INTRODUCTION

Several studies have shown that the receptive fields (RFs) of visual neurons exhibit various types of spatio-temporal inseparability. For example, the inseparability of spatial phase and time is believed to contribute significantly to direction selectivity (Jagadeesh et al. 1993, 1997; Pribe and Ferster 2005; Reid et al. 1987, 1991). Other forms of RF inseparability, including the dynamic sharpening of orientation, spatial-frequency, and binocular disparity tuning functions, have been recently reported (Bredfeldt and Ringach 2002; Chen et al. 2005; Frazor et al. 2004; Menz and Freeman 2003, 2004a,b; Ringach et al. 1997, 2003; Shapley et al. 2003; Xing et al. 2005). A common feature of these data is the fact that the image content at a coarse spatial scale dominates the early part of the response, whereas finer spatial scales dominate later phases of the response. The temporal dependence of the response on spatial scale is consistent with a coarse-to-fine processing of the retinal image. We asked if these features of the responses could arise from the dynamics of RFs early in the visual cortex.

In this report, we describe the dynamics of simple-cell RFs in the primary visual cortex of macaque monkeys by measuring the first-order kernels of simple-cell RFs with subspace reverse correlation (Ringach et al. 1997). The cells’ first-order spatio-temporal kernels were analyzed by fitting two-dimensional Gabor functions, defined as the product of a sinusoidal grating and a Gaussian envelope, at various time delays. We found that simple-cell RFs “shrink” over the time-course of the neural response: the spatial extent of the Gaussian envelope (length and width) and the spatial period of the sinusoidal grating tend to decrease over time. It is possible that the coarse-to-fine processing seen in downstream calculations based on the outputs of simple cells would inherit, to some extent, the tendency to process information from coarse-to-fine spatial scales.

METHODS

Preparation, recording, and optics

All experiments were approved by the UCLA Animal Research Committee and were carried out following National Institutes of Health’s Guidelines for the Care and Use of Mammals in Neuroscience. Acute experiments were performed on anesthetized and paralyzed adult Old-World monkeys (Macaca fascicularis). Initially, the animal was sedated with acepromazine (30–60 μg/kg), anesthetized with ketamine (5–20 mg/kg, im) in the cage, and transported to the surgical suite. Initial surgery and preparation were performed under isoflurane (1.5–2.5%). Two intravenous lines were put in place. A urethral catheter was inserted to collect and monitor urine output, and an endotracheal tube was inserted to allow for artificial respiration. All surgical cut-down sites were infused with local anesthetic (xylocaine 2%, sc). Pupils were dilated with ophthalmic atropine, and custom-made gas permeable contact lenses were fitted to protect the corneas. After this initial surgery, the animal was transferred to a stereotaxic frame. At this point, anesthesia was switched to a combination of sufentanil (2–6 μg/kg/h) and midazolam, or sufentanil (0.15 μg/kg/h) and propofol (2–6 mg/kg/h). We proceeded to perform a craniotomy over primary visual cortex. The animal was paralyzed (pavulon, 0.1 mg/kg/h) only after all surgical procedures, including the insertion of the electrode arrays, were complete.

To ensure a proper level of anesthesia throughout the duration of the experiment, rectal temperature, heart rate, noninvasive blood pressure, end-tidal CO2, SpO2, and EEG were continually monitored by an HP Virida 24C neonatal monitor. Urine output and specific gravity were measured every 4–5 h to ensure adequate hydration. Drugs were administered in balanced physiological solution at a rate to maintain a fluid volume of 5–10 ml/kg/h. Rectal temperature was maintained by a self-regulating heating pad at 37.5°C. Expired CO2 was maintained between 4.5 and 5.5% by adjusting the stroke volume and ventilation rate. The maximal pressure developed during the respiration cycle was monitored to ensure that there was no increment in the airway. A broad spectrum antibiotic (bicillin, 50,000 IU/kg) and anti-inflammatory steroid (dexamethasone, 0.5 mg/kg) were given at the beginning of the experiment and every other day.

In some experiments, a 10 × 10 electrode array (Cyberkinetics, Salt Lake City, UT) with 1- or 1.5-mm-long electrodes was implanted in primary visual cortex. The center of the array was aimed at 6 mm posterior to the lunate sulcus and 8 mm lateral to the midline. In other experiments, extracellular action potentials were recorded with an array of independently movable glass-coated tungsten microelectrodes with exposed tips of 5–15 μm. Electrical signals were amplified.

Address for reprint requests and other correspondence: D. Ringach, Dept. of Neurobiology and Psychology, Franz Hall, Rm 7613, Univ. of California, Los Angeles, CA 90095-1563 (E-mail: dario@ucla.edu).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
and spikes were discriminated using a Cerebus 128-channel system (Cyberkinetics). Spike sorting was performed offline using principal component analysis on the waveform shapes with software developed in our laboratory.

Stimuli were generated on a Silicon Graphics O2 and displayed on a monitor at a refresh rate of 100 Hz and a typical screen distance of 80 cm. The mean luminance was 60 cd/m². A Photo Research Model 703-PC spectro-radiometer was used for calibration. The eyes were initially refracted by direct ophthalmoscopy to bring the retinal image into focus for a stimulus roughly 80 cm from the eyes. Once neural responses were isolated, we measured spatial frequency tuning curves and maximized the response at high spatial frequencies by changing external lenses in steps of 0.25 D. This procedure was performed independently for both eyes.

Kernel estimation, model fitting, and data selection

The spatio-temporal RF of simple cells was measured by means of a subspace reverse correlation method (Ringach et al. 1997). The visual stimulus is a sequence of luminance modulated sinewave gratings varying in orientation, spatial frequency, and spatial phase, presented in pseudorandom order at an effective rate of 50 Hz (each frame is repeated twice). Each image in the stimulus set consisted of a Hartley basis function of size M by M pixels, such that $H_{k_x,k_y}(l,m) = \text{cas}[2\pi(l + k_x m)/M]$ for all $0 \leq l, m \leq M - 1$. Here, $\text{cas} = \sin \alpha + \cos \alpha$, and $k_x$ and $k_y$ represent the spatial frequency of the grating in units of cycles per stimulus side. The stimulus set consisted of all Hartley basis functions $\{ \pm H_{k_x,k_y}\}$ such that $|k_x| \leq k_{\text{max}}$ and $|k_y| \leq k_{\text{max}}$. The maximum spatial frequency for the Hartley basis set was chosen to exceed the range of spatial frequencies that elicited robust responses during initial characterizations with drifting gratings. Spatial frequency in these experiments varied from 1.8 to 10.7 cycles/°. The number of unique images in the stimulus set varied from 336 to 3,360, with a mean of 1,057.3; with four spatial phases for each orientation/spatial frequency, this means that the typical experiment had roughly 260 unique combinations of spatial frequency and orientation. The stimulus contrast in all experiments was considered to be poorly fit, and the data were discarded. The kernels were required to have a high SNR ($>3$), and the entirety of the kernel was required to fall within the stimulus window.

The optimal time delay was defined as the time lag at which the kernel variance peaked. Early and late times were defined as the points closest to this optimal time such that the variance was equal to one half its peak value. The distributions of early, optimal, and late delays are shown in Fig. 1B. Across the sampled population, the medians for the early, optimal, and late times were 44, 54, and 66 ms, respectively. Thus the dynamical effects described below occurred over a time scale of ~20 ms. We refer to the interval bracketed by the early and late delay times as the analysis interval.

To parameterize our measurements of RF size, we fit a Gabor model to the RF kernel obtained from each neuron at each delay, such that for a given delay, $\tau$, RF$(x,y) = A \exp(-x^2/2\sigma_x - y^2/2\sigma_y) \cos(2\pi r/\lambda + \Phi)$, allowing also for an arbitrary rotation and translation of the axes. Two-dimensional Gabor functions consisting of the product of a sinusoid with a Gaussian envelope have long been used to fit simple-cell RFs (DeAngelis et al. 1991, 1993a,b; Field and Tolhurst 1986; Jones and Palmer 1987). The parameters most relevant to the spatial scale of visual processing are length, width, and period. In the Gabor model, the parameters determining RF size—the length ($\sigma_x$) and width ($\sigma_y$) of the Gaussian envelope—are independent of the spatial period of the sinusoid ($\lambda$). Because the values for length and width are returned in terms of the SD of the spatial envelope, the effective RF size can be estimated as roughly $4\sigma$ along each dimension, because 95.45% of the variance of a Gaussian occurs within 2 SD of the center. The mean values for $\sigma_x$ and $\sigma_y$ across the population were 0.269 and 0.268° of visual angle, respectively, yielding corresponding means for the total RF width and length of 1.08 and 1.07°.

We independently fitted two-dimensional Gabor functions to temporal slices of the spatio-temporal kernel by least-squares error minimization. This was done by first fitting the kernel with peak variance and using these parameters as the initial starting point for the parameters at other times within the analysis interval. Optimal parameter values and 95% CIs were computed using Matlab’s nlinfit and nlparci functions. In cases where the width of the CIs exceeded the value of the parameter being estimated within the analysis interval, the kernel was considered to be poorly fit, and the data were discarded. The criteria for inclusion in the dataset were blind to the RF properties of the neuron, other than the signal-to-noise criterion stated above. If the

![FIG. 1. Description of experimental methodology. A: blue curve depicts kernel variance as a function of delay relative to spike occurrence for an example neuron. Signal-to-noise ratio (SNR) value used as a criterion for inclusion in dataset was defined as ratio of peak variance (“optimal” delay at 2) to average of variance in “noise frames” at delays >150 ms. Color panels depict kernels obtained at the early (1), optimal (2), and late (3) delays respectively. B: histograms indicate distributions of early, optimal, and late delays, respectively.](image-url)
CIs for a given parameter estimate at the early and late delays were nonoverlapping, that parameter was considered to have varied significantly over the time-course of the response \( (P < 0.05; \text{Seber and Wild } 1989) \).

We did not observe a strong \( (P < 0.01) \) dependence of changes in RF width, length, or spatial period on the absolute values of these parameters when computed in degrees of visual angle. Because we were primarily interested in relative changes in the model parameters, the parameters for width, length, and period at different time lags were normalized by their values at the optimal delay. When computing the population curves (Fig. 3), we compensated for the variable duration of the analysis intervals in different cells by linearly stretching or compressing the time axis. Because the variance profiles were more consistent in terms of shape than duration, this procedure can be thought of as normalizing the parameter curves at points of similar variance, despite warping them in time. Given this adjustment, all cells contribute to the population composite curves at all points, rather than the cells with the longest analysis intervals dominating the earliest and latest intervals of composites based on simple peak alignment. An alternative analysis where we aligned all responses by the peak in the variance profile provided essentially the same results.

**RESULTS**

Examples of changes in RF size in simple cells of V1 are depicted for three neurons in Fig. 2. The kernels appear in the top row of each set of panels, and the accompanying Gabor fits appear along the bottom row (Fig. 2, A, C, and E). The kernels at the early delay appear larger than those at the late delay. In particular, there seems to be a progressive narrowing of the RF kernel along the axis of preferred orientation, which reflects a decrease in both the RF width and the spatial period. The accompanying plots in Fig. 2, B, D, and F, show how the fitted values for width, length, and period vary as a function of delay. The blue curves indicate the fitted parameter values, within the 95% CIs in green. Note that the changes in these parameters tend to be monotonically decreasing in time and not related directly to the amplitude of the kernels. This finding rules out an “iceberg effect” in the measurements, which would predict an increase in size followed by a decrease in size (Shevelev et al. 1992).

To examine population trends in the dynamics of RF organization, we computed population averages of the parameters (Fig. 3). To facilitate comparisons across cells, we normalized the values of the fit parameters at each time delay by the value obtained at the variance peak (see METHODS). The black line indicates the population mean, and the gray lines indicate \( \pm 2 \text{ SE} \). Because the parameter value at the optimal delay is one for all cells, the SE is necessarily smallest near the center of the graph, which is near the variance peak for all neurons (there is some variation in the peak location because the variance profiles are not perfectly symmetric). The RF width (Fig. 3A), RF length (Fig. 3B), and spatial period (Fig. 3C) showed
For individual cells, we defined a change in one of the parameters as being significant if their CIs at the early and late time lags were nonoverlapping. Significant changes in width, length, and period are indicated by the color-coding of the data in Fig. 4, where red signifies a decrease, green indicates an increase, and black indicates no significant changes. To verify that temporal trends were consistent across the first phase of the response (from early to optimal times) and the second phase of the response (from optimal to late times), we plotted the ratio between the fitted parameters at the early and optimal delays versus the ratio between the optimal and late delays (Fig. 4, A, C, and E). In these scatterplots, a reduction in a parameter from the early to optimal time will result in values above one along the x-axis; a reduction from the optimal to the late time will result in values above one along the y-axis. Thus the clustering of points in the first quadrant of the scatterplots indicates that there is a consistent reduction of the parameters across the first and second phase of the response. The data indicate that RF width, length, and period show a consistent trend toward a decrease over time.

The overall ratio between the early and late time lags in the spatial parameters of the Gabor fits are shown in the histograms of Fig. 4, B, D, and F. The decrease in median RF widths from early to late delays was highly significant (Fig. 4B; Wilcoxon rank sum; $P < 1.5 \times 10^{-7}$). The mean of the early/late ratio for RF width was 1.16 (median = 1.11). The reduction in median RF length was also statistically significant across the population (Wilcoxon rank sum; $P < 2.7 \times 10^{-5}$), although it was less pronounced than that observed for width: the mean length ratio (early/late) was 1.08 (median = 1.06). Similarly, there was a significant decrease in the median spatial period as well (Fig. 4F; Wilcoxon rank sum; $P < 5.1 \times 10^{-11}$). The mean ratio from the early to late delays in period was 1.15 (median = 1.11). For completeness, we also computed the change in RF area by taking the product of the fitted length and width parameters and comparing this product at the early and late decay times. The mean reduction in RF area for the population was 25.5%, indicating that the window through which V1 simple cells view the visual scene decreases substantially (Wilcoxon rank sum; $P < 1.9 \times 10^{-8}$).

The number of significant increases/decreases for width, length, period, and amplitude are 11/42, 12/37, 14/45, and 12/40, respectively (because the early and late delays were matched in terms of kernel variance, the amplitude of the Gabor fit increases from early to late delays because equivalent kernel variance is focused in a smaller area). As one would expect from symmetry considerations, the number of significant increases/decreases in horizontal location (an increase indicates a rightward shift), vertical location (an increase indicates a downward shift), spatial phase, and orientation are 27/22, 16/25, 27/24, and 23/24, respectively. There were no consistent trends at the population level ($P > 0.1$) for these parameters.

The effective number of subregions in each kernel is proportional to the ratio between the RF width and the spatial period, which we define as the normalized width. The median normalized width (Fig. 5A) did not vary significantly from the early to late delay (mean = 1.03; median = 1.03; Wilcoxon rank sum; $P > 0.52$). The median normalized length (Fig. 5B), defined along the same lines, increased slightly (mean = 1.04; median = 1.06; Wilcoxon rank sum; $P = 0.003$). This reflects consistent decreases throughout the analysis interval, indicating that the envelope of the Gaussian of the Gabor model decreased along both dimensions while the spatial period of the underlying sinusoidal component also decreased. We did not see consistent changes in other parameters of the fits, such as RF center and orientation (data not shown).
the fact that changes in length were smaller on average than the changes in spatial period. For similar reasons, the aspect ratio of the RF (Fig. 5C), defined as the ratio of the length to the width, showed a modest (Wilcoxon rank sum; \( P = 0.011 \)) median increase, such that the aspect ratio at the late delay was, on average, 1.08 times as large as the aspect ratio at the early delays. Thus the greater reduction of RF width during the analysis interval causes a progressive narrowing of the RF.

Given the simultaneous reduction in all spatial parameters of the RF, it is natural to ask whether, on individual cells, the changes could be accounted for by a simple scaling of the entire RF. This mechanism would predict that there should be correlations between the observed changes in length, width, and period on a cell-by-cell basis. However, we found no significant correlations in the changes of these parameters (Fig. 6). These results do not support the hypothesis that a single scaling mechanism explains the RF dynamics we observed.

**DISCUSSION**

We showed a type of spatio-temporal inseparability in the first-order kernels of simple cortical cells that has not previously been described in primates. The data show that the envelope and period of the carrier in Gabor fits to the RFs of simple cells decrease over the time of their responses. This phenomenon may underlie, at least partially, the dynamic changes in the peak spatial frequency over time (Bredfeldt and Ringach 2002; Frazor et al. 2004; Mazer et al. 2002) and the sharpening of orientation tuning in some cells (Ringach et al. 1997, 2003; Shapley et al. 2003). In contrast with these reports, which averaged responses over spatial phase and estimated tuning in the Fourier domain, we constructed first-order kernels in the spatial domain, allowing us to show that changes in spatial period occur in parallel with changes in RF size. This is the first demonstration that RF size varies dynamically in macaque V1.

Our findings are partially consistent with previous studies of V1 RFs in anesthetized cats (Suder et al. 2002; Worgotter et al. 1998). Using a stimulus set consisting of flashed dots or bars (bright and dark), Suder et al. (2002) observed both reductions in RF width of ~23% and proportionately greater reductions in width than length. They attributed the shrinkage of V1 RFs to the transition from phasic to tonic input from the lateral geniculate nucleus based on comparisons of RF widths at latencies of 70 versus 200 ms relative to stimulus onset. In contrast, the median late decay time was 66 ms in our population, indicating that the changes in RF size we observed were essentially complete before the interval they identify with the early response had ended. These differences in the observed time-course of RF size changes likely reflect important stimu-
lus differences between the studies. For example, the relatively long stimulus duration (300 ms) and the isolated presentation of individual dots or bars on a uniform background are likely to promote robust onset responses. Because their RF reconstruction technique preserves the shapes of the peristimulus time histogram, nonlinear mechanisms related to recent response history are also preserved and can affect the measured RF width for a given time interval. Our method, which is based on spike-triggered averaging, effectively averages out such response history effects, so the RF dynamics we report are restricted to the linear kernel. As a result, our data show that V1 RFs can exhibit significant reductions in size and period even when the transition from phasic to tonic firing in the LGN is obviated by rapid, sequential stimulus presentation.

The method we used is closer to that of Menz and Freeman (2004a), who used a rapidly presented binary m-sequence of one-dimensional bars to characterize both the monocular and binocular RFs of V1 neurons in anesthetized cats. They observed a dichotomy in the dynamics of binocular and monocular RFs of simple cells, such that the average size of monocular RFs increased, whereas the disparity range for binocular RFs decreased within a 40-ms window. It is possible that the discrepancy between the RF shrinkage we observed and the RF expansion they obtained could reflect differences in stimulus presentation. In this study, all stimuli were presented monocularly. Menz and Freeman (2004a) also reported substantial reductions in the size of RF centers in the LGN when monocular stimuli were used, but the monocular RFs for cortical neurons were obtained by selective cross-correlation with dichoptic stimuli. Species differences in the temporal dynamics of RF structure between cats and monkeys should also be considered. For example, DeAngelis et al. (1993a) found that the optimal delay for RF kernels obtained in cats using a sparse noise stimulus was 68.2 ms, which is slightly larger than the median late delay of 66 ms we observed in the macaque. The average width of the temporal envelope (measured at 0.367 of the envelope peak) was 139.6 ms, which is substantially longer than what we observed (the equivalent calculation yields a mean of 29 ms and a median of 25 ms). Nevertheless, based on these differences, one might expect that RF dynamics would occur more rapidly in the monkey but not in a different direction. Thus it seems more likely that the discrepancy is explained by stimulus differences.

There are a number of potential, nonexclusive mechanisms that could account for the RF dynamics we observed. A central question that needs more study is the degree to which RF size changes in V1 are inherited from RF size changes occurring in the LGN and even the retina. In this regard, it is interesting that Suder et al. (2002) observed similar relative changes (roughly 18% vs. 23%) in the widths of LGN RFs, but concluded that such changes could not explain the absolute changes (0.2 vs. 2°) observed in V1. Another potential explanation that relies solely on the integration of geniculate inputs is that cortical cells may pool parvocellular and magnocellular inputs (cell classes known to have different latencies and receptive field sizes) to generate the observed dynamics (Frazor et al. 2004; Mazer et al. 2002). Frazor et al. (2004) put forward such a model to explain dynamic changes in spatial frequency dynamics. In the cat, even within the same cell class, there is a correlation between response latency and RF size (Weng et al. 2005). Thus it may not be necessary to invoke precise pooling of different inputs to obtain the observed dynamics in cortical cells. Another possibility is that intracortical connectivity contributes to the dynamics of simple cell RFs (Sompolinsky and Shapley 1997). Center-surround interactions across simple cells in the cortex may work to reduce receptive field size and decrease the spatial period (Sabatini 1996; Sabatini et al. 1997). If intracortical feedback is a main mechanism for size dynamics, one would expect RF

FIG. 6. Scatterplots showing correlations among scale-related RF changes. Lines of best fit on each plot were calculated with robust linear regression to minimize effects of outliers. Significance values shown at bottom right were calculated for slopes of regression line, such that a significant value indicates a slope other than 0.
shrinkage to be abolished by inactivation of the cortex (Ferster et al. 1996). It is also possible that RF dynamics may depend on the local structure of the orientation map where the simple cells are embedded (Sabatini 1996; Sabatini et al. 1997; Schummers et al. 2002).

The dynamics of V1 RFs we observed are consistent with a coarse to fine progression in visual processing. It has been argued on theoretical grounds that having low-frequency visual information available early could aid in constraining the analysis of higher frequency visual information, for example, by reducing “false matches” in the context of stereopsis (Marr and Poggio 1979), and there is physiological evidence of mechanisms working along these lines in binocular cells (Menz and Freeman 2003, 2004a,b). Many computer vision algorithms use coarse-to-fine image representations (Adelson and Burt 1982; Koendruck 1984; Witkin 1983), and there is even behavioral evidence that human subjects are less able to distinguish full-bandwidth images from spatially filtered images presented in a low-pass to highpass sequence than images presented from highpass to low-pass (Parker et al. 1997; see Morrison and Schyns 2001 for a critical discussion of coarse-to-fine image processing). The relevance of the physiological effects we and others have observed to psychophysical phenomena will likely depend on their relative time-courses and the range of scale changes. For example, the experiments of Parker (1992, 1997) on a coarse-to-fine progression in the processing of visual scenes used three images of 40-ms duration presented sequentially for 120 ms. When the spectral content of the images was low to high, rather than high to low, subjects rated the picture quality higher and were more likely to claim that a full-bandwidth image was present in the sequence. In our cell population, the typical duration of the analysis interval was between 20 and 25 ms, but it is highly probable that the size dynamics we observed occur over a longer interval (we chose to confine our analysis to delays where the signal to the noise of the kernel was greater than one half the maximum to ensure highly reliably for the fitting procedure). It is also possible that the size dynamics observed in V1, if they are inherited by downstream processing centers, could be spread over a longer time-course, and one more in keeping with psychophysical results. For example, Watt (1987) found improvements in sensitivity for the discrimination of the stereoscopic depth, length, orientation, and curvature of short line segments over a time course of >1,000 ms. Although such effects may have their origins in the RF dynamics of simple cells, additional visual processing is almost certainly necessary to account for increases in performance for viewing durations in excess of hundreds of milliseconds. Finally, the magnitude of spatial scale changes observed in individual cells is rather small compared with the changes needed for a full scale space analysis of the image as done in computer vision (Koendruck 1984; Witkin 1983).

Functionally, changes in the spatial scale of V1 RFs is to achieve a “sparsification” of the neural response. Willmore et al. (2000) have shown that spatial filters derived from principle component analysis of visual images are sparser when their RFs are smaller, for example. We conducted numerical simulations (data not shown) indicating that reductions of RF size of the magnitude we observed can produce a statistically significant reduction in the pairwise correlations among a set of Gabor RF models. However, the extent to which reducing RF size also reduces those correlations depends on a number of factors, such as the spatial overlap and heterogeneity of the RFs themselves. As a result, our simulations suggest that the degree of sparsification produced by changes in RF size will depend on the distribution of RFs among a given network of V1 neurons. The general notion is consistent with the recent demonstration that response profiles of individual neurons in V2 become progressively decorrelated at later stages of the response (Hegde and Van Essen 2004). Because many visual computations rely on the output of simple cells, RF size dynamics could potentially contribute to similar effects in higher visual centers.

GRANTS
This work was supported by National Research Service Award EY-015365-01, National Eye Institute Grant EY-12816, and DARPA HR0011-05-1-0026.

REFERENCES


Witkin A. Scale-space itering. Proceedings of the 8th International Joint Conference on Artificial Intelligence, Karlsruhe, FRG, August 1983.
