stimuli, the RdSTA for natural stimuli, and the value derived from moving sinusoidal gratings. The data sets on the *x*-axis follow the same order from *left* to *right*. Spatial frequency values are given in cycles/degree. Each point represents a separate cell (*n* = 40 points). The 2-sided length of the error bar around each data point is 1/2 SE. Solid line has a slope of one; dashed line shows the best fit taking into account error bars. The values for the best-fitting slope are given in each plot, with their SDs. R^2 values for variance explained by a linear relationship are also given within each plot.

stimulus dimension, including the relevant spatiotemporal filter, may not exceed the overall information carried in the times of occurrence of single spikes. This overall information can be evaluated as (Brenner et al. 2000b)

$$I_{spike} = \frac{1}{T}\int dt \, \frac{r(t)}{\bar{r}} \log_2 \frac{r(t)}{\bar{r}} \qquad (2)$$

where $r(t)$ is the firing rate over multiple repetitions of a single

stimulus segment that is characteristic of the stimulus ensemble of interest and $\bar{r}$ is the mean firing rate. The ratio $I(\mathbf{L})/I_{spike}$ can be used to measure predictive power. Neural responses to 50–150 repetitions of an approximately 11-s-long segment of the natural or noise ensemble were used to compute I_{spike}.

In addition to mutual information, we also computed variance accounted for by a given spatiotemporal filter together with the nonlinear transformation from filter outputs to spike probability. This is done by finding the best nonlinearity for a given filter and the
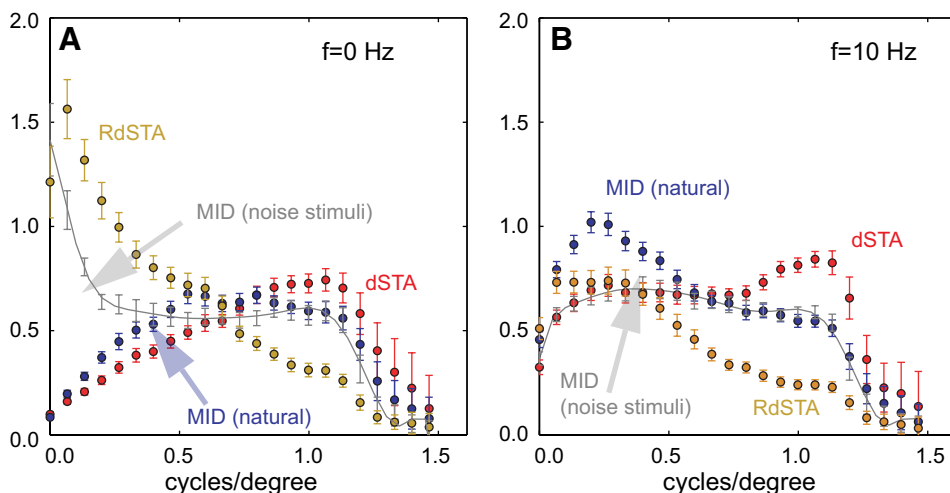
FIG. 5. Population average of amplitude spectra of spatiotemporal filters along their preferred orientation. Filters were derived from neural responses to natural stimuli. Analysis based on the MID filters is shown in blue, analysis for the dSTA filters is shown in red, and analysis for the RdSTA filter is shown in orange. Gray error bars show amplitude spectra for noise MID filters, with a polynomial fit as a gray solid line (Sharpee et al. 2006). *A*: spatial frequency profiles at zero temporal frequency, f = 0 Hz. *B*: spatial frequency profiles for temporal frequency f = 10 Hz. Error bars show SEs. Amplitude spectra $S(k, \omega)$ for each neuron were normalized to have $\sum_{k,\omega} S(k, \omega)^2 = 1$, where the sum is taken across all spatial and temporal frequencies.

recorded sequence of spikes and then computing the predicted amount of variation based on the reconstructed LN model. The two steps can be combined into one equation (Paninski 2003a; Sharpee 2007), which gives the predicted variance normalized by $\bar{r}^2$

$$V(\mathbf{L}) = \int dx P_{\mathbf{L}}(x) \left[\frac{P_{\mathbf{L}}(x|\text{spike})}{P_{\mathbf{L}}(x)} \right]^2 - 1 \qquad (3)$$

This equation is similar to *Eq. 1* for the mutual information and relies on the same probability distributions. Variance accounted by a given LN model cannot be larger than the overall variance in the firing rate

$$V_{\text{max}} = \frac{1}{T} \int dt \left[\frac{r(t)}{\bar{r}} \right]^2 - 1 \qquad (4)$$

As was pointed out by Sahani and Linden (2003) and Machens et al. (2004), both the variance in the firing rate and variance accounted by a given LN model have to be corrected for the positive bias due to finite size of the data set to determine the amount of explainable variance. Procedures for correcting for this bias are described next.

To avoid overestimation in predictive power due to overfitting, mutual information and explained variance were computed using the remaining 1/8 of the data set not used to derive the spatiotemporal filters themselves (this data set was also not used to select the optimal cutoff for RdSTA filters, as described earlier). Mutual information values are positively biased (Brenner et al. 2000b; Paninski 2003b; Strong et al. 1998; Treves and Panzeri 1995). Similar effects happen for variance (Machens et al. 2004; Sahani and Linden 2003). To correct for this bias, we extrapolated the relationship between mutual information (variance) and the inverse of the data set size to the infinite data limit using linear extrapolation based on 80, 85, 90, 95, and 100% of the data from the test set. This procedure was carried out for each jackknife estimate, model, and type of stimulus, as well for the total value of information I_{spike} between unfiltered stimuli and spikes.

We note that the measures of predictive power we are using—the mutual information between filtered stimuli and spikes and the variance in the firing rate by the LN model based on a given spatiotemporal filter—reflect the predictive power based on the best possible nonlinear transformation between filtered stimuli and the spike probability. In other words, the percentage of the information (variance) explained quantifies the best predictive power achievable by a given spatiotemporal filter and arbitrary nonlinearities. Thus although an LN model is more powerful than a linear model by virtue of its nonlinear input–output function, this is not the cause of lower predictive power of the spatiotemporal filters computed in the linear model. Instead, our method assays how accurate a filter a given model (linear or LN) can produce, with an understanding that the predictive power will be

compared taking nonlinear gain functions into account even for spatiotemporal filters computed using the linear model.

RESULTS

We computed the spatiotemporal filters of simple cells probed with natural and noise stimuli according to the assumptions of the linear and LN models. Our goal was to compare how the spatiotemporal filters computed using the linear and LN model changed with the stimulus ensemble. The analysis is based on 40 simple cells in the primary visual cortex recorded in four animals. Spatiotemporal filters of the linear model were estimated as the spike-triggered average (STA) stimulus in the case of white noise stimuli and as the decorrelated STA (dSTA) or its regularized version (RdSTA) for natural stimuli (see METHODS). Spatiotemporal filters of the LN model were estimated as the most informative dimension (MID). In Fig. 2, we show spatiotemporal filters computed according to the linear and LN models for six example simple cells. In agreement with previous findings (David et al. 2004; Felsen et al. 2005b; Sharpee et al. 2006; Smyth et al. 2003), we observed that the various filter estimates were qualitatively similar to each other, even when computed from different stimulus ensembles. This was evident in the overall spatial extent of the filters and in the variation of their peak amplitudes in time. For each spatiotemporal filter, we also show the best nonlinearity that relates stimuli convolved with the filter to the neural firing rate, which is given by associating each filter output value with the mean evoked firing rate averaged over all stimuli having that filter output value.

Orientation selectivity

To compare the spatiotemporal filters quantitatively, we begin by examining preferred orientation values associated with different filters (cf. Fig. 3). We found no significant differences in preferred orientation between the STA and MID filters for white noise stimuli ($R^2 = 0.96$). This is to be expected because for white Gaussian inputs, the STA provides the filter of both the linear and the LN models and non-Gaussian effects in the white noise stimulus ensemble were small. The STA and MID filters should therefore coincide, unless multiple relevant dimensions are present and contribute differently to these filters (see METHODS). Importantly, we found

no significant differences between the preferred orientations of MID filters obtained from neural responses to natural stimuli and those of either the MID filter ($R^2 = 0.92$; $P = 0.46$, paired Wilcoxon test) or the STA filter ($R^2 = 0.90$; $P = 0.77$, paired Wilcoxon test) obtained from neural responses to the noise ensemble. The preferred orientations of filters computed as the dSTA for natural stimuli were much less correlated with, and differed significantly ($P < 0.01$, paired Wilcoxon test) from, those derived from noise stimuli using either the STA or the MID filters for noise stimuli, with $R^2 = 0.18$ and 0.20, respectively. Regularizing the linear filter for natural stimuli (RdSTA) did not produce much improvement, resulting in correlations $R^2 = 0.29$ and 0.28 with the white noise STA or MID filters, respectively. The differences in orientation values derived from the RdSTA and either the white noise STA or MID remained significant ($P < 0.01$, paired Wilcoxon test). Thus the filters produced by the LN model, but not the linear model, from neural responses to natural stimuli produce similar preferred orientations to filters produced by either model from neural responses to noise stimuli.

Similar results are found when comparing preferred orientations computed from neural filters to those determined by neural responses to moving square gratings (Fig. 3). The filters produced by the LN model in response to natural stimuli and those produced by either model in response to noise stimuli showed no significant difference in preferred orientation from the orientations determined from studies with gratings, in agreement with previous studies (Smyth et al. 2003; Usrey et al. 2003). However, the correlation values are smaller than for comparisons between different filters (noise STA: $R^2 = 0.29$, $P = 0.3$, paired Wilcoxon test; noise MID: $R^2 = 0.24$, $P = 0.3$, paired Wilcoxon test, not shown in Fig. 3; natural MID, $R^2 = 0.30$, $P = 0.12$, paired Wilcoxon test). We also note that although estimates of the preferred orientation from gratings and the noise STA agree well for 90% of the cells, the four cells that exhibited a large disparity between the two estimates had either a firing rate or a preferred spatial frequency within the lowest 10% of the population.

In contrast to the performance of filters derived from responses to noise stimuli or from the LN model in response to natural stimuli, the filters produced by the linear model in response to natural stimuli resulted in preferred orientations that were significantly different from those determined by responses to gratings (dSTA: $R^2 = 0.19$, $P = 0.002$, paired Wilcoxon test; RdSTA: $R^2 = 0.24$, $P = 0.005$, paired Wilcoxon test).

Preferred spatial frequency

Next, we examined differences in preferred spatial frequency values (cf. Fig. 4). Altogether there are six estimates of preferred spatial frequency: one value derived from neural responses to moving gratings and five values derived from five different filter estimates for each neuron. The five filter estimates include two for noise stimuli (STA, MID) and three for natural stimuli (dSTA, RdSTA, MID). We present our results as the upper half of a 6×6 matrix of pairwise comparisons. Each row has the same data set on the y-axis and each column has the same data set on the x-axis. Preferred spatial frequency for the two filter estimates derived from neural responses to white noise (STA and MID) were strongly correlated, with

$R^2 = 0.65$ and the value for the best-fitting slope 1.03 ± 0.02 (SD). By comparing filters derived from noise and natural stimuli, we found that preferred spatial frequencies of filters derived from noise inputs were slightly, but statistically significantly, higher than those of filters derived from natural inputs with the LN model. The slope of the best-fitting line (taking into account error bars) was 1.20 ± 0.05 when the noise MID filter was compared with the natural MID filter and 1.09 ± 0.04 when the noise STA filter was compared with the natural MID filter. Measurements derived from neural responses to gratings were not significantly different from those based on the noise MID, somewhat smaller than those based on the noise STA (best-fitting slope of 1.09 ± 0.06), and larger than those based on the natural MID (best-fitting slope of 0.8 ± 0.1). Filters derived from natural inputs under the linear model (dSTA or RdSTA) had preferred spatial frequencies that were poorly correlated with those derived from either of the two noise filters, from the MID filters for natural stimuli, or from neural responses to moving gratings (all $R^2 < 0.03$). We also note that error bars on the preferred spatial frequency values were substantially larger for filters derived from natural inputs under the linear model than for filters derived from natural inputs under the LN model or filters derived from noise inputs.

Spatial frequency profiles

Having found no changes in the preferred orientation between noise and natural stimuli and some change in the preferred spatial frequency, we proceeded to examine differences in the frequency composition of the spatiotemporal filters, measured at the preferred orientation (see METHODS). Previous results (Sharpee et al. 2006) showed that spatial frequency profiles can change profoundly between noise and natural inputs (when estimated as MID filters), without large changes in the preferred spatial frequency values. In Fig. 5 we show the relative frequency composition, averaged across our population of cells, for the three different methods (dSTA, RdSTA, and MID) of estimating spatiotemporal filters with natural stimuli. Even though the dSTA filters are known to be subject to noise amplification, their spatial frequency profiles (shown in red) at low spatial frequency closely resemble the behavior of the MID filters (shown in blue). Some noise amplification does indeed take place at higher spatial frequencies around 1 cycle/degree, and this is more pronounced at the larger temporal frequency f = 10 Hz than at the low temporal frequency f = 0 Hz. A common strategy to deal with the problem of noise amplification at higher spatial frequency is to impose a smoothness constraint, which effectively filters out higher spatial frequencies where signal-to-noise ratio is low. The RdSTA is an example of this strategy. In agreement with previous reports (David et al. 2004; Felsen et al. 2005b; Smyth et al. 2003; Theunissen et al. 2000, 2001; Woolley et al. 2006), spatial frequency profiles of the RdSTA filters (shown in orange) were strongly biased toward low spatial frequencies and, to some extent, to low temporal frequencies.

Thus not only do the three methods of estimating filters with natural stimuli produce spatial frequency profiles that are profoundly different, but also different implementations of the linear model (dSTA and RdSTA) profoundly differ from each other. For comparison, we replot spatial frequency profiles from the noise MID filters in gray solid line (Sharpee et al.

2006). Most notably, for zero temporal frequency, the differences in spatial frequency composition between the dSTA and the RdSTA profiles far exceeded the differences between spatial profiles computed for noise and natural stimuli using the LN model (the MIDs). At low spatial frequencies, the RdSTA shows higher-amplitude spectra than both noise and natural MIDs, whereas the dSTA shows smaller-amplitude spectra than both noise and natural MIDs (although it is close to the frequency composition of the natural MID filters). At higher spatial frequencies, the situation is the reverse. Here, for both 0- and 10-Hz temporal frequencies, the RdSTA shows less high-frequency content than MID filters computed for noise and natural stimuli, whereas the dSTA overestimates the frequency content.

The frequency compositions of the MID filters computed for natural and noise stimuli are approximately identical at higher spatial frequencies. This can be viewed as providing additional support for the computation underlying the LN model because, in the absence of an external smoothness constraint, as in the case of computing the RdSTA, artifacts would tend to accumulate at the higher spatial frequencies, which have much less power than low frequencies in natural stimuli and hence are relatively undersampled.

Similarity of spatiotemporal filters according to correlation coefficients

The previous section showed the much greater reliability of the MID method compared with the various STA methods for measuring spatial frequency selectivity from natural stimuli. Describing the relative sensitivity to different spatial frequencies is an important characterization of neural filters. However, the spatiotemporal filters may also be compared point by point, using correlation coefficients between pairs of filters. We will use the spatiotemporal filters computed for noise stimuli, where the different methods agree, as a reference. On discretization in time and two spatial coordinates, as is necessary for any practical computations, the spatiotemporal filter becomes a multidimensional vector in the stimulus space, where each pixel is a separate dimension (in our case, the dimensionality is $16 \times 16 \times 3$; i.e., 16×16 spatial sampling and 3 time lags). Therefore it is only natural to measure the similarity of two filters as a normalized dot product between them, that is, as a correlation coefficient (CC). Two identical filters have a CC of 1; very dissimilar filters will have a CC of 0. Although there is only one way for the CC between two filters to be 1 (i.e., when they are identical), there are many filters that describe dimensions that are orthogonal to each other and therefore have CC values of 0. We note that the sign of the spatiotemporal filter can always be changed to the opposite if accompanied by a simultaneous inversion of the *x*-axis on the nonlinear gain functions (shown in Fig. 2). For example, a filter with large positive peak together with increasing firing rate with increasing stimulus similarity to the filter is equivalent to a contrast-inverted filter having a negative peak together with firing rate that decreases with stimulus similarity. This means that the sign of the correlation between two filters is not meaningful, so the correlations can be taken to always be nonnegative.

The results of such an analysis are given in Fig. 6. We first compare similarity between spatiotemporal filters of the LN model computed for noise versus natural stimuli (as MID

filters) to the similarity of the filters of the linear model computed for noise stimuli (STA) versus natural stimuli (dSTA). As can be seen in Fig. 6A, for all cells, the spatiotemporal filters of the LN model were more similar to each other across stimulus type than filters of the linear model. This was also true if filters computed for the natural stimuli in the linear model were regularized (Fig. 6B). Because non-Gaussian effects in the white noise stimulus ensemble were small, the STA for noise stimuli also approximates the filter of the LN model and, to that extent, is interchangeable with the noise MID filter. In Fig. 6, *C* and *D*, we use the noise MID filter instead of the noise STA filter to quantify changes in filters between noise and natural stimuli. Here the only difference between *x*- and *y*-axes is in the method for computing filters for natural stimuli. In Fig. 6C we show that for all cells the noise MID filter is closer to the natural MID filter than to the dSTA filter computed for natural stimuli. In Fig. 6D this comparison is carried out between the natural MID filter and the RdSTA. Although there is more variability associated with the RdSTA filters, for most of the cells (37 of 40), the noise MID filter is still closer to the natural MID filter than to the RdSTA filter computed for natural stimuli.

Similarity between spatiotemporal filters obtained with noise and natural stimuli was correlated with both similarity of the corresponding nonlinear gain functions (Fig. 6E) and the average of the predictive power of the natural MID filter in predicting responses to a novel natural stimulus segment and that of the noise MID filter in predicting responses to a novel noise stimulus segment (Fig. 6F). On average there was greater similarity between nonlinear gain functions than between the spatiotemporal filters because all but 7 of 40 cells are above the unity line in Fig. 6E.

To summarize this section, among the three methods of estimating spatiotemporal filters with natural stimuli, the MID method produces filters that are by far the closest to the noise filters. Thus spatiotemporal filters of the LN model appear to be more stable under changes between white noise and natural stimuli than do receptive fields of the linear model.

Predictive power of the linear and LN models

A final criterion for comparing different estimates of spatiotemporal filters is their predictive value. How accurately do filters associated with linear and LN models predict the response to novel stimuli, which were not used to compute the receptive field? Note that, in all cases, we are studying the predictive power of an LN model using the given filter, with the nonlinearity chosen to be optimal for the given filter as described previously. The only difference is how the filter was computed, in particular, whether the nonlinearity was taken into account in computing the filter.

In Fig. 7. we show an example of spike responses to 110 repetitions of the same segment from the natural stimulus ensemble (Fig. 7A) Figure 7, *B* and *C* illustrates comparison between the average firing rate (black line) and its predictions according to the differently reconstructed spatiotemporal filters and nonlinearities (spatiotemporal filters and nonlinearities for this cell are shown in the *middle column*, *bottom row* of Fig. 1). The data from this segment of the natural stimulus were not used in estimating either the filters or the associated nonlinearities. Comparison between actual firing rate and our
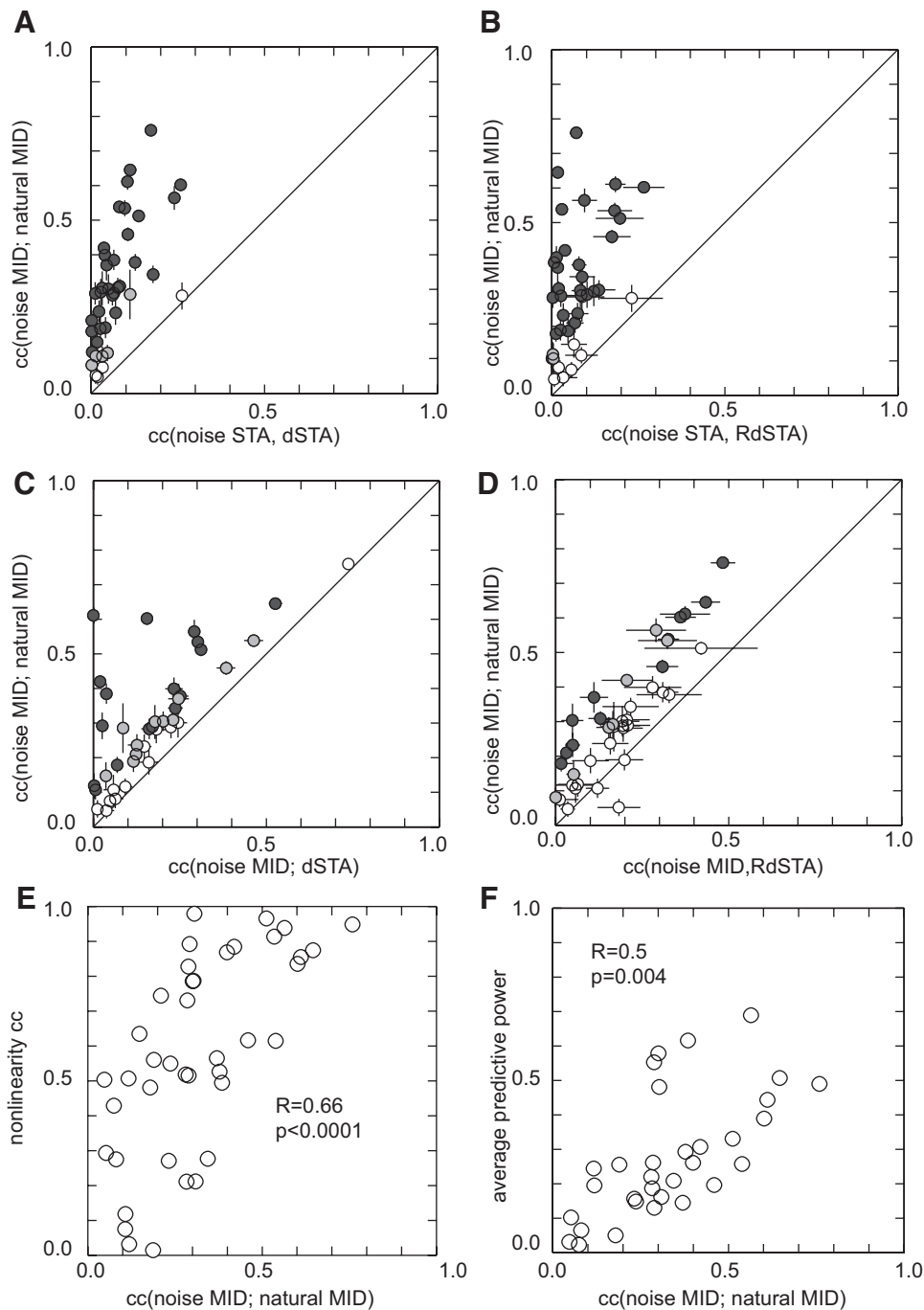
FIG. 6. Correlation between spatiotemporal filters of the linear and LN models with noise and natural stimuli. Similarity between 2 spatiotemporal filters is quantified by their correlation coefficient (CC, or normalized dot product). For all cells, stimuli, and methods for estimating filters, the filter dimensionality was $16 \times 16 \times 3$, 16 pixels horizontally, 16 pixels vertically, and 3 time lags. *A*: CC values between MID filters computed for noise and natural stimuli (*y*-axis) are plotted as a function of CC values between the dSTA filter computed from natural stimuli and the STA filter computed from noise stimuli. *B* is the same as *A*, except that the comparison on the *x*-axis is between the RdSTA filter computed from natural stimuli and the STA filter computed from noise stimuli. In *C* and *D*, the noise filter is computed as the MID compared with both the natural MID filter (*y*-axis) and either the dSTA filter (*x*-axis for *C* or the RdSTA filter (*x*-axis for *D*). Error bars show SDs. Solid line has a slope of one. In *A–D*, color indicates statistical significance of deviation from identical CC values (white, $P > 0.05$; gray, $0.01 < P < 0.05$; black, $P < 0.01$, *t*-test). Greater similarity between spatiotemporal MID filters under natural and noise stimulation is associated with greater similarity in the corresponding nonlinear gain functions (*E*, $P < 10^{-4}$, $R = 0.66$) and greater average predictive power of natural MID for neural responses to a novel segment of natural scenes and of noise MID to responses for novel segment of noise (*F*, $P = 0.004$, $R = 0.5$).

predictions based on the natural MID filter and the corresponding nonlinearity (purple, shown inverted for clarity) is shown in Fig. 7*B*. Although the reconstruction has difficulty predicting very high peaks in the firing rate, more moderate peaks ≤ 30 Hz, as well as the timing of the peaks, can be fairly well reconstructed. In Fig. 7*C* we show an expanded view of the actual firing rates and predictions based on three different filters (MID, dSTA, and RdSTA) and their corresponding nonlinearities.

To quantify the average predictive power, we consider both percentage of explained variance and information, determined as a ratio between variance (information) accounted for by a given filter with the best possible nonlinearity, and the explain-

able variance (information) in neural response. To determine the explainable variance, previous studies have indicated that filters derived from natural stimuli predict responses to natural stimuli better than responses to noise (Sharpee et al. 2006; Woolley et al. 2006) or gratings (David et al. 2004); and, vice versa, filters derived from noise stimuli predict responses to noise better than responses to natural stimuli (Sharpee et al. 2006; Woolley et al. 2006). Therefore we concentrate here on comparing estimates of the power of various filters derived from natural stimuli (MID, dSTA, and RdSTA) for predicting responses to either natural or noise stimuli. In the case of natural stimuli, spikes to be predicted were taken from a novel part of the natural stimulus ensemble that was not used for
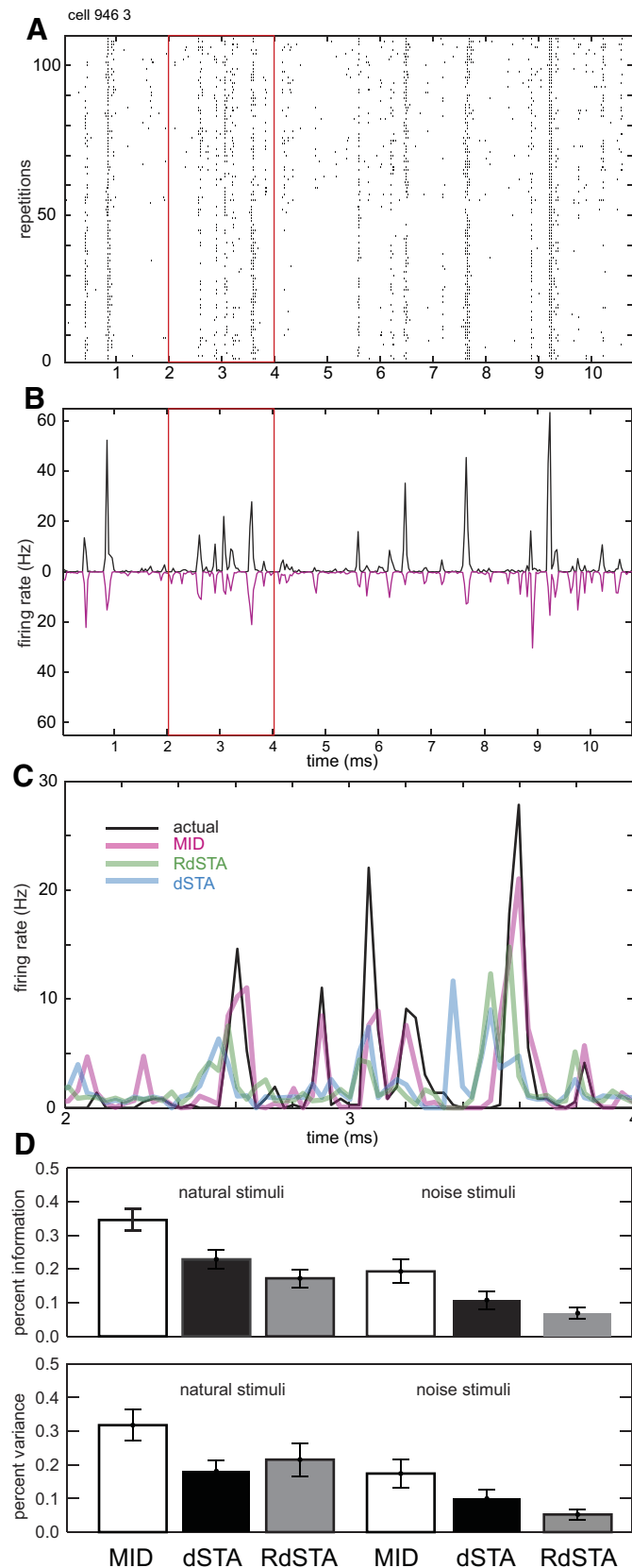
computing the filters or nonlinearities. Predictions for noise stimuli were made for a noise stimulus segment of the same duration as the novel natural stimulus segment.

We found the expected large decrease in predictive power from natural to noise stimuli (Fig. 7D). Beyond that, the MID filters computed from natural stimuli provide better predictions than either the dSTA or the RdSTA computed from natural stimuli. This was true for predictions of responses to both natural and noise stimuli, using either percentage of information or variance to measure predictive power. In the top part of Fig. 7D, we show the percentage of the overall mutual information, I_{spike}, between the entire stimulus and spike trains that was explained by LN models with filters obtained by the different methods. The bottom half of Fig. 7D provides analogous comparisons in terms of the percentage of variance explained. Procedures for determining explainable variance (Machens et al. 2004; Sahani and Linden 2003) are described in METHODS. Both ratios allow one to infer quantitatively how much predictions could potentially improve if the model were expanded to incorporate neuronal sensitivity to more than one stimulus dimension (Agüera y Arcas and Fairhall 2003; Bialek and de Ruyter van Steveninck 2005; Brenner et al. 2000a; de Ruyter van Steveninck and Bialek 1988; Fairhall et al. 2006; Felsen et al. 2005b; Rust et al. 2005; Slee et al. 2005; Touryan et al. 2002, 2005). However, information may be the more appropriate measurement with natural stimuli because of their non-Gaussian properties. We note that measurement of I_{spike} or the overall variance was available only for $n = 32$ neurons in our data set for which a segment of stimulus was repeated a sufficient number of times ($>$50).

On average, the MID filters derived from natural stimuli accounted for $35 \pm 3\%$ (SE) of I_{spike}. Compared with other filter estimates, the MID filters provided significantly better predictions of neural responses to natural stimuli than either the dSTA ($P < 10^{-4}$, comparison between percentages of explained information; $P < 0.001$, comparison between percentages of explained variance) or the RdSTA ($P < 10^{-4}$, comparison between percentages of explained information; $P = 0.01$, comparison between percentages of explained variance). The two-tail paired Wilcoxon test was used for all comparisons. The same effect was observed (cf. Fig. 7) for predictions of responses to noise stimuli ($P < 10^{-4}$ for comparisons with either the dSTA or the RdSTA in terms of percentages of explained information, and $P < 0.002$ for comparisons be-

FIG. 7. Predictive power of the LN models using spatiotemporal filters derived under either the linear or LN models from natural stimuli. A: responses of a simple cell from primary visual cortex to repeated presentations of the same segment of natural scene. Red box delineates the time interval for which we provide an expanded view in C. B: comparison of the actual firing rate (black) with predictions based on the natural MID filter and the corresponding nonlinearity (purple, shown inverted for clarity). C: expanded view of comparison between the actual firing rates and our predictions based on the dSTA (blue), RdSTA (green), and MID (purple) filters together with their corresponding nonlinearities (all shown for this cell in the *middle column, bottom row* of Fig. 1). D: predictions for spikes elicited by natural stimuli are grouped to the *left*, whereas predictions for spikes elicited by noise stimuli are grouped to the *right*. Predictions based on the MID filters are shown with white bars, those based on the dSTA are shown in black, and those based on the RdSTA are shown in gray. *Top*: average predictions as a percentage of the overall information explained by LN models with the different filters and the best possible nonlinearities for these filters. *Bottom*: analogous comparisons using the percentage of variance explained by the LN models based on the same 3 different filters and nonlinearities. Error bars show SEs.

tween percentages of explained variance using either the dSTA or the RdSTA).

The same results were also obtained using a larger fraction of the data (1/4 instead of 1/8) as the test data set for computing jackknife estimates for each kind of a filter. In this case, the MID filters derived from natural stimuli accounted for $31 \pm 3.6\%$ (SE) of I_{spike} when predicting neural responses to a novel set of natural stimuli, which was significantly better than predictions based on the RdSTA ($26.6 \pm 3.5\%$, $P = 0.015$) and those based on the dSTA ($11 \pm 1.6\%$, $P < 10^{-4}$). Similarly, predictions of neural responses to noise stimuli using the MID filters derived from natural stimuli accounted for $18.09 \pm 3.3\%$ and were significantly better ($P < 10^{-4}$) than the corresponding predictions based on the RdSTA filters derived from natural stimuli ($5.3 \pm 1.0\%$) or the dSTA ($4.2 \pm 1.3\%$).

DISCUSSION

Stability of the LN model

Although characterizing neural processing in the natural setting remains one of the ultimate goals of sensory neuroscience, considerable technical difficulties exist for correctly estimating neural receptive fields from natural stimuli (Paninski 2003a; Rust and Movshon 2005; Sharpee et al. 2003, 2004). Here we have examined in detail two models for computing spatiotemporal filters: the linear model and the LN model. With noise inputs, there was very little difference, if any, between the spatiotemporal filters of the two models. This is as expected theoretically for white Gaussian or other uncorrelated (spherically symmetric) stimuli (Chichilnisky 2001; de Boer and Kuyper 1968; Paninski 2003a; Rieke et al. 1997; Sharpee et al. 2004). In contrast, spatiotemporal filters of the linear and the LN models computed with natural stimuli can be profoundly different (Fig. 5). Despite the added complexity in the LN model, it produces spatiotemporal filters that are more stable under changes in the stimulus statistics between noise and natural inputs (Fig. 6). Spatiotemporal filters obtained using the LN model also better predicted spikes elicited by a novel segment of either noise or natural stimuli (Fig. 7), even though predictions based on the spatiotemporal filters computed using the linear model also took into account the best possible nonlinearity for those filters.

The two standard methods of computing spatiotemporal filters of the linear model with natural stimuli have their limitations. As pointed out previously (David et al. 2004; Smyth et al. 2003; Theunissen et al. 2000, 2001; Touryan et al. 2005; Woolley et al. 2006), computing the spatiotemporal filter as the dSTA tends to amplify noise at spatial and temporal frequencies that are relatively undersampled. For natural stimuli, this noise amplification happens at higher temporal and spatial frequencies. In agreement with previous results (David et al. 2004; Smyth et al. 2003; Theunissen et al. 2000, 2001; Touryan et al. 2005), we show here that introducing smoothness constraints and regularization leads to greater predictive power of the RdSTA spatiotemporal filter compared with that of the dSTA. In some cases, however, the RdSTA can produce spatiotemporal filters that substantially overestimate the contribution of low-frequency components (Fig. 5) to neural filtering. In this respect, spatiotemporal filters of the LN model computed as the MID seem to find the middle ground: at low

spatial and temporal frequencies they are similar to the dSTA filters, whereas at higher spatial frequencies their amplitude spectra are intermediate between those of the dSTA and RdSTA filters. At these higher spatial frequencies the MID filters computed for natural and noise stimuli had identical amplitude spectra, which provided additional support for the computations of the LN model.

Theoretical arguments indicate that advantages of computing the MID filters compared with the RdSTA filters should increase with increasing deviations of the stimulus ensemble from the correlated Gaussian distribution (Ringach et al. 1997; Sharpee et al. 2004) or when the neural noise level decreases (Sharpee 2007). One measure of the deviation from a Gaussian ensemble is the kurtosis of the distribution of light intensity values of individual pixels, which is zero for a Gaussian distribution but positive for distributions that are more heavytailed than Gaussian. The natural stimulus ensemble used in this study had a mean kurtosis value across pixels of about 0.4 (range from 0.19 to 0.64) measured for the distribution of light intensity at single pixels across approximately 50,000 frames. By comparison, one can expect to find kurtosis values <0.04 for a sequence of the same size taken from the uncorrelated Gaussian distribution (Press et al. 1992). Our kurtosis measurement is consistent with previous studies of the statistics of natural stimuli where kurtosis values with a mean of about 2 (range 0.2–22) were obtained after correcting for finite-size effects (Thomson 1999). Therefore although our stimulus ensemble is non-Gaussian, typical natural stimulus ensembles are even more strongly non-Gaussian by this measure.

Overall, these findings suggest that stimulus-dependent changes in neural spatiotemporal filters are better characterized by the LN model compared with the linear model. For example, both the dSTA and RdSTA filters computed for natural stimuli show changes in tuning across spatial and temporal frequencies relative to white noise filters, although these changes are generally of opposite sign for the dSTA versus the RdSTA. A point-by-point comparison between spatiotemporal filters, as quantified by the correlation coefficient, also shows that spatiotemporal filters of the linear model differed much more between noise and natural stimuli than did the spatiotemporal filters of the LN model.

Extensions of the LN model

The LN model considered here can be extended to account for multiple relevant stimulus features (Agüera y Arcas et al. 2003; Bialek and de Ruyter van Steveninck 2005; Brenner et al. 2000a; de Ruyter van Steveninck and Bialek 1988; Fairhall et al. 2006; Felsen et al. 2005b; Rust et al. 2005; Slee et al. 2005; Touryan et al. 2002, 2005). Our best predictions using the LN model with a single spatiotemporal filter accounted for about 40% of the overall information contained in the stimulus about the arrival times of single spikes, I_{spike}. Including additional dimensions has the potential to increase the percentage of I_{spike} explained by the model. David et al. (2004) used a model based on two linear filters passed through threshold-linear functions to account for 50% of the variance of responses in monkey V1 cells. Another possible extension is to account for spike history effects, in which the probability of a spike is influenced by prior spikes independent of visual stimulus (Paninski et al. 2004; Pillow et al. 2005).

In any case, it is helpful to use Shannon mutual information as a measure of predictive power, particularly in the setting of natural (non-Gaussian) stimuli, because it measures the degree to which model output accounts for neural response independently of the statistical distributions of these variables. For any distributions, it tells the number of bits of information given, on average, by the model output about the neural response. Although the percentage of variance explained provides an intuitive measure of the size of the error relative to the size of the response, it might not reflect deviations in accounting for the higher-than-second-order moments of the response.

Stimulus-dependent changes in neural spatiotemporal filters have been observed in several sensory modalities, including auditory (Theunissen and Shaevitz 2006; Theunissen et al. 2000, 2001; Woolley et al. 2006), visual (Chander and Chichilnisky 2001; David et al. 2004; Felsen et al. 2005a,b; Hosoya et al. 2005; Sharpee et al. 2006; Smirnakis et al. 1997), and somatosensory (Maravall et al. 2007). It is intriguing that such stimulus-dependent effects increase precision of temporal coding and are tuned to emphasize the most informative features of natural sounds (Theunissen and Shaevitz 2006; Woolley et al. 2005, 2006). In this respect, such stimulus-dependent tuning of auditory neurons is reminiscent of how visual neurons adapt their filtering properties to match the statistics of stimuli (Hosoya et al. 2005; Sharpee et al. 2006; Smirnakis et al. 1997), suggesting that there may be general principles for sensory processing common to different modalities. It will be interesting to know whether taking into account nonlinear transformations between stimuli and spike probabilities when estimating the neural spatiotemporal filter will similarly improve our models of neural coding in other stages of visual processing and sensory modalities, as it does for simple cells in the primary visual cortex.

REFERENCES

Adelman TL, Bialek W, Olberg RM. The information content of receptive fields. *Neuron* 40: 823–833, 2003.

Agüera y Arcas B, Fairhall AL. What causes a neuron to spike? *Neural Comput* 15: 1789–1807, 2003.

Agüera y Arcas B, Fairhall AL, Bialek W. Computation in a single neuron: Hodgkin and Huxley revisited. *Neural Comput* 15: 1715–1749, 2003.

Ahrens MB, Linden JF, Sahani M. Nonlinearities and contextual influences in auditory cortical responses modeled with multilinear spectrotemporal methods. *J Neurosci* 28: 1929–1942, 2008.

Bialek W, de Ruyter van Steveninck RR. Features and dimensions: motion estimation in fly vision. Available online at arxiv.org/abs/q-bio.NC/0505003, 2005.

Brenner N, Bialek W, de Ruyter van Steveninck R. Adaptive rescaling maximizes information transmission. *Neuron* 26: 695–702, 2000a.

Brenner N, Strong SP, Koberle R, Bialek W, de Ruyter van Steveninck RR. Synergy in a neural code. *Neural Comput* 12: 1531–1552, 2000b.

Bussgang JJ. Crosscorrelation functions of amplitude-distorted Gaussian signals. In: *Research Laboratory of Electronics, Technical Report 216.* Cambridge, MA: MIT Press, 1952.

Chander D, Chichilnisky EJ. Adaptation to temporal contrast in primate and salamander retina. *J Neurosci* 21: 9904–9916, 2001.

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

Christianson GB, Sahani M, Linden JF. The consequences of response nonlinearities for interpretation of spectrotemporal receptive fields. *J Neurosci* 28: 446–455, 2008.

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

de Boer E, Kuyper P. Triggered correlation. *IEEE Trans Biomed Eng* 15: 169–179, 1968.

de Ruyter van Steveninck RR, Bialek W. 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, 1988.

Efron B, Tibshirani RJ. *An Introduction to Bootstrap.* Boca Raton, FL: Chapman & Hall/CRC Press, 1998.

Emondi AA, Rebrik SP, Kurgansky AV, Miller KD. Tracking neurons recorded from tetrodes across time. *J Neurosci Methods* 135: 95–105, 2004.

Fairhall AL, Burlingame CA, Narasimhan R, Harris RA, Puchalla JL, Berry MJ 2nd. Selectivity for multiple stimulus features in retinal ganglion cells. *J Neurophysiol* 96: 2724–2738, 2006.

Felsen G, Dan Y. A natural approach to studying vision. *Nat Neurosci* 8: 1643–1646, 2005.

Felsen G, Touryan J, Dan Y. Contextual modulation of orientation tuning contributes to efficient processing of natural stimuli. *Network* 16: 139–149, 2005a.

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

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

Hosoya T, Baccus SA, Meister M. Dynamic predictive coding by the retina. *Nature* 436: 71–77, 2005.

Lesica NA, Jin J, Weng C, Yeh CI, Butts DA, Stanley GB, Alonso JM. Adaptation to stimulus contrast and correlations during natural visual stimulation. *Neuron* 55: 479–491, 2007.

Machens CK, Wehr MS, Zador AM. Linearity of cortical receptive fields measured with natural sounds. *J Neurosci* 24: 1089–1100, 2004.

Maravall M, Petersen RS, Fairhall AL, Arabzadeh E, Diamond ME. Shifts in coding properties and maintenance of information transmission during adaptation in barrel cortex. *PLoS Biol* 5: e19, 2007.

Paninski L. Convergence properties of three spike-triggered analysis techniques. *Network* 14: 437–464, 2003a.

Paninski L. Estimation of entropy and mutual information. *Neural Comput* 15: 1191–1253, 2003b.

Paninski L, Pillow JW, Simoncelli EP. Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. *Neural Comput* 16: 2533–2561, 2004.

Pillow JW, Paninski L, Uzzell VJ, Simoncelli EP, Chichilnisky EJ. Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *J Neurosci* 25: 11003–11013, 2005.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical Recipes in C.* Cambridge, UK: Cambridge Univ. Press, 1992.

Rieke F, Warland D, de Ruyter van Steveninck R, Bialek WB. *Spikes: Exploring the Neural Code.* Cambridge, MA: MIT Press, 1997.

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

Ringach DL, Hawken MJ, Shapley R. Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. *J Vis* 2: 12–24, 2002.

Ringach DL, Sapiro G, Shapley R. A subspace reverse correlation technique for the study of visual neurons. *Vision Res* 37: 2455–2464, 1997.

Ruderman DL, Bialek W. Statistics of natural images: scaling in the woods. *Phys Rev Lett* 73: 814–817, 1994.

Rust NC, Movshon JA. In praise of artifice. *Nat Neurosci* 8: 1647–1650, 2005.

Rust NC, Schwartz O, Movshon JA, Simoncelli EP. Spatiotemporal elements of macaque v1 receptive fields. *Neuron* 46: 945–956, 2005.

Sahani M, Linden JF. How linear are auditory cortical responses? In: *Advances in Neural Information Processing Systems,* edited by Becker S, Thrun S, Obermayer K. Cambridge MA: MIT Press, vol. 15, 2003, p. 125–132.

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

Sharpee T, Rust NC, Bialek W. Maximally informative dimensions: analyzing neural responses to natural signals. In: *Advances in Neural Information Processing Systems*, edited by Becker S, Thrun S, Obermayer K. Cambridge, MA: MIT Press, vol. 15, 2003, p. 261–268.

Sharpee T, Rust NC, Bialek W. Analyzing neural responses to natural signals: maximally informative dimensions. *Neural Comput* 16: 223–250, 2004.

Sharpee TO. Comparison of information and variance maximization strategies for characterizing neural feature selectivity. *Stat Med* 26: 4009–4031, 2007.

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

Simoncelli EP, Olshausen BA. Natural image statistics and neural representation. *Annu Rev Neurosci* 24: 1193–1216, 2001.

Skottun BC, De Valois RL, Grosof DH, Movshon JA, Albrecht DG, Bonds AB. Classifying simple and complex cells on the basis of response modulation. *Vision Res* 31: 1079–1086, 1991.

Slee SJ, Higgs MH, Fairhall AL, Spain WJ. Two-dimensional time coding in the auditory brainstem. *J Neurosci* 25: 9978–9988, 2005.

Smirnakis SM, Berry MJ, Warland DK, Bialek W, Meister M. Adaptation of retinal processing to image contrast and spatial scale. *Nature* 386: 69–73, 1997.

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

Strong SP, Koberle R, de Ruyter van Steveninck RR, Bialek W. Entropy and information in neural spike trains. *Phys Rev Lett* 80: 197–200, 1998.

Theunissen FE, David SV, Singh NC, Hsu A, Vinje WE, Gallant JL. Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network* 12: 289–316, 2001.

Theunissen FE, Sen K, Doupe AJ. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci* 20: 2315–2331, 2000.

Theunissen FE, Shaevitz SS. Auditory processing of vocal sounds in birds. *Curr Opin Neurobiol* 16: 400–407, 2006.

Thomson MG. Visual coding and the phase structure of natural scenes. *Network* 10: 123–132, 1999.

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

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

Treves A, Panzeri S. The upward bias in measures of information derived from limited data samples. *Neural Comput* 7: 399–407, 1995.

Usrey WM, Sceniak MP, Chapman B. Receptive fields and response properties of neurons in layer 4 of ferret visual cortex. *J Neurophysiol* 89: 1003–1015, 2003.

Webster MA, Georgeson MA, Webster SM. Neural adjustments to image blur. *Nat Neurosci* 5: 839–840, 2002.

Webster MA, Mizokami Y, Svec LA, Elliott SL. Neural adjustments to chromatic blur. *Spat Vis* 19: 111–132, 2006.

Woolley SM, Fremouw TE, Hsu A, Theunissen FE. Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat Neurosci* 8: 1371–1379, 2005.

Woolley SM, Gill PR, Theunissen FE. Stimulus-dependent auditory tuning results in synchronous population coding of vocalizations in the songbird midbrain. *J Neurosci* 26: 2499–2512, 2006.