Neuron
Article

Adaptation to Natural Binocular Disparities in Primate V1 Explained by a Generalized Energy Model

Ralf M. Haefner[1,*] and Bruce G. Cumming[1]
[1]Laboratory for Sensorimotor Research, National Eye Institute/NIH, 49 Convent Drive, Building 49/2A50, Bethesda MD 20892, USA
*Correspondence: ralf.haefner@gmail.com
DOI 10.1016/j.neuron.2007.10.042

SUMMARY

Sensory processing in the brain is thought to have evolved to encode naturally occurring stimuli efficiently. We report an adaptation in binocular cortical neurons that reflects the tight constraints imposed by the geometry of 3D vision. We show that the widely used binocular energy model predicts that neurons dedicate part of their dynamic range to impossible combinations of left and right images. Approximately 42% of the neurons we record from V1 of awake monkeys behave in this way (a powerful confirmation of the model), while about 58% deviate from the model in a manner that concentrates more of their dynamic range on stimuli that obey the constraints of binocular geometry. We propose a simple extension of the energy model, using multiple subunits, that explains the adaptation we observe, as well as other properties of binocular neurons that have been hard to account for, such as the response to anticorrelated stereograms.

INTRODUCTION

Evolutionary pressures have acted on a brain processing inputs from the natural world. Natural inputs are not random but are structured in particular ways. Visual inputs have significant spatial correlations and a characteristic Fourier power spectrum ($1/f^2$). Auditory inputs, on the other hand, can be divided into separate classes, each with their own particular structure, e.g., vocalizations which tend to be harmonic and more steady-state or environmental sounds that are nonharmonic and more transient (Smith and Lewicki, 2006). One should therefore expect that instead of being equally capable of processing *any* inputs, the brain should have specialized to best deal with these natural stimuli. This is a key prediction of the efficient coding hypothesis (Attneave, 1954; Barlow, 1961), which has been assumed to be a general design principle for the brain. It states that one driving force behind the evolution of the brain is the goal to represent a maximum of information about the external world using a minimum of energy and neural resources.

Recent studies have provided evidence for this hypothesis in both the auditory (Smith and Lewicki, 2006) and visual domain (Field, 1987; Schwartz and Simoncelli, 2001; Simoncelli and Olshausen, 2001). In particular, they show that visual processing is specialized for the Fourier statistics of monocular natural images and exploits the redundancy introduced by the spatial correlation in natural images. In this paper, we investigate an even more striking redundancy that is a consequence of stereovision. Our two eyes view the world from slightly different vantage points. The resulting differences between the two eyes' images (binocular disparities) are used by the brain to create a three-dimensional percept of the world. At the same time, the images in the left and the right eye are highly redundant: in most cases local regions of the right image will be closely matched by simply translating the appropriate part of the left image. What we call a "simple translation" is a very specific subset of all the possible ways in which the image in the left eye could in principle differ from that in the right eye. In the laboratory, it is possible to create binocular images that employ all possible left-right differences: those that occur in the real world—natural ones (or *physical disparities*)—and artificial ones (*nonphysical disparities*). Formalizing these differences allows us to embed both physical and nonphysical stimuli within the same continuous stimulus space and hence to explore the response of binocular neurons within that space.

At the earliest point where inputs from the two eyes converge onto single neurons (striate cortex), the firing rate of many neurons depends upon binocular disparity (Barlow et al., 1967; Nikara et al., 1968; Poggio and Fischer, 1977). The responses of disparity selective neurons are well explained by the very successful binocular energy model (Cumming and DeAngelis, 2001; Ohzawa et al., 1990) in which the output from receptive fields in both eyes is linearly combined by V1 simple cells and this sum is then passed through an output nonlinearity. This simple scheme naturally places limits on the extent to which binocular regularities in the inputs can be exploited. We explore these limits using a simple stimulus that allows the gradual transition from physical to nonphysical stimuli to be studied: the sum of two sinusoidal luminance gratings.

First, we show that the binocular energy model predicts maximum responsiveness to nonphysical left and right image pairs. Second, we recorded responses of disparity-selective neurons in the striate cortex of awake monkeys. Approximately 42% of the neurons behave as the energy model predicts, with either

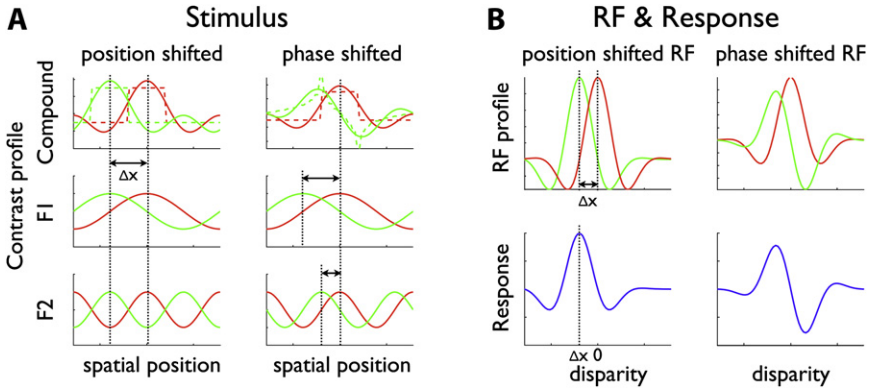

Figure 1. Geometric Constraints

(A) Luminance profiles of a compound grating consisting of two sine-wave gratings with frequency-ratio 1:2 are shown, as presented to the left eye (solid green) and the right eye (solid red). The left column shows the two patterns translated in space with respect to each other. This fixed spatial offset corresponds to a greater interocular phase difference for the higher frequency component than for the lower frequency one. (Ratio of interocular phase difference = frequency ratio.) The right column displays each component shifted by the same phase difference, resulting in different patterns in left and right eyes. The top left panel illustrates a physical stimulus, while the top right panel shows a nonphysical one which would not be produced by natural viewing. The dashed lines show the effects of phase or position shifts on the luminance profile of a bar. The phase shift in the top right column clearly leads to a change in shape (dashed green versus red in top right panel).

(B) The top two panels illustrate the principle differences between the receptive field (RF) profiles (left eye in green, right eye in red) of a binocular neuron which signals disparity: position shift (left) and phase shift (right). The bottom panels show the response of an energy model with such RFs to position disparity in a broadband stimulus: even-symmetric tuning curve for a neuron with position disparity and odd-symmetry for a neuron with phase disparity.

their maximum or their minimum response elicited by nonphysical combinations of interocular phase difference. The other 58% of neurons deviate from the model's predictions, and these deviations are systematically in the direction that exploits the regularities in the natural inputs, as predicted by the efficient coding hypothesis.

Finally, we show how this specialization for physical disparities can be explained by extending the existing standard model, if real neurons represent the sum of multiple energy model-like subunits. This shows that combining multiple subunits can be a useful strategy for exploiting the statistical regularities of visual inputs. It might thereby provide an explanation for another recent observation (Rust et al., 2005) that indicates the presence of multiple excitatory and suppressive spatiotemporal subunits within V1 receptive fields.

RESULTS

Energy-Model Predictions

The energy model is part of a family of generalized linear-nonlinear (LN) models describing the responses of V1 neurons (Chichilnisky, 2001; Rust et al., 2005). During the linear stage of an LN neuron, the 2D image is multiplied with the neuron's 2D receptive field, yielding a single scalar value that represents the effect of the image on the RF. The scalar value is then passed through an output nonlinearity (typically parameterized with an exponent), the result of which determines the spike rate of the neuron. In the binocular energy model, there are two separate receptive fields (RF), one for the left and one for the right eye, and their scalar responses are summed linearly, and this sum is then passed through an expansive output nonlinearity. The output nonlinearity means that the neuronal activity is enhanced when left and right RFs produce similar responses, and it is this that confers disparity selectivity on the cell. The top row of Figure 1B shows two examples for such receptive fields—green for the left and red for the right eye (1D cross-sections, the second dimension is typically well approximated by a Gaussian and can be safely ignored for the following discussion). The RFs are well described by Gabor functions. There are two principle ways in which the cell can encode disparity: (1) the left and right receptive fields are identical up to a translation ("position disparity"; Figure 1B top-left) or (2) receptive fields in the two eyes are related by a fixed phase shift applied to all frequency components ("phase disparity"; Figure 1B top-right) (Ohzawa et al., 1990; DeAngelis et al., 1991; Fleet et al., 1996). Let's take the example of a position shift. At the bottom left of Figure 1B we see the response (solid blue) to a broadband stimulus (stimuli in the real world are "broadband" in the sense that they consist of many Fourier components across a broad range of frequencies), with the displacement ("stimulus disparity") between the left and the right plotted along the x axis. The response peaks when the displacement between the images coincides with the displacement of the receptive fields. The model cell in the right column of Figure 1B (solid blue) has a phase disparity, and this produces a characteristically asymmetrical shape in the disparity tuning curve. Figure 1A illustrates the difference between position and phase shift in the stimulus. The solid lines in the top left panel in Figure 1A show the profiles of a compound grating consisting of two sine-wave gratings with frequency ratio 1:2, displayed individually in the panels below. The profile in the left eye (green) is identical to that in the right eye (red) apart from a positional displacement of Δx. It is important to note that as a consequence of an identical positional displacement Δx in both component gratings, the *phase* shift is different for each component—for a fixed position shift Δx, the phase shift $\Delta\varphi_i$ between the left and right Fourier component of frequency f_i is proportional to their frequency: $\Delta\varphi_i = \Delta x \cdot f_i$ (Fleet et al., 1996). In the right column of Figure 1A, on the other hand, the luminance profile in the left eye was phase shifted compared to the right eye (an identical interocular phase difference is added to each frequency component). However, as the two lower panels illustrate for F1 and F2, an identical phase shift implies different physical displacements. The implication for a broadband pattern can be seen in the top left and right panels of Figure 1A (dashed): while a position-shifted broadband pattern (here an example bar) is identical in the left and right eye up to translation, the

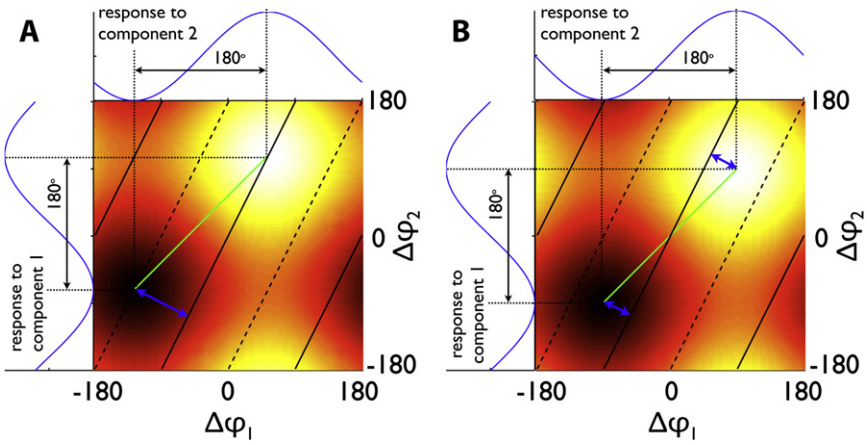

Figure 2. Energy-Model Response

In both panels, the energy-model response is shown in the parametric space spanned by varying the interocular phase difference for both component gratings independently from each other: on the x and y axis of this stimulus space we plot the interocular phase differences between the left and right eye for component 1, $\Delta\phi_1$, and component 2, $\Delta\phi_2$. Since the stimulus (hence the response) has a period of 360 degrees in both dimensions, we show it for [−180 deg …180 deg]. The solid black parallel lines are part of the same continuous line in this periodic space showing combinations of interocular phase difference that correspond to pure spatial translations of the compound grating stimulus. This line passes through the origin and has a slope of 2 due to the 1:2 frequency ratio in our stimulus. The dashed lines depict stimuli that are anticorrelated, i.e., those stimuli for which the contrast is reversed in one eye with respect to the other, prior to translation. The green line connects the response peak and trough, and the blue double arrows indicate the distances to the physical disparity line. The energy model responses to the individual component gratings are shown on the left and top of each panel.

(A) Energy model with a pure position disparity.

(B) Energy model with a pure phase disparity.

phase-shifted broadband pattern has a different shape in the left from the right eye, something that never occurs in the real world. Although real surfaces (e.g., tilted planes or curved surfaces) can give rise to disparities that are not simple translations, no solid surface produces equal interocular phase differences in all its Fourier components.

In order to explore the relationship of those two regimes—phase and position disparities—in both the stimulus and the RFs, we have used the simplest stimulus that allows both to be defined, a compound grating that consists of two frequency components, f_1 and f_2. Figure 2 shows the energy model response for the two example neurons in Figure 1B—one with RFs related by a pure position disparity (Figure 2A) and one with RFs related by a pure phase disparity (Figure 2B). The neuronal response is given in a 2D space spanned by the interocular phase differences for each component independently: the interocular phase difference for grating 1 is plotted on the x axis and that for grating 2 on the y axis. Each physical displacement of the image in the left with regard to the right eye implies a pair of interocular phase differences $(\Delta\varphi_1, \Delta\varphi_2)$ where $\Delta\varphi_1/\Delta\varphi_2 = f_1/f_2$. We chose a value $f_2/f_1 = 2$ for our empirical study. Disparities that result from the physical depth of planar patches in natural viewing therefore form a line of slope $f_2/f_1 = 2$ in this space. We refer to this line as the "physical disparity line." Stimuli away from this line do not occur in natural viewing (are nonphysical).

In the Experimental Procedures (Energy Model Predictions), we show that the response of the energy model to a drifting compound grating is linear in its components: the response to a compound grating is simply the sum of the responses to the component gratings individually. Since the response of an energy model to a single grating is a sinusoidal function of disparity (Fleet et al., 1996), the peak and the trough of the combined response will be separated by 180 degrees along each dimension (see Figures 2A and 2B). Hence, the line connecting peak and trough (green in Figures 2A and 2B) has a slope of 1. This means that peak and trough can never both lie on the physical disparity line at the same time—at least one of them has to occur for nonphysical stimuli. The two examples in Figure 2 illustrate this: for energy model neurons with no phase disparity, the maximum response is elicited by a naturally occurring disparity (Figure 2A). However, in such neurons, eliciting the minimum response requires a nonphysical stimulus, one where each Fourier component is shifted by 180 degrees, i.e., inverted (dashed lines in Figures 2A and 2B). This means that only part of the dynamic range is dedicated to stimuli that actually occur in the real world. The same is true for an energy model constructed with a pure phase disparity (Figure 2B). The maximum response is elicited when $\Delta\varphi_1 = \Delta\varphi_2$ and equals the RF phase disparity, a stimulus that is necessarily away from the physical disparity line. Note that the response in Figure 2B is simply a translated version of Figure 2A. This is true in general: any changes in either position or phase disparity correspond to a mere translation of the response surface but do not alter the relative location of maximum and minimum responses. The relative location is also unaffected by changes in the static output nonlinearity of the energy model. Thus, all variations on the energy model (including hybrid models that combine phase disparity and position disparity) exhibit a common feature: the separation between maximum response and minimum response is 180 degrees of phase on both axes. This shared feature of energy models implies that part of a neuron's dynamic range is devoted to parts of this space that do not correspond to realistic stimuli.

Neuronal Responses

We examined the responses of disparity selective neurons (recorded from the striate cortex of two awake fixating monkeys) to such stimuli, constructed by summing two sinusoidal luminance gratings, with spatial frequencies in the ratio 1:2. The top row in Figure 3 illustrates the responses of one example cell, which closely matches the predictions of the energy model.

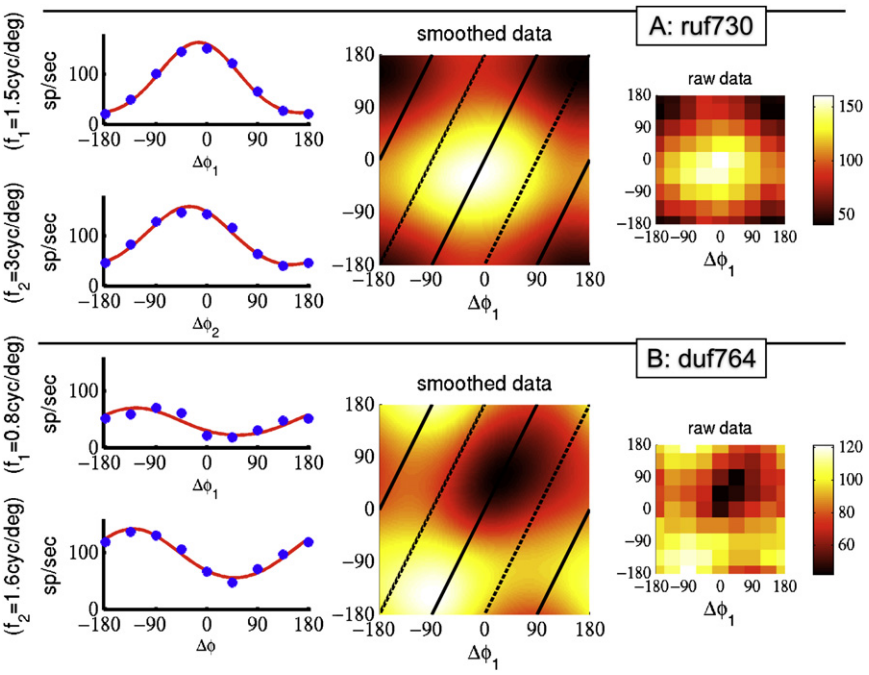

Figure 3. Two Example Cells

The graphs in the left-most column show the responses to the individual gratings of which our compound gratings are composed. The blue markers represent the data (SEMs are smaller than markersize), and the red lines depict the best-fitting sinusoids plus output exponent. The right two panels show the neuronal response to the compound gratings, smoothed and unsmoothed. (A) shows an example that behaves like the energy model (S = 81 degrees). The neuron in (B) concentrates its response near the physical disparity line (in a way the energy model cannot, S = 16 degrees). Solid and dashed lines represent physical and anticorrelated disparities, as in Figure 2.

The response to all combinations of interocular phase difference is well predicted by the responses to individual components, as expected from the energy model. Its response peaks for a physical stimulus, as implied by a model with no phase disparity and, as a consequence, the lower end of its dynamical range is dedicated to nonphysical stimuli. The bottom row in Figure 3 shows responses of a different cell that did not follow the model predictions. This cell exploits the properties of naturally occurring disparities—both peak and trough lie close to the physical disparity line. Thus, all its dynamic range is devoted to physical stimuli.

The degree to which a single neuron exploits the properties of physical disparities can be conveniently quantified by the shortest distance from the physical disparity line to the location of the peak and of the trough (blue arrows in Figure 2). The sum S of these distances is 0 if both peak and trough lie on the line. Since the responses of all energy models are identical up to translation in our space, S is the same for all of them. From the geometry of the space it follows that $S = 90\,\text{deg} \cdot \sin \arctan 2$ or about 80.5 degrees for the energy model (note the shortest distance is not aligned with either axis, and the value of S depends upon the slope of the physical disparity line). Because both axes represent circular variables, the maximum possible value for this sum S is ~161 degrees, the shortest distance between adjacent sections of the physical disparity line. We can therefore use S as a measure for how well adapted any one neuron is to physical images. For the example neuron in Figure 3 that resembles the energy model S = 81 degrees while for the other, adapted, neuron S = 16 degrees. Note that S is strictly positive and S = 0 is only achieved when both peak and trough lie exactly on the physical disparity line; any deviation due to sampling noise increases its value so that the observed value of S would be greater than 0 even for a perfectly adapted neuron.

We show the locations of peaks and troughs for our sample of 45 neurons in Figure 4. It is striking that the locations of the peaks are clustered closely around the physical disparity line, indicating some specialization for the properties of natural binocular images. This observation alone can readily be explained by the energy model, as long as the majority of neurons have only small phase disparities. However, if that were the case, then the response troughs would necessarily cluster *away* from the physical disparity line as explained above. What we find, however, is a distribution of troughs with two maxima, one near the physical disparity line and one far away (Figure 4C, blue). This implies that there is a subpopulation of neurons that has both peaks and troughs near the physical disparity line (like our example in Figure 3B). The deeper implication is a specialization for natural binocular images in those neurons—dedicating the majority of the dynamic range to physical disparities—one that the energy model cannot explain. Despite this phenomenon, the clustering of peak responses around physical disparities is much stronger than the clustering of troughs. This suggests that the coding scheme used by the brain may attach a special significance to response maxima. The distribution of troughs shows a second peak around ±80.5 degrees, compatible with an energy-model-like population of cells with peaks near the physical disparity line, and troughs therefore far away from it.

The adaptation to physical disparities that we observe automatically implies a deviation from the energy model, but the converse is not true: deviations from the energy model do not in general produce this adaptation—deviations can also increase the distance to the physical disparity line so that less, not more, of the dynamic range is devoted to naturally occurring stimuli. In order to quantify deviation from the energy model independent of adaptation, we use another measure of distance in the interocular phase space: the Euclidian distance of the observed peak from the energy model prediction, given the observed trough. In Figure 5A we compare those two measures—the distance to the energy model prediction and the total distance to the physical disparity line. First, we note that the upper half of the space potentially occupied by neurons (borders are dashed) is almost completely empty and all cells lie either close to the centerline (energy model) or below it. This demonstrates the power of the pressure for disparity-selective neurons to be adapted in this

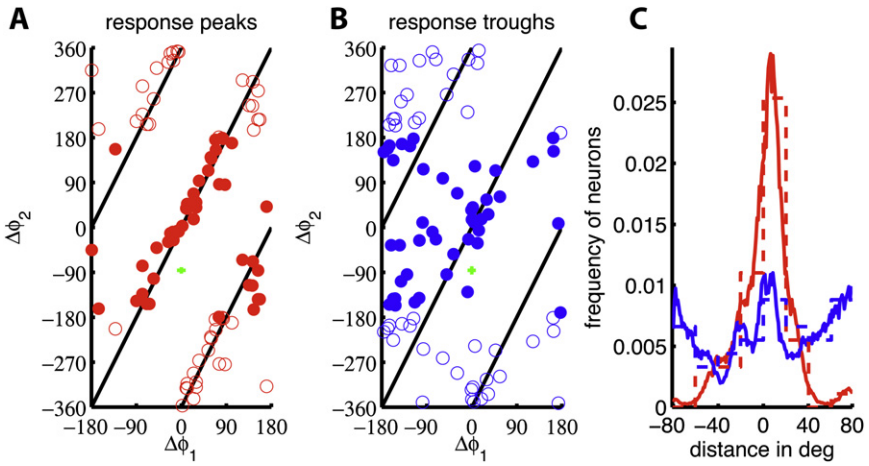

Figure 4. Population Plot of Peak and Trough Locations

(A and B) The location of peaks ([A], red) and troughs ([B], blue) in the 2D space spanned by the interocular phase difference in each component (cf. Figure 2) is shown. Note that we double the y axis to display the full, unbroken physical disparity line going through the origin. Hence, each data point is displayed twice (once filled, once open). The green crosses show the size of the median SEM.

(C) The distribution of distances of the peaks (red) and troughs (blue) to the physical disparity line is shown. The dashed lines represent a histogram of the raw data. The solid lines show the distribution of the 1000× resampled data.

particular way. We find that whenever neurons deviate from the energy model, they do so to decrease *S* and systematically devote more of their dynamic range to signaling naturally occurring disparities. The correlation between the two measures is highly significant (Spearman's $r = -0.77; p = 4 \cdot 10^{-9}$). In our sample of 45 neurons, 17 (38%) show a significant adaptation ($p < 0.05$, by resampling, magenta dots in Figure 5). Of the 20 neurons that deviate significantly from the energy model, all but one do so in the direction of increased adaptation. The marginal distribution of neurons along the y axis (Figure 5B) confirms this finding on the population level. The blue solid curve shows the distribution of the 1000× resampled data (raw data in cyan). By assuming that values $S > 80.5$ degrees are due to sampling noise around energy-model-like neurons (true $S = 80.5$ degrees), we can estimate the overall distribution of energy-model-like cells (dashed blue curve, see Experimental Procedures). The area under the dashed curve yields the total percentage of energy-model-like neurons in our population: 42%. The remaining 58% are at least partially adapted ($S < 80.5$ degrees).

A New Model

The new specialization of disparity-selective neurons we find can be achieved by a relatively simple extension of the energy model (illustrated in Figure 6). Conceptually, an adapted cell can be constructed by combining two cells (C+ and C−), each of which is a traditional energy model plus an expansive static output nonlinearity. Here, input from one of the cells (C−) is subtracted from the response of the other (C+). The result is that the location of the response maximum is largely determined by the position of the peak for C+, and the response minimum is largely determined by the location of the peak for C−. Figure 6A demonstrates this for compound gratings: if the peaks for C+ and C− lie close to the physical disparity line (produced by position disparity), then the model cell reproduces the adaptation we observe. Figure 6B shows the response of the same model to random dot stereograms (RDS, red lines). Both C+ and C− are tuned-excitatory (TE) and incorporate an expansive output nonlinearity (an exponent of 3 in this example). The third panel of Figure 6B shows the total response of the new model: its

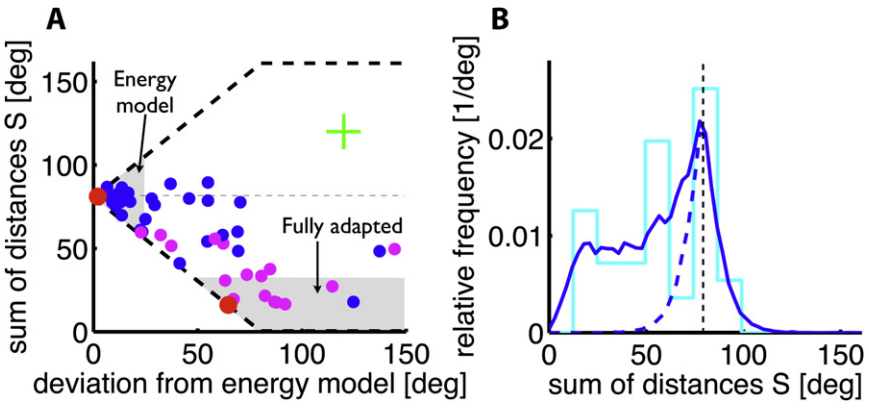

Figure 5. Population Summary of Adaptation

(A) The total distance, *S*, from the physical disparity line (quantifying adaptation) as a function of deviation from the energy model. Note that neurons that deviate from the energy model lie below the center line, showing adaptation to natural images (correlation: Spearman's $r = -0.77; p = 4 \cdot 10^{-9}$). Magenta circles denote cells for which the adaptation is significant (*S* significantly smaller than 80.5 degrees). The green cross shows the size of the median SEM. The red circles indicate our example cells from Figure 3. The energy model prediction for *S* is 80.5 degrees; for fully adapted cells *S* approaches zero. (*S* is inevitably greater than zero since the measure is the sum of two positive numbers.) The two gray areas show how far from the theoretical prediction 68% of cells would fall in the case that all scatter was the result only of sampling noise. The dashed black lines result from geometric constraints—it is impossible for a cell to lie outside them.

(B) Contains histograms showing the distribution of our neurons depending on their adaptation (cyan solid line indicating the raw data). The solid blue line shows our estimate of the actual distribution (resampling 1000 times). The dashed blue line was obtained by mirroring the distribution of neurons (solid blue) right of the energy model prediction (dashed gray line). It represents an unbiased estimate of the subpopulation of energy-model-like neurons in our sample.

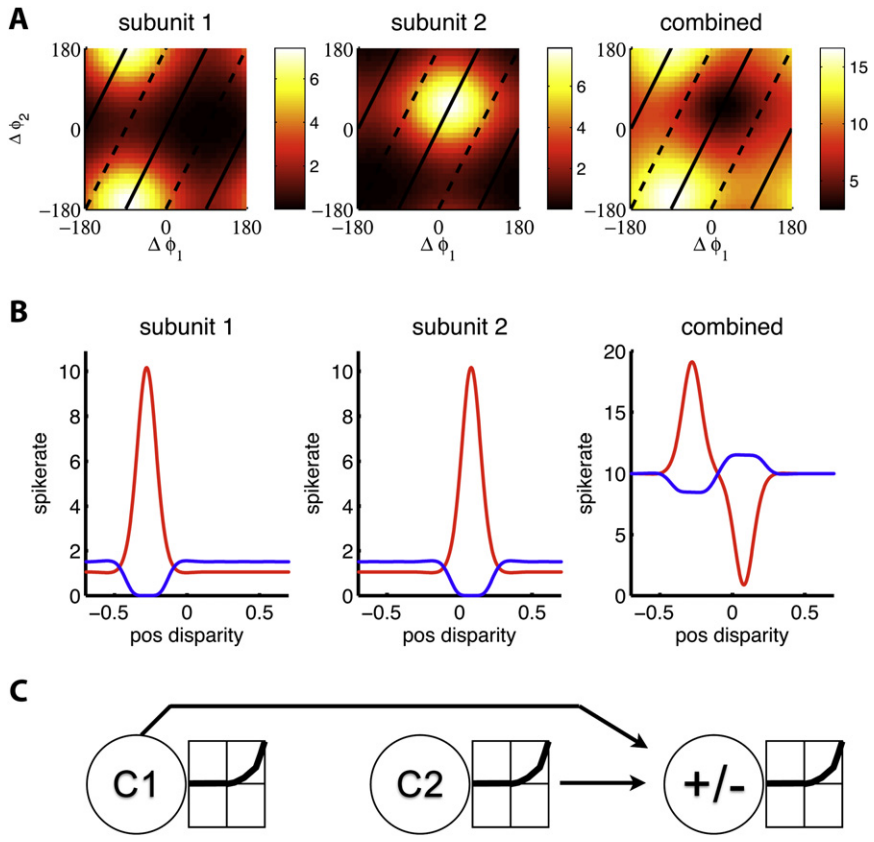

Figure 6. Proposed Subunit Model

The responses to compound gratings (A) and RDS (B) of an example subunit model (C) are shown. The left two panels of (A) and (B) show the responses of the two individual subunits C+ and C− that contribute to the model response (right panel in [A] and [B]). In (B), the red lines are responses to RDS and the blue lines to anticorrelated RDS. Both subunits are energy model neurons (plus output exponent of 3) that only differ in their preferred position disparity. Both are tuned excitatory. Our model subtracts the response of the second subunit from the first subunit, yielding the odd-symmetric tuning curve on the right (the baseline response to large disparities can easily be provided by monocular input or a difference in the baseline responses of the subunits). C+ and C− can be either disparity-selective simple or complex energy model cells, with identical responses in (A) and (B).

preferred and antipreferred disparities are inherited from its subunits and its overall tuning curve is odd-symmetrical, very much as expected from an energy model with phase disparity. We refer to these disparity selective input neurons (C+ and C−) as "subunits" below. In order to avoid confusion, we shall refer to the inputs to the energy model itself as binocular simple cells.

Note that although the parameters for our example model shown in Figure 6 were not fitted to real data, they were chosen to yield realistic responses to gratings and RDS. The particular values for the disparities and spatial frequency tuning were chosen such as to resemble the data of the example cell in Figure 3B. The preferred disparity of subunit 1 is given by the location of the peak in the 2D response surface in Figure 3B and the preferred disparity of subunit 2 by the location of the trough. The RFs of both subunits overlap extensively in each eye, but the exact location does not affect disparity responses. We also emphasize that the spatial frequency tuning curve is identical for both subunits: the adaptation is purely due to combining subunits followed by an additional output nonlinearity, not because of different frequency responses of the subunits.

It is not essential that one subunit provide a subtractive input. A similar result is obtained if a binocular cell of the tuned-inhibitory (TI) type (one whose disparity tuning function is dominated by a trough) (Poggio and Fischer, 1977) is summed with a cell C+ (TE), as shown in Figure S1, available online. In this case the TI cell must incorporate a compressive output nonlinearity. What both implementations have in common is that one subunit contributes a pronounced peak in the response, and a separate subunit contributes the pronounced trough. We find that when linearly combining two energy-model cells, it is important that we include a static output nonlinearity in addition to the squaring inherent to the energy model for the two cells—linearly combining two energy-model units without an additional output nonlinearity cannot produce an adaptation to encoding physical disparities.

The simplest form of our subtractive model assumes that all the RFs in both subunits are identical up to a translation and assumes that all subunits have exactly balanced inputs from both eyes. Deviations from these simplifying assumptions have little impact on the binocular responses we predict but can profoundly affect other properties of model neurons, such as the RF properties expected during monocular stimulation. In our simplest subtractive model, cancellation of the inputs from two subunits can even lead to no monocular responses in either eye (resembling the striking property of many disparity selective neurons in which the binocular response is much greater than either monocular response). However, if the model simple cells do not have exactly balanced inputs from the two eyes (commonplace in real neurons [Hubel and Wiesel, 1962; Prince et al., 2002; LeVay and Voigt, 1988; Smith et al., 1997; Ohzawa and Freeman, 1986b]) it is to easy produce monocular receptive fields that resemble real complex cells. If the simple cells contributing to the subtractive subunit have stronger inputs from one eye, while the inputs to the excitatory subunit are dominated by the other eye, the result will be a neuron that is disparity selective yet appears monocular (no response to stimulation in one of the eyes). Such responses have been noted in many studies (Ohzawa and Freeman, 1986b; Read and Cumming, 2004a; Smith et al., 1997) and are naturally explained in our new model. Ocular imbalance is not necessary to produce monocular responses in the additive version of our model (TE plus TI).

So far, the only assumptions we have made about the nature of C+ and C− concern their disparity selectivity. Even if C+

and C− were binocular simple cells, out model still produces an adaptation. Such a neuron would be classified as simple by the criteria used in physiological experiments (Kagan et al., 2002; Skottun et al., 1991). As a result, at least two different implementations can yield adapted complex cells. In the first, hierarchical one (illustrated in Figure 7A), the adapted neuron receives its input from two nonadapted complex cells, which each receive inputs from binocular simple cells. A second implementation does not require an additional hierarchy of complex cells as shown in Figure 7A. If all simple cell subunits, used to construct the two complex cells, are combined in one step to form a complex cell, an adaptation is produced, provided each simple cell subunit is augmented by an output nonlinearity. The additional output nonlinearity of the simple cell enhances the peak (or trough) of the disparity response. Note that only two traditional binocular simple cells of different preferred disparity (S1a and S2a) are strictly necessary to produce an adapted simple cell in either implementation. Four simple cells (S1a and -b and S2a and -b) are necessary to create a complex cell.

In both formulations, two key features of the original energy model are preserved: processing is linear up to the point of binocular combination, and a static output nonlinearity generates sensitivity to disparity. However, the original energy model also assumes (primarily for mathematical tractability) that all simple cell subunits share the same phase disparity and position disparity (Ohzawa et al., 1990). The empirical deviations shown in Figure 5 force us to abandon this assumption.

The extensions we propose have a profound impact on the long-standing debate concerning the relative roles of position disparity and phase disparity in generating the responses of complex cells (Anzai et al., 1999a, 1999c; Cumming and DeAngelis, 2001; Fleet et al., 1996; Nieder and Wagner, 2000; Ohzawa et al., 1997; Prince et al., 2002; Qian, 1997; Wagner and Frost, 1993). Our results demonstrate for the first time that the responses of many neurons cannot be summarized with a single pair of values for phase and position disparities. The second example in Figure 3 illustrates this point: an analysis of responses to physical disparities (a cross-section along the physical disparity line) reveals a pattern of odd symmetry that has been used to infer the existence of a phase disparity (Cumming and DeAngelis, 2001; Ohzawa et al., 1990). Our examination of responses to other combinations of interocular phase differences reveals that this inference is incorrect when the responses are better described as the result of more than one position disparity or more than one phase disparity.

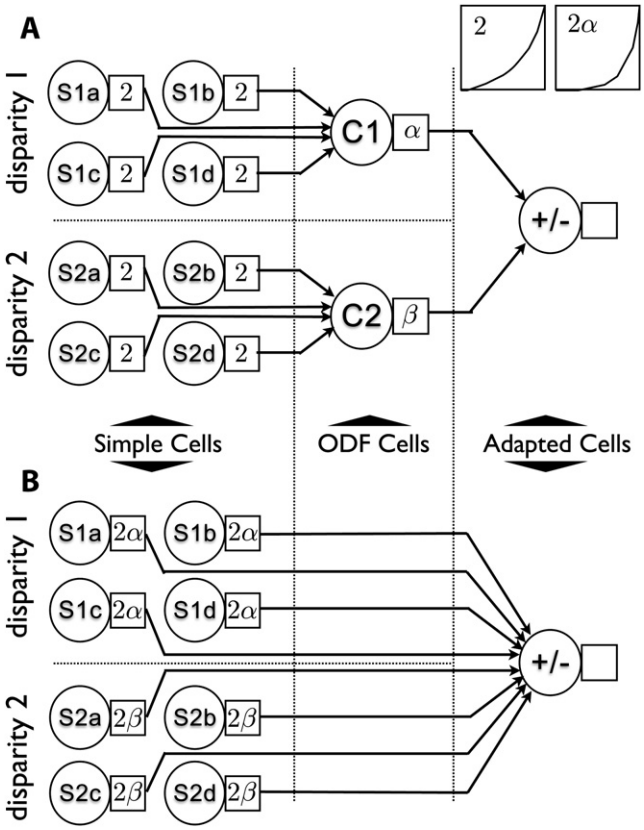

Figure 7. Two Implementations of Subunit Model

(A) The adapted neuron receives input from two energy model complex cells (ODF) followed by output exponents alpha and beta. Each of two complex cells C1 and C2, with different preferred disparities, receives input from four half-squaring simple cells S1a to S1d whose receptive fields are shifted by $\pi/2$ with respect to its neighbor.

(B) The adapted neuron receives input directly from the same half-squaring simple cells each followed by an output exponent α or β, respectively. Note that in both implementations, (A) and (B), the relevant feature of each individual disparity response (peak or trough depending on whether TE or TI cell) is enhanced by the same overall output exponents 2α (subunit 1) and 2β (subunit 2). It is this common feature that produces adaptation in both models. Note further that two binocular simple cells S1a and S2a are completely sufficient to achieve adaptation in a simple cell with disparity responses identical to those of the complex model in Figure 6. S1b and S2b are only needed to produce full-wave rectification (adapted complex cell) and S1c, S1d, S2c, and S2d in order to complete quadrature pairs, ensuring phase invariant responses to gratings.

This Model Explains Earlier Results

Finally, these data and the extended energy model we propose throw new light on one of the few existing observations concerning nonphysical disparities. Using stimuli in which the left eye's image is the photographic negative of the right eye's image (anticorrelated stereograms), several groups have reported that neurons still show disparity selectivity but that the strength of this selectivity is weaker than for correlated stereograms (Cumming and Parker, 1997; Nieder and Wagner, 2001; Ohzawa et al., 1997). Two modifications of the energy model have been proposed to explain this attenuation. One model (Lippert and Wagner, 2001) pointed out that passing the output of an energy model through an additional static nonlinearity can explain the result in some cells (although this cannot work for cells with odd-symmetric tuning [Read et al., 2002]). A second model proposed that a threshold is applied to monocular responses before binocular combination (Read et al., 2002).

Our data provide a powerful test of these models because responses to anticorrelated stereograms are part of the same continuous space as responses to correlated (physical) stimuli. They are a cross-section through our data running parallel to the physical disparity line (dashed lines in Figure 2). Models with thresholded monocular responses generate diagonally elongated regions in plots like Figures 2 and 3, unlike any we observed in real neurons. Simply adding an output exponent to the energy model does not change the locations of the peaks

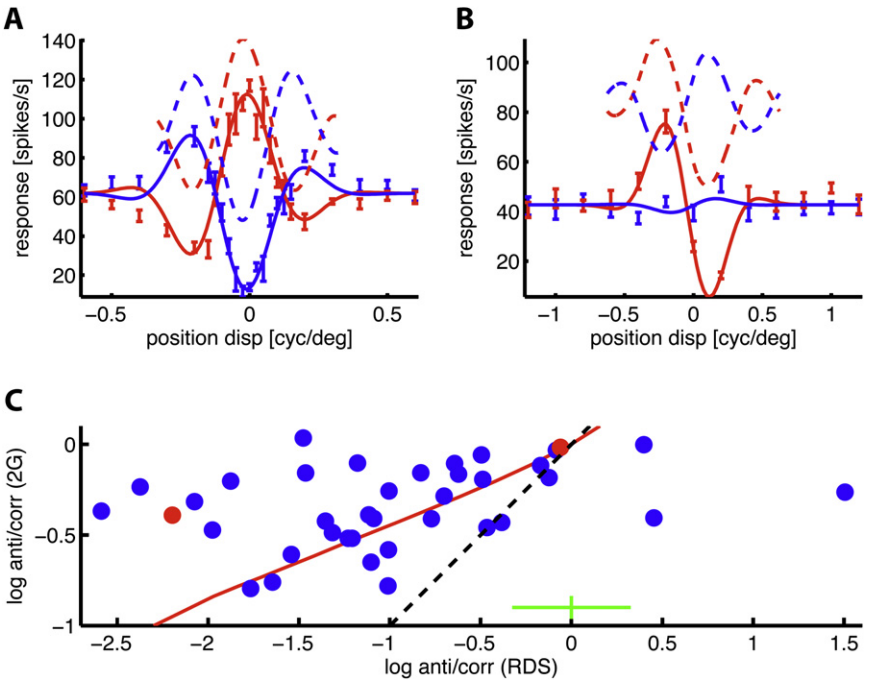

Figure 8. Comparison between Responses to Correlated and Anticorrelated RDS and Compound Gratings

(A and B) Responses to anticorrelated stimuli. For the example neurons from Figure 3, the response is shown to RDS (red, solid), anticorrelated RDS (blue, solid), and correlated and anticorrelated compound gratings (red and blue, dashed). For the RDS responses, error bars (SEM) and the best Gabor fit are shown. The compound-grating responses are cross-sections of the smoothed 2D response surface shown in Figure 3, along the physical disparity line (correlated) and its 180 degree phase-shifted equivalent (anticorrelated), respectively. Note that the dashed and solid curves are not expected to coincide even for the energy model since the stimuli are different.

(C) Population summary of attenuation caused by anticorrelation. Response attenuation for anticorrelated RDS is compared with that of compound gratings (2G). The dashed line is the identity line. Energy model-like neurons with no additional output exponent lie around (0;0), corresponding to no attenuation for either stimulus. The positions of the example neurons shown in Figures 3 and 7 are indicated by red dots. The red line traces an example subunit model response, obtained by varying only the output exponents of each subunit from 3.9 (lower left) to 0.9 (upper right) in the model shown in Figure 6. The green cross shows the median SEM. The reason for the relatively large error in the x direction is that cells with weak modulation to anticorrelated RDS produce very wide confidence intervals on this log scale.

and troughs and so cannot explain the results shown in Figures 3, 4, and 5. Thus, our data demonstrate that current explanations of the responses to anticorrelated stereograms are inadequate. Our model (Figure 6), by virtue of placing both response peak and trough on the physical disparity line, inevitably produces an attenuated response to anticorrelated compared to correlated patterns (blue versus red lines in Figure 6B). Thus, this same model also captures the main features of the same neuron's response to anticorrelated RDS (Figure 8B).

Therefore, a further test of our model is provided by comparing the responses to anticorrelated compound gratings with the responses to anticorrelated RDS. In Figures 8A and 8B, we compare the responses to correlated and anticorrelated compound gratings by taking cross-sections of the response along the correlated (physical disparities) and the anticorrelated line (solid and dashed, respectively, in Figure 2). The first example cell shows a very good agreement between the compound gratings and the RDS response in accordance with the energy model. The second example neuron shows an attenuation for both stimulus types; however, the attenuation for the anticorrelated RDS response is significantly stronger than that for the anticorrelated compound grating response. Figure 8 shows that this second pattern was typically observed across the population of our cells: the attenuation is generally larger for RDS than it is for compound gratings, and the two are correlated (Spearman's $r = 0.38$; $p < 0.02$). This stronger attenuation for RDS results because RDS contain many more frequency components than the two in our compound gratings, allowing the nonlinear interaction between them to further accentuate the effect. Our model replicates this behavior. The red line in Figure 8 represents one example family of our models, obtained by starting with the example model in Figure 6 and varying the output exponent for each complex cell subunit (from 0.9 in the upper right to 3.9 in the lower left). This line illustrates how our model replicates two aspects of the data: (1) the attenuation for RDS is stronger than for compound gratings, and (2) when keeping everything else constant and varying only the output nonlinearity, both attenuations are correlated. Please note, however, that the line we show represents only one of many families: the location and slope depend upon other model parameters (including the shape of the nonlinearity, the spatial frequency tuning of the RFs, and the baseline firing rate). Consequently, even our model does not predict a strong correlation between the attenuation for RDS and for compound gratings for a randomly sampled population of neurons. Nonetheless, a number of neurons lie above the line produced by our model. Whether they can be explained by invoking more than two subunits or by including other known properties like contrast-gain control requires further investigation.

DISCUSSION

Over small regions of the visual field, the difference between the two eyes' images is well described by a simple translation. This means that, when decomposing the two images into their

different frequency components, the interocular difference in the spatial phase of each frequency component is proportional to its frequency. Existing models of disparity selectivity, in particular the energy model, do not take this restriction into account, devoting part of their dynamic range to stimuli that never occur in the real world. We find that some neurons share this property of the energy model and are best tuned to nonphysical stimuli. This striking failure to adapt responses to the properties of physical binocular images is inevitable in the energy model and provides a powerful confirmation of it. However, we also find that more than half of our V1 neurons deviate from the energy model and do so in a way that exploits the properties of physical binocular images. We show that they concentrate their dynamic range in the regions occupied by realistic stimuli, demonstrating an adaptation of cortical neurons.

Furthermore, neurons that show no such adaptation (or incomplete adaptation) tend to have their maximal responses elicited by physical stimuli (Figure 4). This necessarily results in response minima elicited by nonphysical stimuli. This difference between the locations of minima and maxima suggests that the brain attaches special significance to high firing rates.

We propose two alternative models to explain the adaptation we observe: (1) combining the outputs of two (or more) energy model-like complex cells, with different phase or position disparities, in a new stage of binocular processing, or (2) allowing complex cells to receive input from binocular simple cells with different phase disparities, position disparities, and output nonlinearities. Model (1) has the advantage that it can explain why almost half of our neurons are not adapted to natural stimuli—they are necessary as building blocks. Model (2) does not need an additional processing stage and might provide a more natural explanation for the fact that our distribution of neurons displays a continuum of degrees of adaptation. Although model (2) does not explain why so many cells fail to show adaptation, it may be that responses to unnatural combinations of interocular phase may be useful for some other functions. One possibility is that these cells help solve the correspondence problem (determining how to match image features between left and right images) precisely because they identify signals that cannot correspond to a real world object (Read and Cumming, 2007). Both extensions of the energy model also reconcile the model with the earlier observation that responses to disparity in anticorrelated stereograms are weaker than responses to correlated stereograms (Cumming and Parker, 1997; Ohzawa et al., 1997; Nieder and Wagner, 2001).

The adaptation we describe cannot be explained by other, already known effects like dynamic changes in spatial frequency tuning (Bredfeldt and Ringach, 2002). As we show, the energy model response to a drifting compound grating is the sum of its responses to the component gratings individually, each of which is sinusoidal. As the spatial frequency tuning changes over time, only the amplitude of those sinusoidal responses is affected, not their shape. This means that the relative location of peak and trough in the 2D response to compound gratings is also unaffected—separated from each other by 180 degrees along each dimension. Indeed, this is true for any frequency combination, any spatial frequency tuning, and any output nonlinearity.

Another feature of real neurons that our model explains naturally concerns the binocular responses and ocularity of neurons. It has long been known that many neurons only respond to stimuli in one eye when probed monocularly but are actually disparity-selective when tested binocularly (LeVay and Voigt, 1988; Ohzawa and Freeman, 1986a, 1986b; Read and Cumming, 2004a; Smith et al., 1997). This is a direct consequence of the subtractive implementation of our model. Consider the following case: two TE cells with identical RFs at identical locations in both eyes are subtracted from each other. The result is that both subunits will cancel each other perfectly (C+ minus C−) in both eyes. If we then allow C+ to have stronger input from the left eye, while C− has stronger input from the right eye, the neuron will respond to monocular stimulation in the left eye, but not in the right eye.

A quantitative and more detailed examination of the relationship between the monocular and binocular responses can be achieved by an analysis of spike-triggered covariance (de Ruyter van Steveninck and Bialek, 1988; Rust et al., 2005) which reveals functional subunits (eigenvectors). This is possible even for model neurons that would be classified as simple cells. Such an analysis can also distinguish between the additive and subtractive version of our model. We have confirmed in simulations that the additive version produces only excitatory subunits, while the subtractive version also produces inhibitory subunits. This is in agreement with recent physiological data (Rust et al., 2005) that found excitatory and inhibitory eigenvectors for most simple and most complex cells in monkey V1. Since those experiments were conducted with monocular stimuli, it is not possible to say whether the inhibitory eigenvectors had the binocular properties predicted by our model. A spike-triggered analysis of binocular interactions in complex cells of the cat (Anzai et al., 1999b, 1999c) did not include an analysis of the monocular responses, so it is not clear that those data are at odds with our model.

Hence, allowing for multiple subunits that differ from each other in their disparity tuning changes our understanding of the relative roles played by position and phase disparity in complex cells. In the context of the original energy model, responses to a range of disparities can be used to deduce phase and position disparity (even in hybrid models that contain a mixture of both) (Anzai et al., 1999c; Cumming and DeAngelis, 2001; Fleet et al., 1996; Nieder and Wagner, 2000; Ohzawa et al., 1997; Prince et al., 2002; Qian, 1997; Wagner and Frost, 1993). Our new data show that for many disparity-selective cells it is not possible to summarize their responses with a single pair of phase and position disparities and that attempts to do so can yield misleading results. For example, odd-symmetric tuning curves ("near/far" types) have generally been thought to indicate the presence of phase disparity (Fleet et al., 1996; Freeman and Ohzawa, 1990; Ohzawa et al., 1990). However, we show that odd symmetry can be achieved by combining subunits with different position disparities and data like that shown in Figures 3, 4, and 5 suggest that this combination is important, even in the striate cortex. If our model uses position disparities that differ only in their horizontal components, it resembles one that was proposed to explain the observation that V1 neurons encode a wider range of horizontal than vertical disparities (Cumming, 2002; Read and Cumming, 2004b)—another specialization for the statistics of the binocular input.

We further note the similarities between our subunit model for disparity-selective V1 neurons and the original motion energy

model for direction selectivity (Adelson and Bergen, 1985). In both cases, the first stage is provided by the output of V1 simple cells selective to a particular stimulus feature (there a difference of position in time, here a difference of position between the eyes). And in both cases, the next step pools over a population of those simple cells. However, in the last step of the motion energy model, the responses of two pools of neurons signaling opposite directions are subtracted from each other ("opponency"). This opponent stage has not been used in previous versions of the binocular energy model. Our model incorporating subtraction is closely similar to the motion energy model with opponency, although differs somewhat because our two subunits are not in general symmetrical about zero disparity. Interestingly, this agrees with psychophysical measures of disparity opponency, in which opponent channels were not symmetrical about the fixation plane (Cormack et al., 1993). Although this similarity is interesting, it is important to note that the psychophysical result could still occur even if V1 contained only energy model-like neurons: one simply has to propose that opponency occurs later in the pathway. Similarly, the observation that stereoscopic adaptation produces a depth aftereffect (Blakemore and Julesz, 1971) is a necessary feature of our model but is compatible with phase disparity encoding in V1.

The model we propose may also be relevant to understanding some of the transformations applied to disparity signals in extrastriate cortex. It has long been recognized that odd-symmetric tuning curves are more frequently encountered in extrastriate cortex than in striate cortex (Cumming and DeAngelis, 2001; Poggio, 1995; Poggio et al., 1988). Our model provides an explanation for how this could be achieved in the projection from striate cortex to extrastriate cortex, and our data demonstrate a way in which this suggestion can be tested.

The idea that complex and simple cells in striate cortex receive input from multiple subunits with different spatiotemporal properties recently received direct support from a spike-triggered covariance analysis (Rust et al., 2005). Our results suggest that one reason that multiple subunits are combined is to ensure that the dynamic range of neuronal signals is matched to the range of inputs encountered naturally. This requires a combination of subunits appropriate to match the properties of the binocular inputs. The same principle might explain the recent finding that neurons give stronger responses to natural than to artificial monocular images (Felsen et al., 2005). This could result if multiple subunits were similarly matched to the statistics of monocular natural images. Thus, the unexpected specialization we show here might provide a general mechanism by which neural signals can be matched to the properties of natural occurring inputs.

The efficient coding hypothesis is one of the few unifying, crossmodal theories for the sensory system. Its main prediction, that neurons in the sensory system should be adapted to natural inputs, has been tested for a number of specific features inherent to auditory and monocular visual signals (Smith and Lewicki, 2006; Field, 1987; Schwartz and Simoncelli, 2001; Simoncelli and Olshausen, 2001). The tests in the visual domain have been on comparatively subtle statistical aspects of the stimuli, like short-range spatial correlations and the shape of the power spectrum. In this study, we tested a specialization to a particularly strong redundancy caused by the binocular nature of visual inputs. The results (1) confirm the efficient coding hypothesis by way of an adaptation for some neurons in our sample and (2) for the other cells provide a strong confirmation of the energy model mechanism in generating disparity signals in the brain. We propose an extension of the energy model as a unified account of the behavior of both types of neurons.

EXPERIMENTAL PROCEDURES

Stimulation and Recording

We recorded extracellularly from isolated V1 neurons in two awake, fixating rhesus macaque monkeys. Details of the experimental protocol are described elsewhere (Read and Cumming, 2003). All protocols were approved by the Institute Animal Care and Use Committee and complied with Public Health Service policy on the humane care and use of laboratory animals.

In brief, monkeys perched in a primate chair with their heads immobilized. A digitally controlled microdrive was used to lower electrodes (glass coated PtIr, FHC Inc.) through the dura into V1. Spikes were isolated based on the height and width of the electrical waveforms. All electrical activity crossing a manually set voltage threshold was saved for offline analysis with a precision of 0.1 ms. The positions of both eyes were recorded with scleral search coils (implanted under general anesthesia). Stimuli were presented on two Eizo Flexscan F980 monitors (96 Hz framerate, mean luminance 41.1 cd/m^2, contrast 99%, viewing distance 89 cm, 1 pixel subtended 1.1 min arc at the eye) viewed through a Wheatstone stereoscope. Gamma correction was adjusted to achieve a linear luminance response.

Trials lasted approximately 2 s, during which the monkey was required to maintain fixation within a 0.8° window. If the monkey successfully maintained fixation throughout the trial, he received a water reward. Each trial consisted of four 420 ms stimulus presentations, separated by 100 ms blank periods. If the monkey broke fixation before the end of the trial, the successfully completed stimulus presentations were saved, but the monkey did not receive a reward. Responses were measured as the mean spike rate over the entire stimulus presentation, beginning 50 ms after stimulus onset. Minimum response fields (MRF) were mapped manually with bars and gratings, and all subsequent stimuli were centered over the MRF and were made substantially larger than the MRF to ensure that the receptive field was covered. Preferred orientation and spatial frequency were measured with patches of grating presented to the dominant eye, and these values were then used when measuring disparity selectivity to a single binocular grating. Disparity selectivity was also examined with dynamic random dot stimuli. Neurons exhibiting disparity selectivity to either gratings or RDS were then examined with compound gratings. (We also examined a few cells that were not disparity selective to single gratings or RDS. Such cells never exhibited selectivity for interocular phase difference in compound gratings, so were excluded.)

Compound gratings were generally constructed from summing two 50% contrast gratings (frequencies in the ratio 1:2) that spanned the preferred spatial frequency. In a few cases, it was necessary to alter the frequencies because the preferred spatial frequency defined with monocular gratings did not produce the strongest disparity selectivity. The stimuli were typically square patches 4° × 4° across, and the spatial frequencies used ranged from 0.1 to 6 cpd (depending on cell preference). The distribution of preferred orientations was approximately uniform. Eccentricities were in the range 2.5° to 8.5°. Eight values of interocular phase difference were applied to each component (=64 combinations), but the absolute phase (mean of the two eyes) for each component was set randomly at the start of each stimulus. Thus, the absolute phase and the relative phase between components in either eye were unaffected by changes in the interocular phase difference. The temporal frequency for each component was inversely proportional to its spatial frequency so that each eye saw a fixed pattern drifting through an aperture. Responses to the same interocular phase differences were also measured for the individual component gratings. All 80 stimulus conditions were interleaved in a pseudorandom sequence. A minimum of five stimulus repetitions was required, and the mean value was 13.6 repetitions per stimulus.

Analysis

We examined responses in 64 neurons that showed significant selectivity to either RDS or single binocular gratings ($p < 0.05$ on one-way ANOVA). Nineteen of these were excluded from further study because they did not show significant disparity selectivity to one of the two component gratings when tested alone ($p < 0.05$ by one-way ANOVA). All 45 of the remaining neurons showed significant selectivity for interocular phase in the compound grating.

In order to determine the location of maxima and minima, the responses were smoothed with a two-dimensional Gaussian filter (σ = 36 degrees), and the extrema of the smoothed response were used. The locations of these extrema were then used to compute the two indices (described in the Results) quantifying deviations from the energy model.

Data on responses to RDS were available for 43 of the 45 neurons. Five of these did not show significant responses to RDS, so we were able to compare the effects of anticorrelation (in RDS and compound gratings) in 38 neurons. To quantify the response attenuation caused by anticorrelation in compound gratings, we simply compared peak-to-peak response magnitudes for the appropriate slices through smoothed response surfaces like those in Figure 3. To quantify this attenuation in anticorrelated RDS, we simultaneously fit two Gabor functions, allowing only the amplitude and phase to differ between the two stimuli (Cumming and Parker, 1997):

$$f_{\text{corr, anti}}(x) = b + a_{\text{corr, anti}} \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right)\cos(2\pi f_x(x-x_d) - \phi_{\text{corr. anti}})$$

where b is the response to uncorrelated stimuli, a_{corr} and a_{anti} are the amplitudes, x_0 is the location of the centroid (equal to preferred disparity in the case of tuned even-symmetric cells), f_x is the preferred spatial frequency, and ϕ_{corr} and ϕ_{anti} are the phase disparities for the correlated and anticorrelated responses. The ratio $a_{\text{anti}}/a_{\text{corr}}$ then defines the attenuation.

Energy-Model Predictions

In the classical energy model for disparity-selective complex cells (Ohzawa et al., 1990, 1997) the output of two pairs of binocular simple cells with different RFs in both eyes are combined. The RFs in each pair are the inverse of each other, and the RFs of the two pairs are phase-shifted by 90 degrees. The preferred position and phase disparity of the cells is then given by the position and phase difference between the RFs in left and right eye which are assumed to be identical in all four binocular simple cells. Following the notation of Read et al. (2002), we can write the output of a monocular receptive field as

$$r = \iint \mathrm{d}x\,\mathrm{d}y\, I(x,y)\,\rho(x,y)$$

where $I(x,y)$ is the luminance of the image in that eye and $\rho(x,y)$ is the receptive field. For a grating of frequency f and phase φ, this becomes

$$r(f,\varphi) = \tilde{\rho}^{O}(f)\sin\varphi + \tilde{\rho}^{E}(f)\cos\varphi \quad (1)$$

where $\tilde{\rho}^{E,O}$ is the even/odd part of the Fourier transform of the receptive field. For clarity in the following derivation and without restricting generality, we assume position and phase disparity to be zero, one binocular pair to have RFs which are purely even (E), and one pair whose RFs are purely odd (O). Using the labels L and R to distinguish between left and right eye, the complex cell response to a compound grating becomes

$$C = \left(r_L^E + r_R^E\right)^2 + \left(r_L^O + r_R^O\right)^2 \quad (2)$$

with

$$r_L^O = A\sin_{\varphi 1} + B\sin_{\varphi 2} \quad r_R^O = A\sin(\varphi_1 + \Delta\varphi_1) + B\sin(\varphi_2 + \Delta\varphi_2)$$
$$r_L^E = A\cos_{\varphi 1} + B\cos_{\varphi 2} \quad r_R^E = A\cos(\varphi_1 + \Delta\varphi_1) + B\cos(\varphi_2 + \Delta\varphi_2) \quad (3)$$

where indices 1 and 2 refer to the component gratings of frequencies f_1 and f_2 and $\Delta\varphi_{1,2}$ to the interocular phase difference for component grating 1 and 2, respectively. $A = \tilde{\rho}(f_1)$ and $B = \tilde{\rho}(f_2)$ are given by the Fourier transform of the two RFs evaluated at frequencies f_1 and f_2 (Equation 1). Substituting Equation 3 into Equation 2 yields

$$C = 2A^2(1+\cos\Delta\varphi_1) + 2B^2(1+\cos\Delta\varphi_2) + 2AB[\cos(\varphi_1-\varphi_2) + \cos(\varphi_1-\varphi_2+\Delta\varphi_1) + \cos(\varphi_1-\varphi_2-\Delta\varphi_2) + \cos(\varphi_1-\varphi_2+\Delta\varphi_1-\Delta\varphi_2)]$$

The first two terms in this sum are the responses of the complex cell to each of the two component gratings alone. The third term depends both on the interocular phase differences $\Delta\varphi_{1,2}$ and the difference between the two absolute phases $\varphi_1 - \varphi_2$.

For a compound grating, $\varphi_1 - \varphi_2$ varies as a function of physical displacement. Therefore, the energy model response to compound gratings varies as a function of position, even though the response is position-invariant when either grating is presented alone. However, since $\varphi_1 - \varphi_2$ takes on all possible values between 0 and 360 degrees for each pair $(\Delta\varphi_1, \Delta\varphi_2)$, the third term averages to zero for drifting compound gratings. Hence, the average response of the energy model is linear in drifting gratings of differing spatial frequencies:

$$\langle C(\Delta\varphi_1,\Delta\varphi_2)\rangle_t = 2A^2(1+\cos\Delta\varphi_1) + 2B^2(1+\cos\Delta\varphi_2) = C(\Delta\varphi_1) + C(\Delta\varphi_2)$$

where $C(\Delta\phi) = 2\tilde{\rho}(f)^2 \cdot (1+\cos\Delta\phi)$ is the energy model response to an individual grating with identical receptive fields in left and right eye. Because this result depends on averaging over absolute phases, it can be seen that the same holds for simple cells.

Relaxing the assumption of identical RFs in left and right eyes and introducing a position and a phase disparity Δx_{RF} and $\Delta\phi_{\text{RF}}$ between the RFs in the left and right eye has the effect of simply translating the response pattern (for an example see Figure 2):

$$\langle C(\Delta\varphi_1,\Delta\varphi_2)\rangle_t = 2A^2(1+\cos(\Delta\varphi_1 - \varphi_{\text{RF}}(f_1))) + 2B^2(1+\cos(\Delta\varphi_2 - \varphi_{\text{RF}}(f_2)))$$

where the preferred interocular phase difference is $\phi_{\text{RF}}(f) = \Delta x_{\text{RF}} \cdot f + \Delta\phi_{\text{RF}}$.

Computational Details for Figure 5

We obtain an estimate of the frequency distribution of our population by performing a 1000 times resampling with replacement on the original spike rates. The resulting distribution functions $f_i(S)$ obtained for every neuron i are summed to form the distribution function $F(S)$ for our entire population (solid blue curve in Figure 5). In order to estimate the proportion of adapted neurons, we assume that F(S > 80.5 deg) is entirely due to energy-model-like neurons plus sampling noise. This can only overestimate the true number of energy-model-like neurons. Since random noise gives rise to a symmetric distribution around 80.5 deg, the distribution of energy-model-like neurons will be symmetric with respect to 80.5 deg (dashed blue line in Figure 5). The difference between this and the total distribution represents the adapted subpopulation.

Supplemental Data

The Supplemental Data for this article can be found online at http://www.neuron.org/cgi/content/full/57/1/147/DC1/.

ACKNOWLEDGMENTS

This research was supported by the Intramural Research Program of the NIH, National Eye Institute. We thank Denise Parker for excellent animal care.

Received: April 25, 2007
Revised: September 24, 2007
Accepted: October 31, 2007
Published: January 9, 2008

REFERENCES

Adelson, E.H., and Bergen, J.R. (1985). Spatiotemporal energy models for the perception of motion. J. Opt. Soc. Am. A *2*, 284–299.

Anzai, A., Ohzawa, I., and Freeman, R.D. (1999a). Neural mechanisms for encoding binocular disparity: receptive field position versus phase. J. Neurophysiol. *82*, 874–890.

Anzai, A., Ohzawa, I., and Freeman, R.D. (1999b). Neural mechanisms for processing binocular information I. Simple cells. J. Neurophysiol. *82*, 891–908.

Anzai, A., Ohzawa, I., and Freeman, R.D. (1999c). Neural mechanisms for processing binocular information II. Complex cells. J. Neurophysiol. *82*, 909–924.

Attneave, F. (1954). Some informational aspects of visual perception. Psychol. Rev. *61*, 183–193.

Barlow, H.B. (1961). In Sensory Communication, W.A. Rosenblith, ed. (Cambridge, MA: MIT Press), pp. 217–234.

Barlow, H.B., Blakemore, C., and Pettigrew, J.D. (1967). The neural mechanism of binocular depth discrimination. J. Physiol. *193*, 327–342.

Blakemore, C., and Julesz, B. (1971). Stereoscopic depth aftereffect produced without monocular cues. Science *171*, 286–288.

Bredfeldt, C.E., and Ringach, D.L. (2002). Dynamics of spatial frequency tuning in macaque V1. J. Neurosci. *22*, 1976–1984.

Chichilnisky, E.J. (2001). A simple white noise analysis of neuronal light responses. Network *12*, 199–213.

Cormack, L.K., Stevenson, S.B., and Schor, C.M. (1993). Disparity-tuned channels of the human visual system. Vis. Neurosci. *10*, 585–596.

Cumming, B.G. (2002). An unexpected specialization for horizontal disparity in primate primary visual cortex. Nature *418*, 633–636.

Cumming, B.G., and Parker, A.J. (1997). Responses of primary visual cortical neurons to binocular disparity without depth perception. Nature *389*, 280–283.

Cumming, B.G., and DeAngelis, G.C. (2001). The physiology of stereopsis. Annu. Rev. Neurosci. *24*, 203–238.

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

DeAngelis, G.C., Ohzawa, I., and Freeman, R.D. (1991). Depth is encoded in the visual cortex by a specialized receptive field structure. Nature *352*, 156–159.

Felsen, G., Touryan, J., Han, F., and Dan, Y. (2005). Cortical sensitivity to visual features in natural scenes. PLoS Biol. *3*, e342. 10.1371/journal.pbio.0030342.

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

Fleet, D.J., Wagner, H., and Heeger, D.J. (1996). Neural encoding of binocular disparity: energy models, position shifts and phase shifts. Vision Res. *36*, 1839–1857.

Freeman, R.D., and Ohzawa, I. (1990). On the neurophysiological organization of binocular vision. Vision Res. *30*, 1661–1676.

Hubel, D.H., and Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. *160*, 106–154.

Kagan, I., Gur, M., and Snodderly, D.M. (2002). Spatial organization of receptive fields of V1 neurons of alert monkeys: comparison with responses to gratings. J. Neurophysiol. *88*, 2557–2574.

LeVay, S., and Voigt, T. (1988). Ocular dominance and disparity coding in cat visual cortex. Vis. Neurosci. *1*, 395–414.

Lippert, J., and Wagner, H. (2001). A threshold explains modulation of neural responses to opposite-contrast stereograms. Neuroreport *12*, 3205–3208.

Nieder, A., and Wagner, H. (2000). Horizontal-disparity tuning of neurons in the visual forebrain of the behaving barn owl. J. Neurophysiol. *83*, 2967–2979.

Nieder, A., and Wagner, H. (2001). Hierarchical processing of horizontal disparity information in the visual forebrain of behaving owls. J. Neurosci. *21*, 4514–4522.

Nikara, T., Bishop, P.O., and Pettigrew, J.D. (1968). Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex. Exp. Brain Res. *6*, 353–372.

Ohzawa, I., and Freeman, R.D. (1986a). The binocular organization of complex cells in the cat's visual cortex. J. Neurophysiol. *56*, 243–259.

Ohzawa, I., and Freeman, R.D. (1986b). The binocular organization of simple cells in the cat's visual cortex. J. Neurophysiol. *56*, 221–242.

Ohzawa, I., DeAngelis, G.C., and Freeman, R.D. (1990). Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. Science *249*, 1037–1041.

Ohzawa, I., DeAngelis, G.C., and Freeman, R.D. (1997). Encoding of binocular disparity by complex cells in the cat's visual cortex. J. Neurophysiol. *77*, 2879–2909.

Poggio, G.E. (1995). Mechanisms of stereopsis in monkey visual cortex. Cereb. Cortex *5*, 193–204.

Poggio, G.F., and Fischer, B. (1977). Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. J. Neurophysiol. *40*, 1392–1405.

Poggio, G.F., Gonzalez, F., and Krause, F. (1988). Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. J. Neurosci. *8*, 4531–4550.

Prince, S.J., Cumming, B.G., and Parker, A.J. (2002). Range and mechanism of encoding of horizontal disparity in macaque V1. J. Neurophysiol. *87*, 209–221.

Qian, N. (1997). Binocular disparity and the perception of depth. Neuron *18*, 359–368.

Read, J.C., and Cumming, B.G. (2003). Measuring V1 receptive fields despite eye movements in awake monkeys. J. Neurophysiol. *90*, 946–960.

Read, J.C., and Cumming, B.G. (2004a). Ocular dominance predicts neither strength nor class of disparity selectivity with random-dot stimuli in primate V1. J. Neurophysiol. *91*, 1271–1281.

Read, J.C., and Cumming, B.G. (2004b). Understanding the cortical specialization for horizontal disparity. Neural Comput. *16*, 1983–2020.

Read, J.C., and Cumming, B.G. (2007). Sensors for impossible stimuli may solve the stereo correspondence problem. Nat. Neurosci. *10*, 1322–1328.

Read, J.C., Parker, A.J., and Cumming, B.G. (2002). A simple model accounts for the response of disparity-tuned V1 neurons to anticorrelated images. Vis. Neurosci. *19*, 735–753.

Rust, N.C., Schwartz, O., Movshon, J.A., and Simoncelli, E.P. (2005). Spatiotemporal elements of macaque v1 receptive fields. Neuron *46*, 945–956.

Schwartz, O., and Simoncelli, E.P. (2001). Natural signal statistics and sensory gain control. Nat. Neurosci. *4*, 819–825.

Simoncelli, E.P., and Olshausen, B.A. (2001). Natural image statistics and neural representation. Annu. Rev. Neurosci. *24*, 1193–1216.

Skottun, B.C., De Valois, R.L., Grosof, D.H., Movshon, J.A., Albrecht, D.G., and Bonds, A.B. (1991). Classifying simple and complex cells on the basis of response modulation. Vision Res. *31*, 1079–1086.

Smith, E.C., and Lewicki, M.S. (2006). Efficient auditory coding. Nature *439*, 978–982.

Smith, E.L.R., Chino, Y.M., Ni, J., Ridder, W.H.R., and Crawford, M.L. (1997). Binocular spatial phase tuning characteristics of neurons in the macaque striate cortex. J. Neurophysiol. *78*, 351–365.

Wagner, H., and Frost, B. (1993). Disparity-sensitive cells in the owl have a characteristic disparity. Nature *364*, 796–798.