Vision Res. Vol. 32, No. 2, pp. 203–218, 1992
Printed in Great Britain.

0042-6989/92 $5.00 + 0.00

Directionally Selective Complex Cells and the Computation of Motion Energy in Cat Visual Cortex

ROBERT C. EMERSON,* JAMES R. BERGEN,† EDWARD H. ADELSON‡

Received 7 March 1990; in revised form 10 July 1991

We applied a set of 1- and 2-bar tests to directionally selective (DS) complex cells in the cat's striate cortex, and compared the responses with those predicted by two computational models. Single-bar responses and 2-bar interactions produce distinctive patterns that are highly diagnostic. The observed responses are quite similar to those predicted by a basic (non-opponent) motion-energy model [Adelson & Bergen (1985) Journal of the Optical Society of America A, 2, 284–299]. However, they are not consistent with an opponent combination of energy models, nor are they consistent with any stage of the classic Reichardt model. In particular, the Reichardt model (as well as opponent combinations of energy models) predicts a separable *space–time symmetry in the 2-bar interaction that is not observed in our measurements, while the non-opponent energy model predicts an* inseparable, *oriented interaction very similar to the measured cortical responses. Comparisons between model and measurements suggest possible mechanisms of spatial receptive-field organization and of nonlinear transformations.*

Motion Receptive field Model Directional selectivity Visual cortex Complex cells Nonlinear Motion energy Vertebrate Cat

INTRODUCTION

Interest in processing of signals related to image motion by single visual neurons has been intense since the late 1950s. In the cat's visual cortex, Hubel and Wiesel (1959) showed that many neurons responded more strongly to one direction of bar motion than to the opposite direction, i.e. they were directionally selective (DS). At about the same time, Reichardt and colleagues (e.g. Reichardt, 1957, 1961) began studying optomotor responses in the fly, and proposed a motion-opponent DS model that has been successful in many domains. More recently van Santen and Sperling (1985) have described an "elaborated" version of the Reichardt model that incorporates front-end bandpass filters in space and time to better fit known properties of the *vertebrate* visual system than the lowpass and delay filters originally used to model the fly.

It has been shown (Poggio & Reichardt, 1976) that many different motion computations are formally related; and in the literature these various models are sometimes grouped together to emphasize their similarities. But in our present work we are interested in making possible distinctions between models. Therefore, to avoid confusion, we will follow the common convention of using the term "Reichardt model" to refer only to the classic Reichardt computation, in which bandpass filters are optional, but a four-quadrant multiplier is obligatory.

A number of investigators have proposed that the initial stage of DS could be achieved with linear spatiotemporal filters (Fahle & Poggio, 1981; Watson & Ahumada, 1983, 1985; Adelson & Bergen, 1985; Burr, Ross & Morrone, 1986). Furthermore, there is recent evidence for corresponding single cells in the visual cortex (Reid, Soodak & Shapley, 1987; McLean & Palmer, 1988, 1989; Hamilton, Albrecht & Geisler, 1989), but see Discussion about their less obvious role in motion processing. Although such linear filters cannot generate a higher *mean* response for motion in the preferred direction, their gradual progression of temporal phase with position may produce stronger *modulation* for that direction. Adelson and Bergen (1985) have proposed a model in which such a linear stage is followed by a nonlinear interaction that produces an unmodulated (phase-independent) signal corresponding to "motion energy". The model is consistent with a variety of psychophysical motion phenomena, and can plausibly be implemented with neural hardware.

*Department of Ophthalmology/Box 314, and Center for Visual Science, University of Rochester, Rochester, NY 14642, U.S.A.
†SRI/David Sarnoff Research Center, Princeton, NJ 08543-5300, U.S.A.
‡MIT Media Lab and Department of Brain and Cognitive Science, E15-384, Cambridge, MA 02139, U.S.A.

Our interest here is to study the properties of DS *complex* cells (Hubel & Wiesel, 1962), and to relate their behavior to that of candidate models. It has been noted that the motion-energy model is closely related to the Reichardt model, and that in certain cases it is possible to construct opponent combinations of basic energy models that are formally equivalent to Reichardt models (Adelson & Bergen, 1985; van Santen & Sperling, 1985; cf. Poggio & Reichardt, 1976). However, we show here that the basic energy stage of the Adelson and Bergen model is not equivalent to any stage of the Reichardt model. This difference creates an opportunity to distinguish between the models.

Previously, we have applied two spatiotemporal tests to DS complex cells in the cat's striate cortex (Emerson, Citron, Vaughn & Klein, 1987b). We have now predicted responses of Reichardt and energy models to the same two tests. Each model contains many stages, and each stage of each model predicts distinct behavior. Although the final stage of an opponent combination of two basic energy models can be identical to the final stage of fundamentally opponent models such as the Reichardt model, at every other stage the two models differ in unambiguous ways. As it turns out, among these single-subunit models, the observed neural behavior agrees well with only one stage of one model, namely, the basic (non-opponent) energy stage of the motion-energy model. We therefore offer the basic motion-energy model as a plausible candidate for DS complex cells in cat. Some of these results have previously been presented briefly (Emerson, Bergen & Adelson, 1987a; Emerson & Bergen, 1989a,b).

METHODS

Surgical preparation

Details about the surgical preparation of adult cats have been reported elsewhere (Emerson & Gerstein, 1977a; Citron, Emerson & Ide, 1981; Emerson & Coleman, 1981; Emerson *et al.*, 1987b). Briefly, animals were surgically anesthetized with barbiturate, and prepared for extracellular recording of single neurons in area 17 with high impedance glass micropipettes. After installing a fluid-tight chamber onto the skull, the exposed dura was reflected, and cortical oscillations were restricted by filling the sealed chamber with high viscosity silicone fluid. During the experiment, the animal was paralyzed, and anesthetized lightly with a continuous 1–2 mg/kg/h of thiopental sodium I.V. Heart rate was monitored continuously to ensure that anesthetic dosage was sufficient to keep the animal in a relaxed and unresponsive state. Contact lenses focussed the eyes onto a tangent screen or video monitor 115 cm distant.

Visual stimuli

After locating and classifying receptive fields (RFs) with conventional moving edges and bars, and with flashing bars (Citron *et al.*, 1981; Emerson & Coleman, 1981; Emerson, 1988), we inserted a mirror into the path to the tangent screen and centered an 8° by 11° video monitor on the RF. We then measured the impulsive responses to single and paired, optimally oriented bars by presenting a counterbalanced sequence of dark and light bars, simultaneously at 16 contiguous 0.5° by 11° bar positions across the RF. This 16-bar random "white-noise" stimulus has been described in detail previously (Citron & Emerson, 1983). Each bar could change its luminance randomly and independently of the other bars every video frame (every 16 msec) over three luminances, 444 cd/m^2 (light bar), 222 cd/m^2 (mean luminance), and 0 cd/m^2 (dark bar). As all possible combinations of dark and light bars could occur in a given frame, the stimulus included bars of both contrast signs, each moving up and down in short sequences, and at a wide range of speeds. All cell measurements shown here except in Fig. 1A were isolated from components of the response to this rich spatiotemporal stimulus.

Physiological data

The times of neural impulses were measured with a resolution of 1 msec, and used for estimating the responses to two tests, a single-bar presentation (Fig. 1C–E) and the nonlinear interaction between responses in a two-bar presentation (Fig. 2). The Wiener-like kernels used for these estimates have been described in detail elsewhere (see Appendix of Emerson *et al.*, 1987b). They represent a "time-locked response", similar to a PST histogram, and are calculated, as in a PST histogram, by crosscorrelating the neural response with the single or paired stimuli of interest. Here, the nonlinear interaction plots were limited to (Wiener) second order, which is the lowest-order nonlinear interaction. In the context of a Wiener kernel, the order implies the number of stimulus events that contribute to the crosscorrelation, and for intensive nonlinearities the calculation can generate a squared term by including the product of two bars that coincide in space and time.

Although second-order may seem too low to characterize 1- and 2-bar responses in a real complex cell, which could be of order higher than two, (1) Wiener kernels are orthogonal, and therefore up to their order do the best job possible in capturing nonlinear behavior (Marmarelis & Marmarelis, 1978). That is, low-order Wiener kernels partially incorporate any existing nonlinearities of higher-degree. (2) In an earlier paper we compared a second-order measurement with one that included third- and fourth-order Wiener kernels (Fig. 1K vs J of Emerson *et al.*, 1987b). We found essentially no fourth-order (Fig. 5E of that paper), and the estimates look qualitatively the same, which suggests that any third-order nonlinearities may affect the amplitude but not the shape of the interaction. And (3), a 2-bar stimulus is all that is needed to test completely the second-order model under consideration. The nonlinear residual from a 2-bar linear superposition test is identical to the model's Wiener second-order kernel (p. 153 of Marmarelis & Marmarelis, 1978), and provides an

equivalent test with which to compare the second-order kernel of the measured cell.

Model data

Model data were derived from digital simulations of 1- and 2-bar tests of the models discussed by Adelson and Bergen (shown in Fig. 18 of their 1985 paper). Responses were simulated by convolving temporally and spatially the stimuli of interest with the model filter functions and then subjecting the resulting signal to the appropriate linear and nonlinear transformations. Impulse responses of the two linear *temporal* filters used in the early stages are shown here in Fig. 3B. They were drawn from equation (1) of Adelson and Bergen (1985), with the addition of β to adjust the weighting of the undershoot relative to the earlier positive phase

$$f(t) = (kt)^n \exp(-kt)[1/n! - \beta(kt)^2/(n+2)!] \quad (1)$$

where n takes values of either 6 or 9, k was 1.5, and β was 0.9. Responses were sampled in 16 periodic increments in time, t, beginning with the time of stimulus presentation, and also 16 samples in one-dimensional (1D) space. The number of samples plotted was chosen to match the physiological measurement.

In our first calculation of 1- and 2-bar model responses (Emerson *et al.*, 1987a) we used the odd- and even-symmetric *spatial* Gabor functions of Watson and Ahumada (their Fig. 7B and Fig. 6B in 1983 & 1985, respectively), and of Adelson and Bergen (Fig. 10A in 1985). Here, we have been able to obtain a nearly identical model-result by using front-end filters that preserve the center-surround, even-symmetric RF structure that we and others have measured in the retina and LGN of the cat (e.g. Rodieck, 1965; Citron *et al.*, 1981; Citron, Emerson & Levick, 1988). We have restricted the form of these early spatial filters to an even-symmetric Gabor function of $\sigma = 2.5$ (see Fig. 3B,C), because spatially odd-symmetric RFs do not occur until the visual cortex. We obtained the approximate 90° spatial phase-shift needed between these two filters by offsetting the relative spatial position by 1. Therefore, the serial order and nature of transformations in a well-fitting model will be directly applicable to geniculocortical connectivity and coding in the mammalian visual system.

RESULTS

Cortical measurements

Results shown here are representative of the five DS cells out of seven striate cells from the complex family on which we have performed detailed spatiotemporal white-noise analysis. We have chosen as our most detailed example the same DS "C" cell (Henry, 1977) that we have analyzed previously for its subunit properties (Emerson *et al.*, 1987b). Two other examples of "B" cells from the complex family (Henry, Lund & Harvey, 1978) are shown in Discussion, and a nondirectional B cell is shown in Fig. 6 of Emerson *et al.* (1987b). DS *simple* cells (Hubel & Wiesel, 1962) show a similar dependence on nonlinear interactions (Emerson & Citron, 1988, 1989, 1992), but they will be considered in detail elsewhere.

To compare physiological measurements with models, we have selected 1- and 2-bar tests that probe the system carefully by eliciting different signatures in different RF types (Citron *et al.*, 1981; Emerson, 1988), and also for DS vs non-DS cells (Emerson *et al.*, 1987b). Bars were longer than the RF, which makes the experiment one-dimensional in space, as are the models. The first step is to show that a neuron will demonstrate DS responses to a stepwise moving stimulus that can be exactly mimicked by our discrete 16-bar CRT display. Figure 1A shows that stepwise movement of a conventional light bar across all 16 positions of the CRT (nine of which are shown in Fig. 1B) can elicit strongly DS responses, even when the bar remains at each position for 64 msec. This is a general result for DS members of either simple or complex cortical cell families (Emerson & Gerstein, 1977b; Emerson & Coleman, 1981).

Cortical 1-bar tests

To study DS we began by reducing the stepwise movement sequence to its separate single-bar spatial elements, shown in Fig. 1C. Estimates of responses to a 16-msec presentation at each single bar position in this test are shown in Fig. 1D in PST-histogram format. Here, dark bars were chosen because they elicited stronger responses. DS complex cells show the same preferred direction for light and dark bars. These Wiener second-order-kernel estimates of the 1-bar response include the nonlinear intensive or "diagonal" term of the Wiener second-order kernel, which in such a nonlinear neuron provides a considerably more accurate estimate of the dark-bar response than if only the first-order, or "linear" term was used (for details see Methods; Marmarelis & Marmarelis, 1978; Emerson *et al.*, 1987b; Mancini, Madden & Emerson, 1990).

The homogeneity in time and symmetry in space of the brief increases in firing rate around positions 7–12 for this neuron indicate that DS in this unit can not be explained by considering responses to single bars. That is, there is nothing in the single-flash responses to indicate DS, and a linear superposition of 1-bar responses in the two possible spatiotemporal sequences at a given speed would generate a similar prediction for either upward or downward motion (e.g. Fig. 1E of Emerson & Citron, 1989; or Fig. 1F of Emerson & Citron, 1992). A correlate of this property is that these 1-bar responses, when plotted in the contour format of Fig. 1E, show response domains that lack oblique orientation in space–time (Adelson & Bergen, 1985). In the remaining figures, we use contour plots for two-dimensional (2D) functions because they illustrate well the shapes of these 2D relationships.

Cortical 2-bar tests

As 1-bar tests do not explain DS in these cells, the cells' response must depend on nonlinear interactions

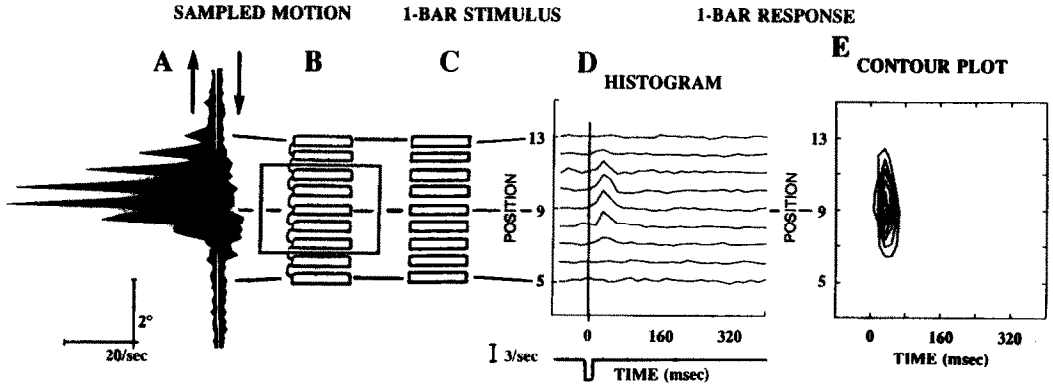

FIGURE 1. Sampled motion in a directionally selective (DS) striate complex ("C") cell (cell 1–3). (A) Averaged responses to 180 upward (left bins) and downward (right bins) stepwise sequences of a conventional light bar moving across the RF, as shown in B, at a mean speed of 7.8 °/sec (note upward preference). Vertical axis represents position; time–course of the response is reversed for downward movement (note arrows) to align bar positions spatially for opposite directions. Each position was maintained on the video screen for 64 msec; bin resolution, 16 msec. (B) Successive contiguous positions of the 0.5° × 11° bar moving stepwise across the RF (large rectangle). (Schematic bars of B & C are shorter and narrower than the bars used for experimental mesurement, which assumed contiguous positions, and were longer than the RF.) (C) Individual positions 5–13 of the sequence in B. Such isolated stimuli comprise, spatially, the first-order elements of a moving stimulus. (D) Second-order Wiener estimate of (time-locked) responses to single 16-msec stationary dark-bar flashes at each position represented in C, including an intensive component. Time-course plotted with 16-msec resolution, referred to the presentation time. (E) Same data as in D, but plotted as contours to emphasize the lack of oblique orientation in space vs time. Contour lines map out the locus of points with the same response value. Contour slices are chosen at equally spaced values, except that the contour at zero is never drawn in these plots because it is dominated by noise. Amplitude calibrations in all figures are expressed as firing rate (spikes/sec) at the bar-luminance used, and were calculated using the maximum 16-msec bin. Range of data here and in D −0.351 to +2.64/sec. The positive range in E was divided into six contour intervals, each 0.440/sec. These values apply specifically at this luminance and, except for A (which was the only conventional measurement reported in this paper), to the highly gain-adapted conditions of the white-noise experiment in which the responses were measured (for details see Methods of Emerson *et al.*, 1987b). All white-noise measures for this cell represent 35 min of random-stimulus presentation. Mean luminance was 222 cd/m^2, with 100% modulation. Retinal eccentricity was 11.7°.

between at least two positions in space and time. Therefore, our second test used two bars to capture the elements of motion to which 1-bar tests are blind: direction (spatiotemporal order) and speed (spatial and temporal interval). Because we already know the responses to each single bar from the above test, the new information desired is the nonlinear interaction between the individual responses that make up this elemental sample of motion. Figure 2 shows an example of a linear superposition test for a bar presented at position 10 followed by a bar at position 9, a sequence in the downward, nonpreferred direction for this cell. The stimuli are represented in space and time as small circles in the boxes of Fig. 2A,B, where the vertical axis shows spatial location of the presentation and the horizontal axis shows the time of the presentation, always referred to the bar at the center of B, called the "reference bar". Below each box is shown the response of this same DS complex cell to the two impulsive bar presentations (stimulus time–course indicated below the response). Figure 2C represents the 2-bar stimulus as a pair of circles, the open circle representing the "neighbor bar", which can take any value over the full range of positions and times within the space–time coordinate box. The fixed reference event is represented by the solid circle. Below the spatiotemporal bar-pair is shown, not the measured 2-bar response, but the Wiener second-order estimate of the *difference* between the linear superposition of separate bar responses and the raw response itself, i.e. the nonlinear residual, or interaction. The negative interaction shown in Fig. 2C strongly suppresses the expected 1-bar response to the second bar (Fig. 2B), which helps explain the weak responses for downward bar motion in Fig. 1A. (The amplitude in Fig. 2C is doubled with respect to that in A & B to illustrate the similar time–course of the interaction.)

To appreciate the velocity dependence of the spatiotemporal interaction around reference bar 9, we need to explore the result of sampling with the neighbor bar (open circle) the entire 2D spatiotemporal region within the box of Fig. 2C. The significant information about the interaction below the box is confined to the first 80 msec after the second pulse, the event that provides the first opportunity for nonlinear interaction. This property requires defining a new temporal variable, τ, whose origin is at the occurrence of the second event. Previous work has shown that the τ-course of a DS interaction is monophasic and independent of the temporal bar separation ($\Delta\tau$) or of the spatial bar separation (ΔS) (see Fig. 2 of Emerson & Vaughn, 1987 for a $\Delta\tau$ vs τ plot, and Discussion below). Therefore, to allow plotting the interactions over 2D coordinates that include space, henceforth, we will suppress the τ variable by summing the interaction over the first 80 msec to derive a single number (scalar) for each 2-bar combination (except in Fig. 5B, which had a 48 msec summing period; see legend).

Figure 2D shows a contour plot of the summed interaction in a format that will be used below for comparing measured with modeled results over the full

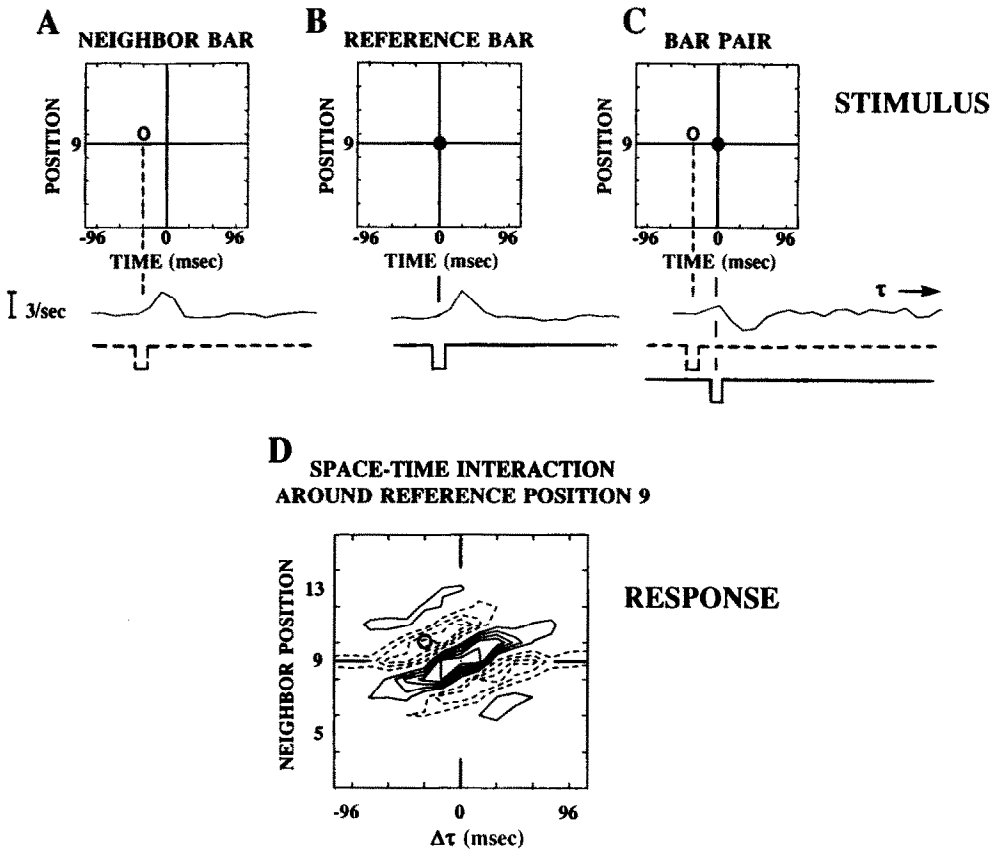

FIGURE 2. Calculating a 2-bar nonlinear interaction for the same DS neuron. (A) Location of the variable, or "neighbor" bar in space–time is designated with an open circle (here, a dark bar at position 10). The neuron's response, taken from Fig. 1D, is shown below in the correct temporal position, along with the time–course of the stimulus, 32 msec before the bar in B. (B) Location of the fixed, or "reference" dark bar at position 9, but always shown as a solid circle at the center of the spatiotemporal coordinates, with the time–course of its response and the stimulus shown below. (C) Spatiotemporal representation of this particular dark-bar pair. The time–course of the nonlinear residual from a superposition test of the two bars is shown below as a function of time, τ, after the presentation of the later stimulus. The interaction is calculated as the Wiener second-order spatiotemporal kernel, which is the second-order estimate of the response to the bar-pair minus the linear sum of responses in A and B. It is plotted at twice the amplitude of A and B to emphasize the similarity in time–course. Therefore, the negative interaction in the downward, null direction for this two-position sequence, alone, suppresses to about one-half the expected response to the later occurring bar at position 9 to B. (D) Spatiotemporal dependence of nonlinear interactions around reference position 9, plotted over the full two-dimentional (2D) range of 16 possible 0.5° *neighbor* positions, and 15 temporal separations ($\Delta\tau$) covering ± 112 msec. Dashed contour lines denote negative regions of the interaction. Kernel values, such as in C, have been summed over the first 80 msec (5 bins) following the later stimulus to create a scalar value that can be plotted in D. Interaction values have been normalized from the 5-bin sum to the peak time-bin of the interaction by multiplying by 0.391, a shape factor determined experimentally for this cell. Range of normalized data: -0.910 to $+1.12$/sec, with six positive contours. The location of the particular suppressive interaction illustrated above in A–C is denoted by an open circle at the bottom of the valley in D, at the same 2D location as the neighbor bar in A,C. Responses of this cell were DS in Fig. 1A because upward movement elicited strong nonlinear facilitation along the positive ridge, whereas the many pairs of bars constituting a downward movement sequence suppressed responses that were expected from the 1-bar tests (see text). The incomplete suppression shown here by the low amplitude in C is caused both by our limiting the contribution to only a single pair of positions, and by this second-order measurement's inability to exclude light-bar effects, which were weaker. Inclusion of third-order effects predicted stronger suppression (see Fig. 1 K vs J of Emerson *et al.*, 1987b).

spatiotemporal region. Solid contours denote the region of constant positive interaction above the expected linear response (mainly facilitation in the shape of an elongated, elevated ridge). Dashed contours denote negative regions of the nonlinear interaction (suppression in the form of two elongated valleys). The strongest suppression is marked by the open circle at the bottom of the upper valley in D, and corresponds to the bar-pair illustrated in C.

As suggested previously (Emerson *et al.*, 1987b), a compelling interpretation for the obliquely oriented ridge-valley structure of the interaction in Fig. 2D is that it provides a basis for velocity selectivity in single neurons. Because the coordinates of this plot are space vs time, each possible constant-velocity bar-traverse includes nonlinear interactions occupying loci only along a straight line that intersects the reference position at the center of the plot. At this position during sampling a traverse, the neighbor and reference bars coincide ($\Delta S = \Delta\tau = 0$), which tests intensive nonlinearities through a doubly intense bar. The slope of the line equals the bar velocity in °/sec. If the neighbor bar, or any moving bar, is stepping in the upward direction at the appropriate speed to create a trajectory with an average slope matching that of the positive interaction ridge (here, about $+16$°/sec), it will benefit maximally

from nonlinear facilitation. At that velocity strong facilitation is elicited between each position in the sequence and its brief presence when it occupied the reference position. Trajectories associated with other velocities can evoke strong suppression in the valleys, especially for negative slopes, which occur for the null direction. See Discussion for consequences of this shape for filtering properties in the frequency domain.

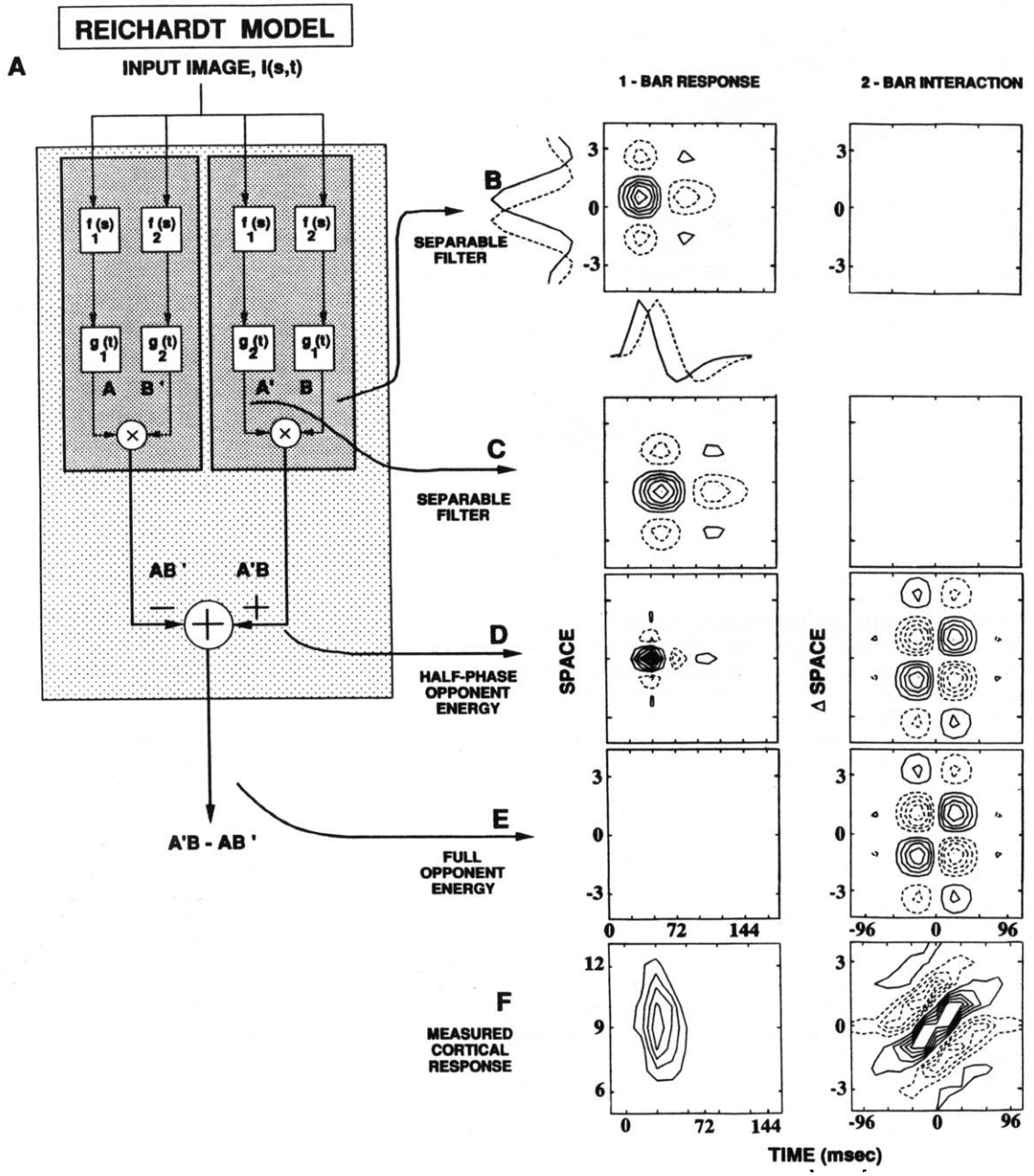

FIGURE 3. (A) Version of the Reichardt model that incorporates a layer of bandpass spatial filtering [the $f_n(s)$'s], followed by a layer of bandpass temporal filtering [the $g_n(t)$'s] to achieve an elaborated model that is appropriate for mammalian vision. (B) Separable spatiotemporal impulse response resulting from the cascaded spatial and temporal filters indicated by the *solid* curves in B (see text for details). Coordinates and contours as in Fig. 1E, but with dashed regions negative. Model contour intervals have been chosen arbitrarily to best illustrate the 2D shape of the responses, but are always equal. (C) Separable spatiotemporal impulse response resulting from the cascaded spatial and temporal filters indicated by the *dashed* curves in B. This filter is approximately in quadrature in space and time with respect to that shown in B. (D) Multiplying the outputs of the separable filters of B,C produces a nonlinear response, seen in the 2-bar interaction (right). Note that the 1-bar response (left) also remains separable. All modeled 2-bar interactions are in the format of Fig. 2D, but to improve the s/n and remove the slight asymmetries in interactions around individual reference positions, they have been averaged over the 16 reference positions that contributed to the nonlinear interaction. (E) Subtracting the two "half detectors" cancels 1-bar responses, and shows the same 2-bar motion opponency as in D. (F) Physiological measurements differ from every level of the Reichardt model, both for 1- and for 2-bar responses. Wiener-kernels for the measurement of F, right have been averaged over reference-positions 7–11, and then normalized to the scale of the maximal interaction bin for reference position 9 of Fig. 2C by multiplying unaveraged values of Fig. 2D by an experimentally determined 1.38, in addition to the factor of 0.391 used to correct for temporal summation. Range of data in F, left: same as in Fig. 1E; normalized range in F, right: −0.864 to +1.07/sec. D,E, right and F, left plotted with five contours; B, C, and F, right used six; and D, left needed eight to see the secondary regions.

The shape of this interaction is the same for reference positions chosen across the RF (Fig. 3 of Emerson *et al.*, 1987b). Therefore, in complex cells, velocity selectivity seems to depend on a family of similar, overlapping subunits, whose characteristics can be seen best by averaging the interactions around all active positions across space, each centered around the reference bar (Fig. 4 of Emerson *et al.*, 1987b; right sides of Figs 3F and 4H here for the same unit as in Fig. 2D). We have used this spatially averaged interaction, called the "motion kernel" to compress the available 4D information from 2-bar tests of motion into a single 2D plot. Below, we use this construct to compare measured with modeled interactions. Although we have averaged across all space for the modeled 2-bar interactions, the range is not critical because unaveraged measured or modeled 2-bar interactions show similar structure with a maximum near the reference bar, even for simple cells (e.g. Fig. 3 of Emerson & Citron, 1989, 1992).

1- and 2-bar tests of the Reichardt model as compared with the cortex

As the 2-bar interaction of Fig. 2D and its spatial average show a characteristic shape that may be diagnostic for the underlying mechanisms, we applied both 1-bar and 2-bar tests to two models that depend on different mechanisms, a motion-opponent model, as exemplified by the Reichardt model (1957), and the unidirectional motion-energy model of Adelson and Bergen (1985). We begin in Fig. 3A with the geometry of a motion-opponent model whose front end matches the elaborated Reichardt detector proposed by van Santen and Sperling (1984, 1985). A difference from earlier descriptions is that we have regrouped the elements pictorially in A to emphasize the functional subunits and serial stages of the model. Two columns of boxes in the righthand portion of the figure show responses in space vs time for elements of the pictorial model, the left column for 1-bar inputs, and the right column for 2-bar interactions.

Within each darker rectangle in A is a half-phase subunit that depends on the product of two space–time-separable 2D linear filters, shown in B & C, each of which is generated by the outer product of one of the spatial and one of the temporal functions shown to the left of B and between B & C, respectively. The resulting center-surround even-symmetric filters are shifted in space with respect to each other (see Methods) and have different temporal phase to achieve approximate spatiotemporal quadrature (discussed also by van Santen & Sperling, 1985). Each of the half-detectors, which produce the signals AB′ and A′B, are motion opponent, but with phase-dependent oscillations. When the difference is taken at the last stage, the final signal is also motion-opponent, but the phase-dependency has been removed.

As the first stages are linear, they exhibit no nonlinear interactions; thus the panels at the right of B & C are blank. The nonlinear interaction first appears in panel D-right, after the multiplicative stage in the half-detector. Note that although the 1-bar responses in B–D (left), are displayed in the same dimensions as Fig. 3F-left, they differ considerably from the unimodal complex-cell impulse–response by showing phase sensitivity, i.e. alternating positive and negative regions in space and time. The 2-bar interaction shows a DS preference for upward motion, but also indicates full motion opponency. That is, the space–time separable interaction shown in D-right (and also for the motion-opponent stage of E-right) shows as strong, but negative, response interactions for downward as for upward movement. This fundamental motion opponency has been pointed out by van Santen and Sperling (1985). When the front-end spatiotemporal filters are bandpass, as here, motion opponency occurs at the half-detector level, as well as at the final output stage. By the time we reach this fully opponent stage in E, the model shows no responses at all to single bars, and shows antisymmetric motion-opponency for bar pairs (see below). This restriction of the response to moving stimuli is a well-known property of Reichardt models (e.g. Poggio & Reichardt, 1976), and in a mathematical sense, specializes them for carrying purely directional information.

The Reichardt model makes several clear predictions about the 2-bar interaction: (1) the interaction should be zero along the $\Delta S = 0$ and $\Delta\tau = 0$ axes; (2) the interaction should be separable in space and time; and (3) the interaction should be antisymmetrical about the spatial and temporal axes, i.e. the interaction retains the same amplitude but opposite sign when the sign of ΔS or $\Delta\tau$ is inverted, which generates a "checkerboard" pattern. None of these predictions is upheld in the data. In addition, no stage correctly predicts the unimodal character of the 1-bar response. The incompatibility derives from a fundamental difference in symmetry of the interactions, and would not disappear with changes in spatial or temporal filter parameters.

Comparison of measured cortical responses with the energy model

As neither the 1-bar nor 2-bar cortical tests matched any stage of the Reichardt model, we next tested the motion energy model, as proposed by Adelson and Bergen in 1985. Figure 4A shows a drawing of the energy model presented in Fig. 18B of the 1985 paper, except that it has been reorganized pictorially to facilitate comparison of the functional subunits and serial stages of information-processing with those of the Reichardt model in Fig. 3A. (The positions of the mirror-symmetric unidirectional models also have been exchanged to agree with the upward preferred direction of the Reichardt model in Fig. 3.) In Fig. 4B,C the same linear separable filters as in the Reichardt model of Fig. 3B,C shows no 2-bar interaction, nor does their linear sum, shown in D-right. However, the responses of D and I differ from that in Fig. 3D by being generated through summation rather than multiplication of these front-end separable filters (Watson & Ahumada, 1983). The resulting filters in D-left and I are obliquely oriented, i.e. space–time inseparable, and differ relative to each other by being shifted in phase by about 90° (i.e. in quadrature) in both space and time.

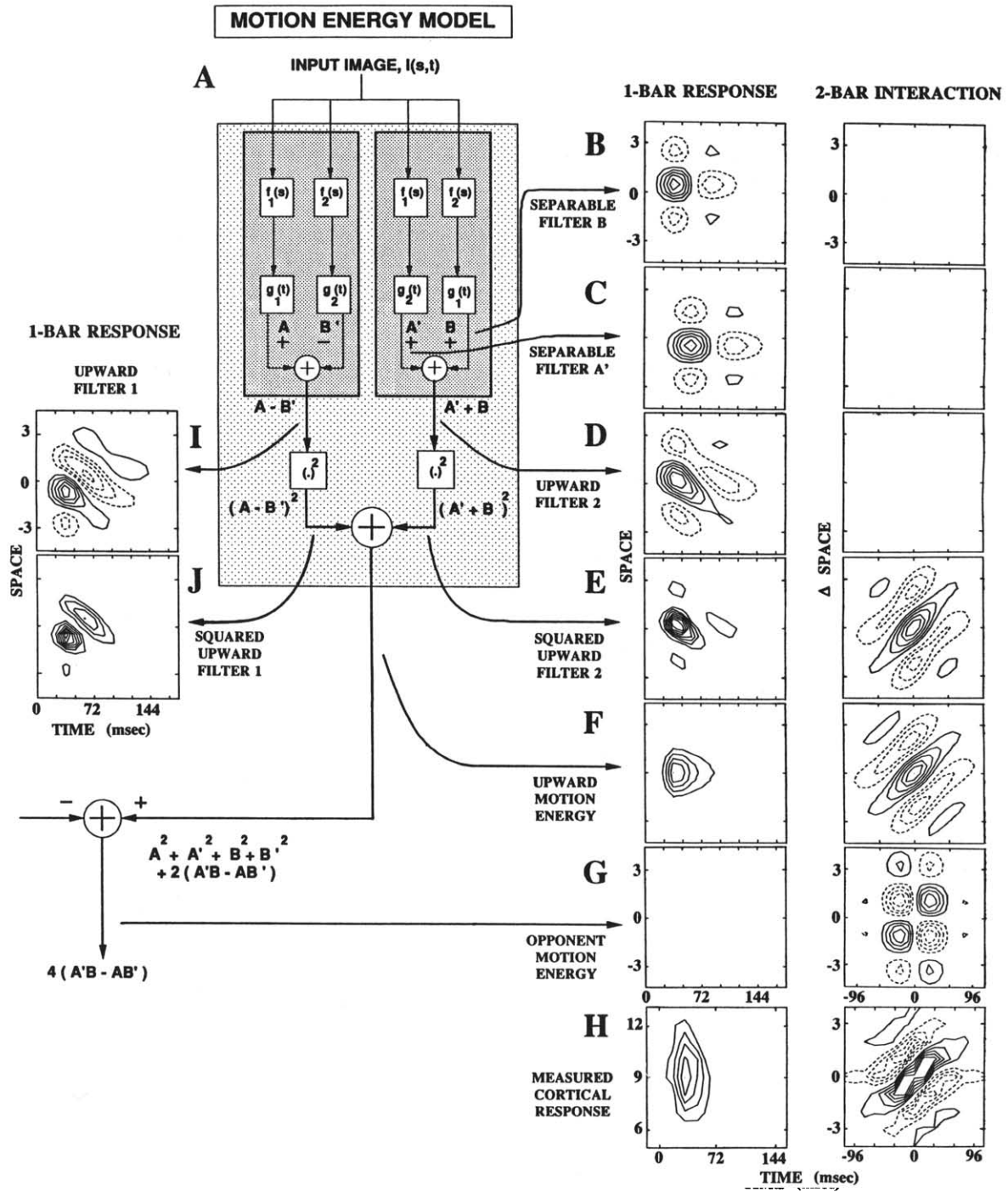

FIGURE 4. (A) Structure of the basic motion energy model in same format as in Fig. 3. B,C same separable front-end filters as in the Reichardt model (Fig. 3). (D,I) *addition* of filter outputs, rather than multiplication in Fig. 3D, creates *inseparable* linear filters that are shifted in both spatial and temporal phase relative to each other by 90°, i.e. they are in spatiotemporal quadrature. (E,J) Squaring retains much of the oblique orientation, and generates DS 2-bar interactions that are inseparable. (F) Combination of the squared outputs from quadrature partners generates agreement with both 1- and 2-bar physiological measurements in H. The smaller extent of model 1-bar responses, however, suggests multiple subunits in the complex cell. (G) Combination with output of a mirror-symmetric energy model generates responses that are identical to those of the Reichardt model in Fig. 3E, and inconsistent with the observed responses. (H) Cortical measurements agree with model results only at the unidirectional energy stage of F because of basic symmetry differences with other stages of the energy model. All data plotted with six contours, except that F, left, G, and H, left used five contours, and E, left and J needed eight contours.

The stages following the squaring operation, shown in Fig. 4E,J are the first stages to generate a nonlinear interaction. The interaction roughly resembles that at the unidirectional energy stage (F-right), which matches the measured cortical response (H-right), but the 1-bar responses in E-left and J display incorrect predictions of spatiotemporal structure. The problem is solved in the unidirectional energy response, shown in F-left, by summing the two squared responses to give a phase-independent measure of unidirectional motion energy. The decay period of the temporal function in equation (1) was chosen to ensure that the duration of modeled 1-bar

responses matched the cortical responses, and the spatial and temporal oscillation-period matched those of the measured 2-bar interactions.

The 2-bar interactions shown in Fig. 4E,F (right), however, retain the phase dependence of the input filters in D-left and I, because initial summation of the two bar-responses occurs before the squaring operation. The augmentation and cancellation at the two phases of the input filters from which the two 1-bar responses are sampled creates the interaction. The preferred velocities implied separately by the slopes of the oriented domains in the 1-bar test of D-left vs the 2-bar test of E-right (both in °/sec) agree with each other because the slope values are inversely related. That is, the arbitrary ΔS-convention that we chose for 2-bar interactions dictates an upward directional preference for *positive* slopes, in agreement with the enhanced modulation from upward moving images convolved with filters having *negative* 1-bar filter slopes (D-left and I; cf. Fig. 7 of Adelson & Bergen, 1985). The unidirectional energy stage, F, is labeled "upward" here, but obviously could refer to any movement direction, depending on the spatially 2D orientation of the input filters.

The major result of this study is the satisfying agreement between modeled 1- and 2-bar responses in Fig. 4F, with cortical responses of H, measured with the same DS complex cell as shown in Figs 1 and 2. As shown in Appendix A, the similarity between modeled and measured 2-bar interactions, made more reliable by averaging across space, indicates that the autocorrelation functions of the subunit filters are similar. Therefore, the spatiotemporal periodicity (frequency response) and spatial extent of complex-cell RF-subunits match those of the model subunit filters in Fig. 4B–D,I. But if the complex cell had only a single pair of subunit filters, such as those in the single-subunit model of Fig. 4, then the vertical extent of its 1-bar response in H would match that of F. Its larger extent is strong evidence that complex cells use pooling from an array of spatially distributed subunits (corroborating Fig. 8 of Movshon, Thompson & Tolhurst, 1978; and Fig. 3 of Emerson *et al.*, 1987b; also see Discussion here). Except for this difference in spatial extent of 1-bar responses, and minor differences in shapes of contour domains, it appears that the energy model provides an excellent representation of DS cortical responses to our 1- and 2-bar tests.

Finally, the output of an opponent formulation of two basic energy models is shown in Fig. 4G. This is the site at which the response is equivalent to the output of the Reichardt model. After subtracting the response from a mirror-symmetric, downward-selective energy unit coming from the left (structure not detailed), the 1-bar test elicits no response at all, which differs from both the measured cortical response and the modeled unidirectional energy response. This lack of 1-bar responses occurs because the opponent combination represents the difference between the energy term for each direction; and flashing bars, counterphase-modulated gratings, and many similar stimuli exhibit equal upward and downward energy.

Even more significant is the discrepancy in shape of the 2-bar opponent interaction in G-right from measured and modeled unidirectional results. As in the Reichardt model of Fig. 3E, downward motion is predicted to elicit strong negative responses. On the other hand, measured downward responses of Fig. 1A and corresponding measured interactions that would be integrated along a trajectory with a negative slope in Fig. 4H-right, indicate approximately zero response. Adelson and Bergen (1985), and van Santen and Sperling (1985) have pointed out that the similarity between *opponent* responses in an opponent formulation of energy models (Fig. 4G) and the Reichardt model (Fig. 3E) reduces to a formal equivalence when, as here, the models are built with appropriate spatial and temporal filters. However, the basic (unidirectional) motion-energy stage has no equivalent in the Reichardt model, and the only location in the energy model at which both 1- and 2-bar tests agree with the physiology is at the unidirectional energy stage shown in Fig. 4F.

The Reichardt model, like the opponent formulation of two energy models, takes the difference between two halves as its final step. But the signals being subtracted are quite different in the two cases. In the opponent formulation of energy models, the penultimate signals are phase-independent, non-opponent (unidirectional) motion signals; in the Reichardt model, they are phase-dependent and motion-opponent signals. In the opponent version of the energy model, the final stage *produces* opponency; in the Reichardt model, the opponency is present from the start, and the final stage produces phase-independence.

DISCUSSION

We have used a pair of sensitive tests to compare measured directionally selective (DS) complex cells in the striate cortex of the cat with the performance of two motion models on these same tests. Motion sequences elicit DS responses in a large proportion of cortical neurons (Emerson & Coleman, 1981), and in both models, but responses to 1- and 2-bar tests differ considerably between the two models.

The range of complex-cell responses

Before exploring physiological implications and comparing models, it is helpful to know the range of cortical cell properties one encounters. Of the five DS out of seven complex cells studied in detail, Fig. 5A is the most similar, and B is the most different from the unit illustrated in Figs 3 and 4. The two remaining DS cells resembled Fig. 5A. The chief difference between the cell in Fig. 5B and all of the others is a lack of an intensive nonlinearity at the point where $\Delta S = \Delta\tau = 0$. This lack of a purely intensive nonlinearity indicates a departure from a pure squaring operation, and may signify a strong saturation at otherwise large output amplitudes. Note also the large slope value for interaction domains in B, which indicates that this unit should have a higher preferred speed than the unit in A. Differences aside, all

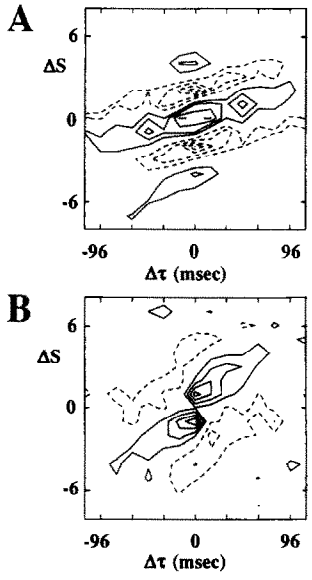

FIGURE 5. Other directionally selective complex cells. (A) 2-bar interaction for complex-family "B"-cell (Henry *et al.*, 1978) number 6–6–32 represents 4 of the 5 DS complex cells studied extensively. Interaction was averaged over six active positions of the RF, and summed over an 80-msec τ-period after the second stimulus. Note similarity to plots of Figs 2D and 4H, right. Range of raw kernel values: -3.14 to $+2.54$/sec in five negative contours (not normalized for temporal or spatial averaging). (B) 2-bar interaction for B-cell number 6–2–6. Interaction was averaged over six active positions of the RF, and was summed over a 48-msec τ-period beginning 32 msec after the second stimulus. Chief differences from A are higher preferred speed and small intensive nonlinearity (at $\Delta S = \Delta\tau = 0$, see text). Range of raw kernel values: -0.748 to $+1.38$/sec in five positive contours. Measurements for each cell represent 6.5 min of random-stimulus presentation.

five units show elongated, obliquely oriented domains, whose slopes agree with their preferred direction.

This oriented Gabor-like structure in space vs time is the signature of a nonlinear filter that is space–time inseparable, and has certain correlates in the frequency domain, as discussed by Daugman (1985), and Emerson *et al.* (1987b). All the energy lies in either the odd or even quadrants, which limits neural responses to a range of spatial and temporal frequencies for motion in one direction, showing no response for the null direction. That is, the neuron is a unidirectional spatiotemporal filter both in space–time and in frequency, but does not respond to image velocity *per se*, independently of spatial and temporal frequency (Fig. 10 of Watson & Ahumada, 1983; Fig. 11 of Adelson & Bergen, 1985; and Fig. 8 of McKee, Nakayama & Silverman, 1986). Hamilton *et al.* (1989, p. 1290), and Emerson *et al.* (1987b, p. 49) each have discussed how the spatiotemporal filters can be inseparable in either space–time or 4-quadrant frequency domains, but appear separable in frequency when only one frequency-quadrant is viewed (for the preferred direction).

The spatial-subunit makeup of DS complex cells

The disparity in spatial extent between modeled and measured 1-bar responses in left sides of Fig. 4F vs H indicates that DS complex cells have more than one spatial subunit in their RFs, because our matching of model 2-bar interactions to measured ones constrains the size but not the number of the model *subunits*. To determine the minimal number of subunits that would explain the spatially broad complex-cell responses, we have generated a quantitative prediction in Fig. 6B for the measured 1-bar responses of Fig. 6A as the coursest set of equally separated subunits of the shape shown in Fig. 4F-left that would smoothly approximate the physiological measurement. Availability of 2-bar measurements provides a rare opportunity to investigate the convergence of RF subunits at one stage onto the next. With a spatial resolution that is twice that of the physiological measurement, the model of Fig. 6B shows that it was necessary to have at least four subunits, each separated by one spatial step of the physiological measurement (e.g. 1.5 steps was too coarse). Because the decay-time, $1/k$, of the model filters in Fig. 4B,C was chosen slightly too long, it was decreased by 10% for this figure only.

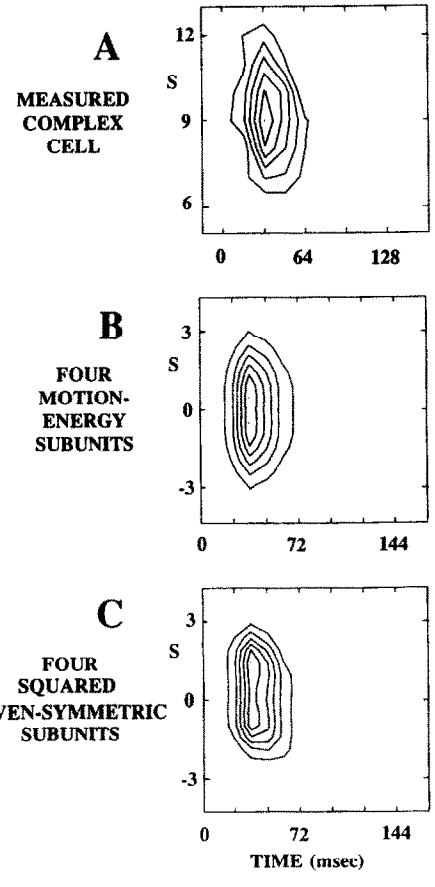

FIGURE 6. Constitution of complex-cell spatial subunits. (A) Measured 1-bar response for C-cell number 1-3. Parameters of Fig. 4H, left, except that all plots here used six contours. (B) Modeled responses for 1-bar upward motion-energy responses of Fig. 4F, left, but replicated four times in space, with a constant separation of one physiological position (two model positions). Responses are similar to physiological measurement of A, but required that the separable-filter time-constant ($1/k$) be shortened from 0.67 in Fig. 4 to 0.60 to better match the physiological decay period. (C) Same geometry as above, except with squared upward response of Fig. 4E, left, the squared output of the spatially even-symmetric inseparable upward filter. Some of the oblique orientation in contours of Fig. 4E, left survived the spatial pooling, but differences among A, B, and C are too small to distinguish reliably between the two models.

Of course, more subunits over the same spatial extent would be indistinguishable from the coarsest set, and would only dictate more overlap between each subunit.

Once we admit that a cortical cell pools information from a number of spatial subunits, we must consider whether the phase independence in our 1-bar measurements now *requires* the odd-even inseparable filter-pairs of the energy model shown in Figs 4D-left, I, and schematized in Fig. 4A. For example, in another system (the lobular plate of the fly), Egelhaaf and colleagues (Egelhaaf & Reichardt, 1987; Borst & Egelhaaf, 1989) have shown that *spatial integration* of like-subunits that are distributed over space performs the same function, removing the phase-dependence of earlier stages. We already know that the lack of agreement in the 2-bar interactions of separable models such as the Reichardt model removes its connectivity from consideration in the striate cortex, because the 2-bar interactions we have measured and modeled reflect the properties of the subunits themselves, independently of how they are pooled. But to determine whether pooling of the *single* squared 1-bar response of the spatially even-symmetric upward filter in Fig. 4E-left could also explain 1-bar responses of Fig. 6A, we have modeled in Fig. 6C the same squared response, integrated over the same coarse spatial array as shown in Fig. 6B. Although the model of Fig. 6C retains more of the oblique orientation of the subunit's squared response than the model of B, the similarity of the models and the noise in the cortical measurement of Fig. 6A preclude a reliable assessment of the need for summation of pairs of squared subunits with quadrature phase.

It should not be surprising that 2-bar interactions for pooled, shifted versions of single or paired quadrature filters are similar, because the autocorrelation properties of the linear quadrature-pair filters (Fig. 4D-left, I) are similar. We know this because Appendix A shows that spatially averaged 2-bar interactions in right sides of Fig. 4E,F are simply the average autocorrelation functions of the single or combined filters, respectively; and those 2-bar interactions are similar. To examine still further the correspondence between physiology and the model, we have investigated the third dimension, τ (the time–course of the interaction), which is present in Fig. 2C, but missing in all other 2-bar interactions shown here. As mentioned above in the context of suppressing this variable to allow spatiotemporal plots, the time–course of physiological interactions for a single pair of positions is found to be independent of the temporal separation. $\Delta\tau$. This property leads to separable 2-bar interactions when they are plotted in $\Delta\tau$ vs τ (Fig. 2 of Emerson & Vaughn, 1987), and as it turns out, characterizes models with either the single-subunit quadrature-pair of Fig. 4F, or the four-subunit models based on Figs 4E-left or J (not shown). This separability can not be mimicked by any single linear–nonlinear sequence of operations; and yet it is fully explained by the convergence of a single pair of model spatiotemporal filters in quadrature. The significance of this observation is that because the physiological and the model 2-bar interactions have the same τ-course as the 1-bar responses, they contribute no time–course of their own, but rather modulate the amplitude or *gain* of the response to the second bar (see Larkin, Klein, Ogden & Fender, 1979), increasing it for sequences in the preferred direction, and decreasing it for the null direction. Apparently convergence of as few as two different phases is an effective way of modulating the response amplitude for movements in opposite directions without changing the time–course (see Fig. 1H of Emerson & Citron, 1991, for contributions of these spatiotemporal "kernels" to the shapes of movement predictions, even in *simple* cells).

This fit of the model over all three dimensions suggests that spatial pooling, after squaring of a broad variety of possible space–time inseparable subunits could explain cortical-cell responses. This conclusion, in turn, suggests that in the cortex, the definition of the energy model should be broadened to include squaring of space–time inseparable subunits that may be shifted in spatial phase (position), rather than having explicit spatiotemporal quadrature.

Geniculocortical calculation of motion energy

An important function of a well-fitting model is to suggest ways in which the geniculocortical physiology might actually compute these spatiotemporal and intensive transformations. Retinal RFs could account for the spatial properties of the basic energy model (though this need not be the case), and we believe that the temporal phase lag may be provided to the cortex by so called "lagged" X and Y cells, described in the cat's lateral geniculate nucleus by Mastronarde (1987a,b), and by Humphrey and colleagues (Humphrey & Weller, 1988a,b; and Saul & Humphrey, 1990). The nonlinear intensive transformation, the squaring operation, could be accomplished easily as shown in Fig. 7A, by applying a gradual or "soft" threshold (e.g. a half-squarer composed of a half-wave rectifier followed by a squarer) to inverted and noninverted versions of the appropriate signals. Recent systems identification approaches suggest that neural threshold functions in simple cells might actually approach half-squarers (Emerson, Korenberg & Citron, 1988, 1989), and that complex cells approximate a full-squaring operation (Emerson, Korenberg & Citron, 1990a,b, 1992). Pollen, Gaska and Jacobson (1989) also have used half-squarers in a proposed quadrature complex-cell model. Heeger (1991) has discussed a useful application of half-squarers, especially in conjunction with a contrast gain-control.

In light of the above discussion about the phase properties of constituent subunits, the needed quadrature phase-shifts in space, and even the phase inversions, might be accomplished by soft-thresholding the outputs of appropriately shifted spatial samples from periodic spatial filters. The chief impact of adding modeling studies to our physiological measurements has been to simplify vastly our interpretation of seemingly complicated nonlinear spatiotemporal interactions in complex cells by showing that a natural nonlinear transformation

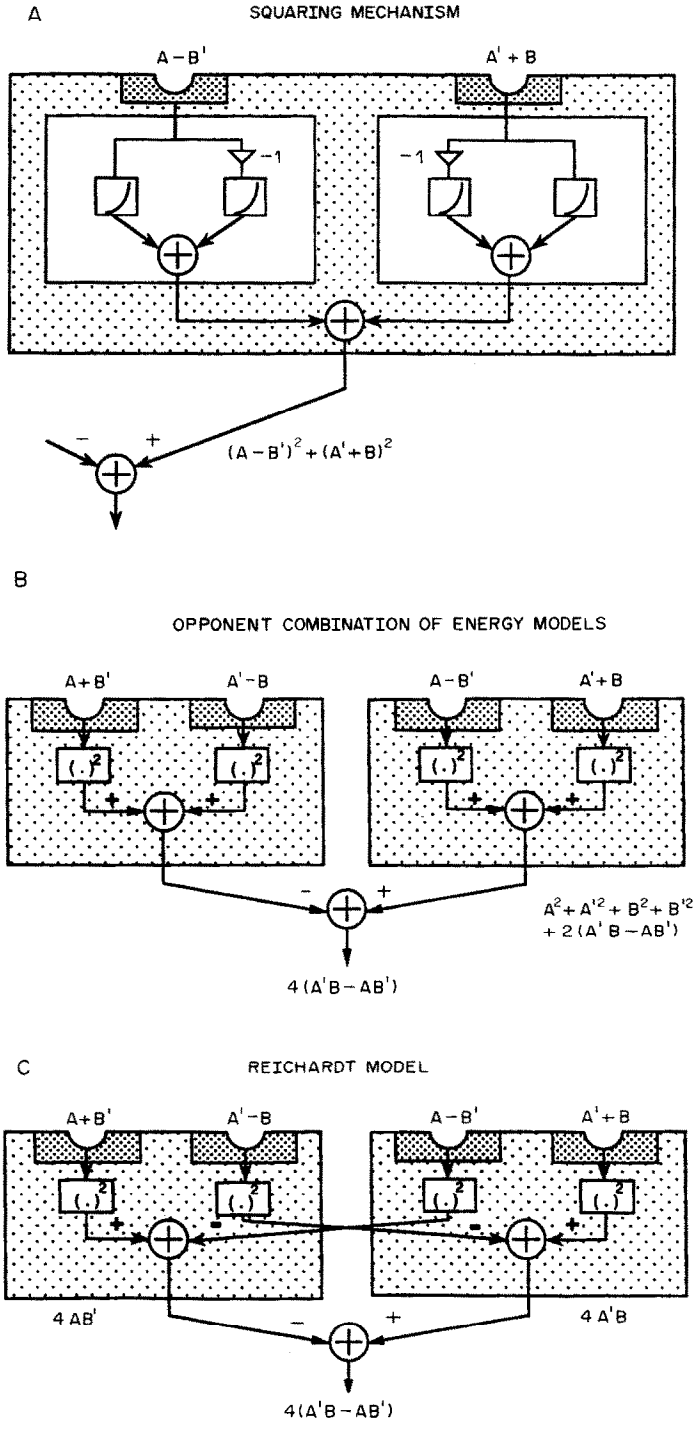

FIGURE 7. Illustration of a possible squaring mechanism and of the relationship in connectivity between the Reichardt and energy models. (A) A combination of inversion and static nonlinearity shows how a threshold-like half-squaring function could be used to carry out the neural squaring operation. This transformation could be as simple as passing signals of appropriate sign through cortical neurons whose spike-firing characteristics have a gradual onset. (B,C) A single exchange between connections of an opponent combination of energy models B is capable of generating both the pure product terms needed at the half-detector stage and the fully opponent output signal of the Reichardt model C. For details see text of Appendix B.

such as a threshold may generate these phenomena by following a linear space–time inseparable filter. The neural identity of these inseparable spatiotemporal filters is our last subject of discussion.

Results shown here and elsewhere have established a need in all DS complex cells for a family of early inseparable filters, whose properties then appear in the 2-bar interactions because of later-occurring nonlinearities. The lone account that dissents from this principle (Baker & Cynader, 1988) was an interaction experiment with two conventional bars. Any 2-bar measurments that employ subtraction of null-sequence responses from preferred-sequence responses (their Figs 3B, 4B & 5) can cancel underlying inseparable spatiotemporal

interactions, as has been discussed already (pp. 49–50 & 56 of Emerson *et al.*, 1987b). However, we also have examined the relatively *separable* raw interaction data for their Figs 3A and 4A, which were calculated with the correct procedure of subtracting the explicitly measured 1-bar responses. A discrepancy remains between their separable and our inseparable interactions; it may be caused by a difference in operating points between their conventional flashing-bars and our white-noise experiments. Recent evidence from experiments with drifting gratings, which we believe to establish a highly adapted operating point similar to that in the white-noise experiment, supports the dependence of DS on *inseparable* filters in striate neurons of both cat and monkey (Hamilton *et al.*, 1989). Figure 7B,C shows the relationships between motion-opponent (separable) combinations of basic energy models and the Reichardt model, and Appendix B provides a more quantitative comparison between the two models.

Although we have demonstrated inseparable subunits in DS complex cells, it is not obvious whether these filters correspond to distinct recordable cells, or are incorporated into the connectivity and dendritic organization of the cortical cell. On the one hand, DS simple cells would seem to make ideal candidates for this function, because of demonstrations of inseparable 1-bar responses in the cat (Reid *et al.*, 1987; McLean & Palmer, 1988, 1989). Simple cells even show the required range of spatial phases to generate quadrature (Pollen & Ronner, 1981; Field & Tolhurst, 1986). And in the monkey, recent measurements of DS simple cells in V1 also suggest spatiotemporal inseparability in 1-bar responses (Jacobson, Gaska, Chen & Pollen, 1989), and agree with an inseparable linear quadrature model preceding any of the nonlinearities in simple cells (Hamilton *et al.*, 1989).

On the other hand, *linear* space–time inseparable models (e.g. Reid *et al.*, 1987) seem to account for only about half of the DS measured in simple cells. A demonstration of simple-cell DS *without* a significant contribution from inseparable *1-bar* filters would support the likely *nonlinear* generation of DS directly from LGN afferents, because simple cells are presumably first-order cortical neurons. We have found that among nine DS simple cells examined, five showed no obvious inseparability in their *1-bar* responses (Emerson & Citron, 1988, 1989, 1992). Furthermore all 14 DS units studied, whether complex or simple, showed inseparable *2-bar* nonlinear interactions of the appropriate slope. This association of DS with inseparable nonlinear subunits, and the absence of DS at the preceding LGN stage suggests that generation of these space–time inseparable subunits may occur independently of any explicitly recordable cortical cell-type that shows the 1-bar properties of a single inseparable subunit. Generation of these inseparable nonlinear subunits for computing motion energy in both simple and complex cells is apparently an important general function of the visual cortex.

SUMMARY

1. Many models have been proposed for directionally selective (DS) cells in striate cortex. It has been proposed that initial stages of DS could be created by the action of linear filters that are spatiotemporally oriented; and several investigators have reported recordings in DS simple cells that are consistent with this function. It has also been argued that phase independence of the sort observed in DS complex cells could be achieved by a motion-energy computation, in which the outputs of linear units in the quadrature phase are squared and summed (Adelson & Bergen, 1985).

2. We have tested DS complex cells in the cat with random flashing bars, and have found that the responses are in good agreement with those predicted by a basic (non-opponent) motion energy model. Both the 1-bar responses and the 2-bar interactions behave in a manner consistent with such a model.

3. We have also tested the predictions of the classic Reichardt model [modified with prefilters as suggested by van Santen & Sperling (1985)]. This model's predictions are not consistent with measurements of either the 1-bar responses or the 2-bar interactions.

4. Both models have several stages, and the experimental methods allow us to test each stage as a candidate model for the DS complex cell. No stage of the classic Reichardt computation is consistent with the DS cell responses, and only a single stage of the energy model is consistent. The broad spatial distribution of subunits in complex cells obscures the phases of individual subunits, but leaves intact the requirement for inseparable filters and for the squaring operation, both provided by the basic motion-energy unit.

5. It is possible to create a fully opponent version of the motion-energy model by placing two basic motion-energy units in a push–pull configuration (Fig. 7B). In this case the opponent energy system would become equivalent to a full Reichardt model (Fig. 3E and 7C). However, such a computation is not consistent with what we observe in striate cortex.

REFERENCES

Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A, 2*, 284–299.

Andersen, R. A., Snowden, R. J., Treue, S. & Graziano, M. (1990). Hierarchical processing of motion in the visual cortex of monkey. *Cold Spring Harbor Symposia in Quantitative Biology, 55*, 741–748.

Baker, C. L. Jr & Cynader, M. S. (1988). Space-time separability of direction selectivity in cat striate cortex neurons. *Vision Research, 28*, 239–246.

Borst, A. & Egelhaaf, M. (1989). Principles of motion detection. *Trends in Neuroscience, 12*, 297–306.

Burr, D. C., Ross, J. & Morrone, M. C. (1986). Seeing objects in motion. *Proceedings of the Royal Society of London B, 227*, 249–265.

Citron, M. C. & Emerson, R. C. (1983). White noise analysis of cortical directional selectivity in cat. *Brain Research, 279*, 271–277.

Citron, M. C., Emerson, R. C. & Ide, L. S. (1981). Spatial and temporal receptive-field analysis of the cat's geniculocortical pathway. *Vision Research, 21*, 385–397.

Citron, M. C., Emerson, R. C. & Levick, W. R. (1988). Nonlinear measurement and classification of receptive fields in cat retinal ganglion cells. *Annals of Biomedical Engineering, 16,* 65–77.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A, 2,* 1160–1169.

Egelhaaf, M. & Reichardt, W. (1987). Dynamic response properties of movement detectors: theoretical analysis and electrophysiological investigation in the visual system of the fly. *Biological Cybernetics, 56,* 69–87.

Emerson, R. C. (1988). A linear model for symmetric receptive fields: Implications for classification tests with flashed and moving images. *Spatial Vision, 3,* 159–177.

Emerson, R. C. & Bergen, J. R. (1989a). Directionally selective neurons calculate motion energy in cat visual cortex. *Investigative Ophthalmology and Visual Science, Supplement, 30,* 425.

Emerson, R. C. & Bergen, J. R. (1989b). Nonlinear analysis of motion energy calculations in cat visual cortex. In Buus, S. (Ed.), *Proceedings of the 15th Northeast Bioengineering Conference; IEEE Cat. No. 89-CH2689-8* (pp. 143–144).

Emerson, R. C. & Citron, M. C. (1988). How linear and nonlinear mechanisms contribute to directional selectivity in simple cells of cat striate cortex. *Investigative Ophthalmology and Visual Science, Supplement, 29,* 23.

Emerson, R. C. & Citron, M. C. (1989). Linear and nonlinear mechanisms of motion selectivity in single neurons of the cat's visual cortex. In Kleinman, D. L. (Ed.), *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics; IEEE Cat. No. 89-CH2809-2* (pp. 448–453).

Emerson, R. C. & Citron, M. C. (1992). Linear and nonlinear mechanisms of motion selectivity in simple cells of the cat's striate cortex. In Pinter, R. B. & Nabet, B. (Eds), *Nonlinear vision: Mathematical analyses and a synthesis.* Boca Raton, Fla: CRC Press.

Emerson, R. C. & Coleman, L. (1981). Does image movement have a special nature for neurons in the cat's striate cortex? *Investigative Ophthalmology and Visual Science, 20,* 766–783.

Emerson, R. C. & Gerstein, G. L. (1977a). Simple striate neurons in the cat. I. Comparison of responses to moving and stationary stimuli. *Journal of Neurophysiology, 40,* 119–135.

Emerson, R. C. & Gerstein, G. L. (1977b). Simple striate neurons in the cat. II. Mechanisms underlying directional asymmetry and directional selectivity. *Journal of Neurophysiology, 40,* 136–155.

Emerson, R. C., Bergen, J. R. & Adelson, E. H. (1987a). Movement models and directionally selective neurons in the cat's visual cortex. *Society Neuroscience Abstracts, 13,* 1623.

Emerson, R. C. & Vaughn, W. J. (1987). Nonlinear spatiotemporal analysis of directionally selective cortical neurons. In Marmarelis, V. Z. (Ed.), *Advanced methods of physiological system modeling* (pp. 172–204). Los Angeles: USC Biomedical Simulations Resource.

Emerson, R. C., Korenberg, M. J. & Citron, M. C. (1988). Measurement of a simple-cell threshold function in cat's striate cortex. *Society of Neuroscience Abstracts, 14,* 899.

Emerson, R. C., Korenberg, M. J. & Citron, M. C. (1989). Identification of intensive nonlinearities in cascade models of visual cortex and its relation to cell classification. In Marmarelis, V. Z. (Ed.), *Advanced methods of physiological system modeling* (pp. 97–111). New York: Plenum Press.

Emerson, R. C., Korenberg, M. J. & Citron, M. C. (1990a). Cortical computation of motion energy: location of the squarer. In Pedersen, P. D. & Onaral, B. (Eds), *Proceedings of the 12th Annual IEEE Conference on Engineering in Medicine and Biology; IEEE Cat. No. 90-CH2936-3* (pp. 1882–1883).

Emerson, R. C., Korenberg, M. J. & Citron, M. C. (1990b). Squaring transformations in complex cells of cat's striate cortex. *Investigative Ophthalmology and Visual Science, Supplement 31,* 397.

Emerson, R. C., Korenberg, M. J. & Citron, M. C. (1992). Identification of complex-cell intensive nonlinearities in a cascade model of cat visual cortex. *Biological Cybernetics.* In press.

Emerson, R. C., Citron, M. C., Vaughn, W. J. & Klein, S. A. (1987b). Nonlinear directionally selective subunits in complex cells of cat striate cortex. *Journal of Neurophysiology, 58,* 33–65.

Fahle, M. & Poggio, T. (1981). Visual hyperacuity: spatio-temporal interpolation in human vision. *Proceedings of the Royal Society of London B, 213,* 451–477.

Field, D. J. & Tolhurst, D. J. (1986). The structure and symmetry of simple-cell receptive-field profiles in the cat's visual cortex. *Proceedings of the Royal Society of London B, 228,* 379–400.

Hamilton, D. B., Albrecht, D. G. & Geisler, W. S. (1989). Visual cortical receptive fields in monkey and cat: spatial and temporal phase transfer function. *Vision Research, 29,* 1285–1308.

Heeger, D. J. (1987). Model for the extraction of image flow. *Journal of the Optical Society of America A, 4,* 1455–1471.

Heeger, D. J. (1991). Nonlinear model of neural responses in cat visual cortex. In Landy, M. S. & Movshon, J. A. (Eds), *Computational models of visual processing* (pp. 119–133). Cambridge, Mass.: MIT Press.

Henry, G. H. (1977). Receptive field classes of cells in the striate cortex of the cat. *Brain Research, 133,* 1–28.

Henry, G. H., Lund, J. S. & Harvey, A. R. (1978). Cells of the striate cortex projecting to the Clare-Bishop area of the cat. *Brain Research, 151,* 154–158.

Hubel, D. H. & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology, London, 148,* 574–591.

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

Humphrey, A. L. & Weller, R. E. (1988a). Functionally distinct groups of X-cells in the lateral geniculate nucleus of the cat. *Journal of Comparative Neurology, 268,* 429–447.

Humphrey, A. L. & Weller, R. E. (1988b). Structural correlates of functionally distinct X-cells in the lateral geniculate nucleus of the cat. *Journal of Comparative Neurology, 268,* 448–468.

Jacobson, L. D., Gaska, J. P., Chen, H. & Pollen, D. A. (1989). 2D spatial/1D temporal white noise stimulus-response cross correlation studies in V1 of the macaque monkey. *Investigative Ophthalmology and Visual Science, Supplement, 30,* 298.

Larkin, R. M., Klein, S., Ogden, T. E. & Fender, D. H. (1979). Nonlinear kernels of the human ERG. *Biological Cybernetics, 35,* 145–160.

Mancini, M., Madden, B. C. & Emerson, R. C. (1990). White noise analysis of temporal properties in simple receptive fields of cat cortex. *Biological Cybernetics, 63,* 209–219.

Marmarelis, P. Z. & Marmarelis, V. Z. (1978). *Analysis of physiological systems: The white-noise approach.* New York: Plenum Press.

Mastronarde, D. N. (1987a). Two classes of single-input X-cells in cat lateral geniculate nucleus. I. Receptive-field properties and the classification of cells. *Journal of Neurophysiology, 57,* 357–380.

Mastronarde, D. N. (1987b). Two classes of single-input X-cells in cat lateral geniculate nucleus. II. Retinal inputs and the generation of receptive-field properties. *Journal of Neurophysiology, 57,* 381–413.

Maunsell, J. H. R. & Van Essen, D. C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *Journal of Neurophysiology, 49,* 1127–1147.

McKee, S. P., Silverman, G. H. & Nakayama, K. (1986). Precise velocity discrimination despite random variations in temporal frequency and contrast. *Vision Research, 26,* 609–619.

McLean, J. & Palmer, L. A. (1988). Contribution of linear mechanisms to direction selectivity of simple cells in area 17 and 18 of the cat. *Investigative Ophthalmology and Visual Science, Supplement, 29,* 23.

McLean, J. & Palmer, L. A. (1989). Contribution of linear spatiotemporal receptive field structure to velocity selectivity of simple cells in area 17 of cat. *Vision Research, 29,* 675–679.

Movshon, J. A., Thompson, I. D. & Tolhurst, D. J. (1978). Receptive field organization of complex cells in the cat's striate cortex. *Journal of Physiology, London, 283,* 79–99.

Poggio, T. & Reichardt, W. (1976). Visual control of orientation behaviour in the fly. Part II. Towards the underlying neural interactions. *Quarterly Reviews of Biophysics, 9,* 377–438.

Pollen, D. A., Gaska, J. P. & Jacobson, L. D. (1989). Physiological constraints on models of visual cortical function. In Cotterill, R. M. J. (Ed.), *Models of brain function* (pp. 115–135). Cambridge, U.K.: Cambridge University Press.

Pollen, D. A. & Ronner, S. F. (1981). Phase relationships between adjacent simple cells in the visual cortex. *Science, New York, 212*, 1409–1411.

Reichardt, W. (1957). Autokorrelationsauswertung als Functionsprinzip des Zentralnervensystems. *Z. Naturforsch. 12b*, 448–457.

Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In Rosenblith, W. A. (Ed.), *Sensory communication* (pp. 303–317). New York: Wiley.

Reid, R. C., Soodak, R. E. & Shapley, R. M. (1987). Linear mechanisms of directional selectivity in simple cells of cat striate cortex. *Proceedings of the National Academy of Sciences U.S.A., 84*, 8740–8744.

Rodieck, R. W. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research, 5*, 583–601.

van Santen, J. P. H. & Sperling, G. (1984). Temporal covariance model of human motion perception. *Journal of the Optical Society of America A, 1*, 451–473.

van Santen, J. P. H. & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A, 2*, 300–321.

Saul, A. B. & Humphrey, A. L. (1990). Spatial and temporal response properties of lagged and nonlagged cells in cat lateral geniculate nucleus. *Journal of Neurophysiology, 64*, 206–224.

Watson, A. B. & Ahumada, A. J., Jr (1983). A look at motion in the frequency domain. *NASA Technical Paper 84352*, 1–10.

Watson, A. B. & Ahumada, A. J., Jr (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A, 2*, 322–342.

Acknowledgements—The authors wish to thank M. C. Citron for help in acquiring physiological data, S. A. Klein for deriving kernel formulations, W. J. Vaughn for physiological kernel calculations, and C. M. Cappiello for preparing figures. We would also like to acknowledge helpful comments on the manuscript from M. Egelhaaf and A. Borst. This work was supported by EY06679 to RCE, and Core Grant EY01319 to the Center for Visual Science. EHA was supported in part by NSF No. IRI 871-939-4.

APPENDIX A

The 2-bar Interaction of a Quadratic System is a Spatiotemporal Autocorrelation Function

In this section we prove a general result concerning the structure of the second-order kernel of a second-order mechanism, one whose output is the sum of the squares of the outputs of a family of linear subunits. Let R_i be a set of n linear subunits with impulse responses $L_i(s, t)$; that is, for an input $I(s, t)$:

$$R_i(t) = \int_0^\infty \int_{-\infty}^\infty L_i(s, \tau) I(s, t - \tau)\, ds\, d\tau.$$

Let

$$E(t) = \sum_{i=0}^{n} [R_i(t)]^2$$

be the output of the mechanism. For a system of this kind, the second-order Wiener kernel and the second-order Volterra kernel are both equal to the second-order superposition residual, K_2 (Marmarellis & Marmarellis, 1978, pp. 148–153; Schetzen, 1980, pp. 38–45). If we let $E_1(t; s_0, \tau_0)$ be the output of the mechanism in response to an impulse input at position s_0 and time τ_0, and $E_2(t; s_0, \tau_0, s_1, \tau_1)$ be the response of the mechanism to a pair of impulses occurring at (s_0, τ_0) and (s_1, τ_1), then the second-order residual is defined as

$$K_2(\Delta s, \Delta\tau) = \int_{-\infty}^\infty \int_{-\infty}^\infty E_2(0; s, \tau, s + \Delta s, \tau + \Delta\tau) - E_1(0; s, \tau) - E_1(0; s + \Delta s, \tau + \Delta\tau)\, ds\, d\tau.$$

In this section we prove that K_2 is just the sum of the spatiotemporal autocorrelation functions of the linear subunits. Formally

$$K_2(\Delta s, \Delta\tau) = 2 \sum_{i=0}^{n} A_i(\Delta s, \Delta\tau)$$

where

$$A_i(\Delta s, \Delta\tau) = \int_{-\infty}^\infty \int_{-\infty}^\infty L_i(s, \tau) L_i(s + \Delta s, \tau + \Delta\tau)\, ds\, d\tau$$

Proof:

Let $R_i(t; s_0, \tau_0)$ be the response of subunit i to an impulse input at position s_0 and time τ_0. Then

$$R_i(t; s_0, \tau_0) = L_i(s_0, t - \tau_0)$$

and

$$E_1(t; s_0, \tau_0) = \sum_{i=0}^{n} [L_i(s_0, t - \tau_0)]^2.$$

Let $R_i(t; s_0, \tau_0, s_1, \tau_1)$ be the response of subunit i to a pair of impulses occurring at (s_0, τ_0) and (s_1, τ_1). By superposition:

$$R_i(t; s_0, \tau_0, s_1, \tau_1) = L_i(s_0, t - \tau_0) + L_i(s_1, t - \tau_1)$$

and

$$E_2(t; s_0, \tau_0, s_1, \tau_1) = \sum_{i=0}^{n} [L_i(s_0, t - \tau_0) + L_i(s_1, t - \tau_1))]^2.$$

From these definitions, it follows that:

$$\begin{aligned} K_2(\Delta s, \Delta\tau) &= \int_{-\infty}^\infty \int_{-\infty}^\infty E_2(0; s, \tau, s + \Delta s, \tau + \Delta\tau) - \\ &\quad E_1(0; s, \tau) - E_1(0; s, + \Delta s, \tau + \Delta\tau)\, ds\, d\tau \\ &= \int_{-\infty}^\infty \int_{-\infty}^\infty \sum_{i=0}^{n} \{[L_i(s, -\tau) + L_i(s + \Delta s, -(\tau + \Delta\tau))]^2 \\ &\quad [L_i(s, -\tau)]^2 - [L_i(s + \Delta s, -(\tau + \Delta\tau))]^2\}\, ds\, d\tau \\ &= \int_{-\infty}^\infty \int_{-\infty}^\infty \sum_{i=0}^{n} [2 L_i(s, -\tau) L_i(s + \Delta s, -(\tau + \Delta\tau))]\, ds\, d\tau \\ &\quad 2 \sum_{i=0}^{n} \int_{-\infty}^\infty \int_{-\infty}^\infty L_i(s, \tau) L_i(s + \Delta s, \tau + \Delta\tau)\, ds\, d\tau \\ &= 2 \sum_{i=0}^{n} A_i(\Delta s, \Delta\tau). \end{aligned}$$

This result is independent of the nature of the functions L_i. However, as a special case it specifies that if the functions L_i consist of a pair of filters in quadrature phase, or are a set of shifted versions of the same function, then the second-order kernel is just the autocorrelation function of this filter. Note that this result also implies that even if the input filters do not have precisely equal power spectra, for example in the case of odd- and even-phase gabor functions, then the second-order kernel is just the average of the various autocorrelation functions. Furthermore, since the second-order kernel depends only on the autocorrelation functions of the linear subunits and not on their spatial distribution, it contains no information about such factors as number, extent of spatial pooling, or degree of overlap of the subunits.

APPENDIX B

Computational Comparison Between Reichardt and Energy Models

Figure 7A shows how easily a neuron could effect the squaring operation needed for calculating motion energy. But the nonlinear transformation in the Reichardt model is also of second degree, and appears to be closely related. One could argue that if a squaring operation is available then so is the product function of the Reichardt model because of the cross product generated by squaring a sum: $(A' + B)^2 = A'^2 + B^2 + 2A'B$. The converse may not be true, however, because of the missing squared terms (see below).

In Fig. 7B is shown schematically an opponent combination of two basic energy units, such as in Fig. 7A, but connected in push–pull, the left unit preferring downward motion. To show that the basic structure of this energy model is not only physiologically realizable, but is also composed of general enough building-blocks to accommodate other configurations, we have expressed in Fig. 7C the entire opponent Reichardt model as a bilateral reconnection of only two paths between the pair of unidirectional energy units in B. A single exchange between outputs of the two inner $(\ )^2$ boxes in B, and the resulting inversion in weighting of the new connections necessary to retain the same output signal from the opponent combination are the only requirements to produce not only the fully opponent output signal (Fig. 3E), but also the pure cross-product terms, $4A'B$ and $4AB'$, that provide the signature of the Reichardt half-detector Fig. 3D.

This observation underscores the seemingly trivial difference in connectivity between the two models, but at the same time, illustrates the importance of the pure squared terms, $A^2 + A'^2 + B^2 + B'^2$, shown at the right side of Fig. 7B for the unidirectional energy model. These terms produce, spatially, an approximately Gaussian shaped "pillow" on which ride the opponent 2-bar interactions of Fig. 3D,E (right). This positive pillow cancels the otherwise purely negative interactions for the null direction (that would be integrated over a trajectory with a negative slope in an opponent model (Fig. 3E right) for downward movement), which creates the unidirectional property of the energy model. This property may be critical for the successful performance of the visual system, as suggested by the success of Heeger's (1987) spatially 2D model for extracting image flow. It was built from unidirectional energy subunits, instead of depending on the opponent-energy signal available from a Reichardt computation.

Note that although the Reichardt detector is not a good model for DS complex cells in the striate cortex of the cat, its appropriateness in other systems, remains an open question. For example, Maunsell and Van Essen (1983) have found that some DS neurons in MT of the monkey show inhibition of the background firing rate for motion in the null direction. Further, Andersen, Snowden, Treue and Graziano (1990) have found mild suppression of preferred-direction responses by null-direction motion that occurs during motion of a transparent cylinder. This mild opponency in MT may be multiplicative rather than subtractive, but is likely to be generated by a combination of separate motion-energy signals originating in V1, because DS cells in V1 of both monkey and cat are known to be unidirectional (Anderson *et al.*, 1990; Emerson *et al.*, 1987b), and monkey MT receives strong input from V1.