

# Neural Basis for Stereopsis from Second-Order Contrast Cues

Hiroki Tanaka and Izumi Ohzawa

Graduate School of Frontier Biosciences and School of Engineering Science, Osaka University, Osaka 560-8531, Japan

Humans and animals use visual cues such as brightness and color boundaries to identify objects and navigate through environments. However, even when these cues are not available, we can effortlessly perform these tasks by using second-order cues such as contrast variation (envelope) of patterns on surfaces. Previously, numerous psychophysical studies examined properties of binocular depth processing based on the contrast-envelope cues and suggested the existence of a stereo system that uses these cues. However, its physiological substrate has not been identified yet. Here, we show that a subset of cortical neurons in cat area 18 show binocular interactions for the contrast-envelope stimuli. These neurons are capable of representing a variety of depths in the three-dimensional space based on the information available from contrast cues alone. Furthermore, these neurons show similar disparity-tuning curves for borders defined by both luminance and contrast cues. This cue-invariant tuning is consistent with a linear binocular convergence model for monocular luminance and contrast-envelope processing pathways.

**Key words:** stereopsis; contrast cues; binocular processing; second-order stimuli; cue invariance; cat area 18; early visual cortex

## Introduction

Borders between areas with different luminances and colors (first-order cues) are key features that our visual system uses for defining shape and identifying objects. However, even when these obvious cues are absent, the visual system can perform the same tasks by using borders defined in the contrast envelope (second-order cues), as demonstrated by two cycles of a clearly visible grating defined by local contrast variations (see Fig. 1, top right) (Chubb and Sperling, 1988; Cavanagh and Mather, 1989).

Previously, a substantial number of psychophysical studies investigated binocular processing of contrast-envelope cues (Hess and Wilcox, 1994; Wilcox and Hess, 1995, 1996, 1997; Schor et al., 1998; Edwards et al., 1999, 2000; Langley et al., 1999; McKee et al., 2004). These studies have shown that we are able to perceive depth based purely on binocular disparities of contrast-envelope cues and suggested that such stereopsis is mediated by nonlinear mechanisms that are distinct from that for extracting luminance-based cues. However, there has not been a physiological investigation that has examined responses of cortical neurons to binocular contrast-envelope stimuli with systematically controlled binocular disparities. In this study, therefore, we address the questions on physiological mechanisms of binocular process-

ing of contrast-defined features and their relationships to the luminance-based binocular mechanisms. Specifically, are there cortical neurons capable of signaling depth when visual features are defined in contrast cues? And, if so, do these neurons signal the same depth in contrast cues as that defined by luminance cues? To address these questions, we examined binocular disparity-tuning properties of neurons in the cat early visual cortex using binocular luminance and contrast-envelope stimuli. It is known that there are cortical neurons selective to spatial frequency or motion of contrast-envelope stimuli (Zhou and Baker, 1993). Such neurons are more frequently found in area 18 than in area 17.

We also want to know what form of neural circuitry implements disparity information processing based on luminance and contrast-envelope cues. A generally accepted monocular model of contrast-envelope-sensitive neurons in the early visual cortex is illustrated in Figure 1 (Zhou and Baker, 1993). These neurons appear to possess a hybrid mechanism consisting of a linear receptive field (RF) (see Fig. 1, left) and a nonlinear pathway (right) in which output of multiple first-stage filters (tuned to fine carrier features) is rectified and integrated by the second-stage RF (tuned to contrast-defined features). Essentially the same model seems to be widely accepted in psychophysical studies (Smith, 1994). How can this model be extended to the binocular system? How and where in the visual system do the luminance and contrast processing pathways converge binocularly? Because there are many possible combinations of connections involving the two eyes, two cue pathways, and two stages of neurons for the contrast-envelope processing pathway, identifying a single true model is not a trivial exercise. Here, we will address these questions physiologically by examining binocular disparity-tuning curves of multiple neurons, using various binocular combinations of stimuli similar to those in Figure 1.

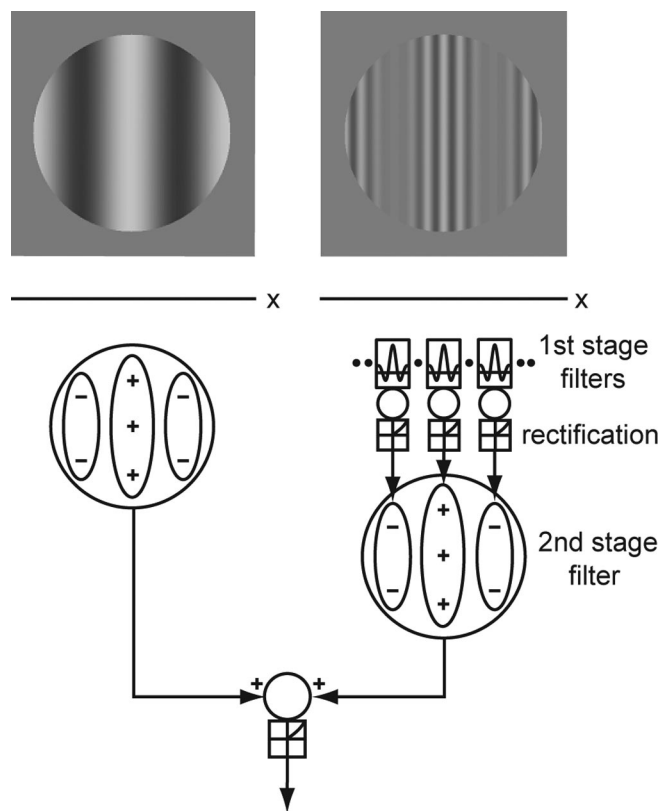
Received Aug. 3, 2005; revised March 9, 2006; accepted March 10, 2006.

This work was supported by Grants 15029230 and 15700258 and by the Project on Neuroinformatics Research in Vision through special coordination funds for promoting science and technology from the Ministry of Education, Culture, Sports, Science, and Technology; and by Grant 13308048, the 21st Century Centers of Excellence Program, and a France–Japan Joint Research Program Grant from the Japan Society for the Promotion of Science. We thank our laboratory members (Shinji Nishimoto, Takahisa Sanada, Rui Kimura, Kota Sasaki, Masayuki Fukui, Miki Arai, Masashi Iida, Tsugitaka Ishida, and Taihei Ninomiya) for their help in experiments and valuable discussions.

Correspondence should be addressed to Dr. Izumi Ohzawa, Graduate School of Frontier Biosciences and School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan. E-mail: ohzawa@fbs.osaka-u.ac.jp.

DOI:10.1523/JNEUROSCI.4379-05.2006

Copyright © 2006 Society for Neuroscience 0270-6474/06/264370-13\$15.00/0



**Figure 1.** A parallel stream model of cat area 18 cells for processing luminance and contrast-envelope stimuli. Cat area 18 cells show similar orientation and spatial frequency selectivity to luminance stimuli (left) and contrast-envelope stimuli (right). Luminance stimuli are thought to be processed through a conventional linear RF (left). The contrast-envelope stimuli are processed with filter-rectify-filter circuits (right), in which fine texture elements (carrier) are extracted at first-stage filters and second-stage filters sum the rectified outputs of the first-stage filters to extract the contrast envelope.

## Materials and Methods

All animal care and experimental guidelines conformed to those established by the National Institutes of Health (Bethesda, MD) and were approved by the Osaka University Animal Care and Use Committee.

**Surgery and apparatus.** Twenty-one normal adult cats (2–4 kg) were prepared for single-unit recording using standard procedures. Details of the surgical procedures were described previously (Nishimoto et al., 2005). Briefly, initial surgical anesthesia was induced and maintained by isoflurane (2–3.5% in oxygen). After insertion of catheters in at least two veins and insertion of a glass tracheal tube by tracheostomy, the animal was secured in a stereotaxic apparatus. A hole (typically 5–7 mm in diameter) was made over the representations of area 18. Then, paralysis was induced by a loading dose of gallamine triethiodide (Flaxedil; 10–20 mg), and the animal was placed under artificial respiration at the rate of 20–30 strokes/min. Paralysis was maintained by continuous infusion of gallamine triethiodide ( $10 \text{ mg} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$ ). Anesthesia for the recording session was maintained by a combination of nitrous oxide (70%), oxygen, and sodium thiopental (Ravonal;  $1 \text{ mg} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$ ) in the infusion fluid. The infusion fluid also contained glucose ( $40 \text{ mg} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$  in Ringer's solution). Electrocardiogram, end-tidal  $\text{CO}_2$ , intra-tracheal pressure, heart rate, and rectal temperature were monitored continuously and maintained at a normal level throughout the experiments. The pupils were dilated with atropine sulfate (1%), and nictitating membranes were retracted with phenylephrine hydrochloride (Neosynesis; 5%). Contact lenses of appropriate power with 4 mm artificial pupils were placed over the corneas.

Lacquer-coated tungsten microelectrodes (1–5 M $\Omega$ ; A-M Systems, Sequim, WA) were used for extracellular recording from single cells in area 18 (Horsley–Clark A4 L4). Two electrodes mounted in a single protective

guide tube were driven in parallel with a single microelectrode drive to increase the chance of encountering neurons and simultaneous recordings of pairs of neurons. The signals from the electrodes were amplified, bandpass filtered (model 1800; A-M Systems), fed to a custom-made data acquisition system (Ohzawa et al., 1996) and an oscilloscope. The data acquisition system consisted of analog-to-digital converters and a spike sorter that sorted signals from each electrode into a maximum of five different classes in real time. The isolated spike data (time resolution, 40  $\mu\text{s}$ ) were sent from the data acquisition system, along with stimulus timing information, to a separate computer that controlled trials and performed preliminary on-line analysis. The data were saved to a file to allow off-line analysis of data.

For each electrode track, electrolytic lesions (5  $\mu\text{A}$ , 10 s) were made at 700–1500  $\mu\text{m}$  intervals while the electrodes were retracted. After all recording sessions, the animal was then given an overdose of pentobarbital sodium (Nembutal) and perfused through the heart with formalin (4% in buffered saline). Coronal sections (40–60  $\mu\text{m}$  thickness) of the visual cortex were made and stained with thionin. The locations of electrode tracks have been identified.

**Visual stimuli.** Stimuli were produced by a Windows-based personal computer controlling a graphics card (Millennium G550; Matrox, Dorval, Quebec, Canada) and displayed on a color cathode ray tube monitor (76 Hz,  $1600 \times 1024$  pixels, mean luminance of 40  $\text{cd}/\text{m}^2$ ; GDM-FW900; Sony, Tokyo, Japan) using only the green channel, placed 57 cm away from the cat's eye. Stimuli were presented dichoptically through front-surface mirrors angled at 45° in front of the animal's eye. In each experiment, the luminance nonlinearity of the display was measured using a photometer (Minolta CS-100; Konica Minolta Photo Imaging, Mahwah, NJ) and linearized by gamma-corrected lookup tables.

We used circular patches of luminance and contrast-envelope gratings as visual stimuli. A spatiotemporal luminance profile of the one-dimensional drifting luminance grating at a point  $x$  and time  $t$  is defined by the following:

$$L_{\text{lum}}(x,t) = L_{\text{mean}}(1 + C \sin(2\pi f_x x - 2\pi w t)),$$

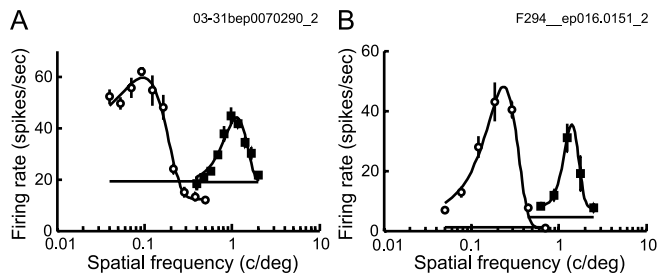
where  $L_{\text{mean}}$  is mean luminance,  $C$  is contrast,  $f_x$  is spatial frequency, and  $w$  is temporal frequency. The contrast was 50%. Contrast-envelope stimuli were composed of a high spatial frequency luminance grating (carrier) with its contrast modulated by a low spatial frequency sine wave grating (envelope) (Zhou and Baker, 1993).

$$L_{\text{env}}(x,t) = L_{\text{mean}}(1 + C \sin(2\pi f_c x - 2\pi f_{tc} t) (1 + m \sin(2\pi f_e x - 2\pi f_{te} t)) / 2),$$

where  $f_c$ ,  $f_e$ ,  $f_{tc}$ , and  $f_{te}$  are spatial frequency of the carrier, spatial frequency of the envelope, temporal frequency of the carrier, and temporal frequency of the envelope, respectively. The contrast of the carrier,  $C$ , was 50%, and the modulation of the envelope,  $m$ , was 100%. The envelope was always drifted at 2 Hz. The carrier was drifted at 0 Hz (stationary) or 2 Hz. These variations are attributable to attempts to increase responses to contrast-envelope stimuli as much as possible. The mean luminance of luminance and contrast-envelope stimuli was the same as that of the background of the display.

**Procedures.** Once one or more neurons were isolated, optimal monocular parameters of the luminance grating were first determined. The orientation and spatial frequency of the envelope for the contrast-envelope stimuli were set to be the same as those for the luminance grating. The carrier spatial frequency was varied and set to an optimal value for activating neurons. If neurons did not seem to respond to contrast-envelope stimuli clearly, we generally terminated measurements with the contrast-envelope stimuli and used the neurons for other studies or moved on to other cells by advancing the electrodes.

If neurons responded to contrast-envelope stimuli, we generally found that neurons show bandpass tunings for the carrier (average bandwidth at half peak height, 1.27 octaves;  $n = 26$ ). Optimal values for the carrier [average peak carrier frequency, 1.16 cycles/degree ( $\text{c}/\text{deg}$ );  $n = 26$ ] were in such a high spatial frequency range that contrast-envelope stimuli contained no luminance energy within the spatial frequency pass-band of neurons for the luminance stimuli, consistent with the previous studies (Zhou and Baker, 1993). Examples of two representative cells are



**Figure 2.** Spatial frequency tunings of two neurons for luminance stimuli and the carrier of contrast-envelope stimuli. Open circles and solid squares indicate the mean discharge rates when the spatial frequency of luminance stimuli and the carrier spatial frequency of the contrast-envelope stimuli were varied, respectively. The spatial frequency of the envelope of the contrast-envelope stimuli was fixed near the optimal values for the luminance stimuli (0.08 c/deg for the neuron in **A**, 0.24 c/deg for the neuron in **B**). Solid curves indicate Gaussian functions fitted to the data points. The optimal spatial frequency and bandwidth at half-height were obtained from the fitted curves. Note that these neurons are narrowly tuned to the carrier spatial frequency. Carrier bandwidths of the neurons in **A** and **B** are 1.10 and 0.66 octaves, respectively. Note also that the response range for the two stimuli did not overlap (we confirmed, during initial manual searches, that luminance stimuli with spatial frequencies above the highest one tested in computer-controlled measurements did not evoke responses). Error bars indicate the SEM. Horizontal lines below the tuning curves indicate the spontaneous discharge rates. Because the two tuning curves were measured separately, the spontaneous discharge rates are not necessarily the same.

shown in Figure 2. We then set the carrier frequency at these optimal values. The average carrier and envelope frequencies of the contrast-envelope stimuli were 0.97 c/deg (SD, 0.36;  $n = 70$ ) and 0.11 c/deg (SD, 0.06;  $n = 70$ ), respectively. The average ratio of carrier/envelope frequencies was 11.0 (SD, 7.6;  $n = 70$ ). On average, the carrier frequency was four times higher than that of the high-cut frequency of neurons for luminance stimuli (high-cut frequency at half peak height, 0.23 c/deg; SD, 0.11;  $n = 59$ ). These values justify the assumption that contrast-envelope stimuli cannot be processed by a linear pathway and require different nonlinear processing pathways. Note also that clear sharp tunings for the carrier spatial frequency rule out the possibility that envelope responses are attributable to point-wise nonlinearities, including those in the display monitor and saturations in early neural stages. These nonlinearities can generate distortion products at the frequency of the envelope. However, because such nonlinearities lack the first-stage filter, there should not be any sharp bandpass tuning for carrier spatial frequency (Zhou and Baker, 1993). This expectation was clearly not the case. The carrier orientation was set at the optimal values of the neurons, although in many cases, neurons were only broadly tuned to the carrier orientation, if at all. The size of grating patches was adjusted to just cover the RFs of the neurons.

We then measured sensitivity to binocular disparity for each of the envelope gratings, luminance gratings, and the cross-cue stimuli, in which the left and right eyes receive different types of stimuli, in a randomly interleaved manner in a single run. For each type of binocular stimuli, binocular disparities (interocular phases) were varied in 30° phase steps over one cycle. For the contrast-envelope stimuli, binocular disparity of the envelope was varied, whereas the carrier phase was fixed at 0 with respect to the center of the stimulus patches. This run also included tests in which the left- or right-eye stimulus was presented alone monocularly and tests for null stimuli (blank stimuli with a luminance that is the same as the average luminance of grating stimuli). The gratings were drifted in the optimal direction of the neuron.

For a small number of neurons, we tested sensitivity to the interocular carrier phase in a separate run. The interocular phase of the carrier was varied in 30° steps over one cycle, whereas that of the envelope was fixed at the optimal value of the neuron. Orientation, spatial frequency, and temporal frequency of the contrast-envelope stimuli were the same as those used in the above main run. Twelve contrast-envelope stimuli and blank stimuli were interleaved. For all runs using luminance and contrast-envelope stimuli, each stimulus was presented for 4 s and repeated typically five times.

**Data analysis.** If responses of a neuron to at least 1 of 12 contrast-envelope stimuli with different interocular phases of the envelope were significantly different from the spontaneous activity level (paired  $t$  test;  $p < 0.05$ ) and were  $> 1$  spike/s, the neuron was judged as being responsive to contrast-envelope stimuli and included in data analysis.

Modulation depth of the disparity (interocular phase) tuning curves is defined as the ratio of the amplitude of F1 to that of the F0 component of the tuning curves. The F1 component is the amplitude of a cycle of sine function fitted to the disparity-tuning curve, whereas F0 is the mean discharge rate to binocular stimulation. The higher the index, the greater the modulation of the responses is as a function of the interocular phase. An index near 0 indicates that the responses are hardly modulated. The index may exceed 1 depending on the shape of the disparity-tuning curve (e.g., if the tuning curve is clipped).

To establish whether the modulation is statistically significant or not, we analyzed the goodness of the fit as follows. Because the binocular stimuli are periodic with a period of 360°, one cycle of sinusoids were fitted for the interocular phase tuning curves using a least-squares criterion. Note that we have never observed cases in which interocular phase tuning curves exhibit more than one cycle of response variations, although it is a theoretical possibility if the system, for example, responds at a doubled temporal frequency (Freeman and Ohzawa, 1988). Next, the mean square of fitted curves about the mean value and the mean square of residuals between the fitted curves and the tuning curves were calculated. We then obtained the ratio of the two values. If this value was larger than the 5% criterion for the  $F$  test [i.e., 95% point of  $F$  distribution with  $m$  and  $n-m-1$  degrees of freedom ( $n$ , number of data (typically 60);  $m = 2$ : number of parameters (amplitude, phase)], we judged the fitting to be acceptable. Although this test is used for linear regression, we confirmed that most of the neurons determined to be significant by this test were also statistically significant for one-way ANOVA ( $p < 0.05$ ; 19 of 25 neurons). Optimal phases are peaks of the fitted sinusoidal curves.

For evaluating cue invariance, it is necessary to compare two phase tuning curves. Normally, we fitted two sinusoids independently to two disparity-tuning curves to calculate the difference of optimal phases between the two curves. However, we also simultaneously fitted two sinusoids to the two tuning curves with their phases constrained to be the same. By comparing these two fits, with independent and common phase values, we are able to assess the statistical significance of the phase difference, thereby producing a metric for cue invariance. Residual variances around these two fits were compared by a sequential  $F$  test to determine whether the phase produces a significance improvement (Draper and Smith, 1998; Thomas et al., 2002). The level of significance was 0.05.

**Simulations.** Computer simulations for binocular versions of the “filter-rectify-filter” model (Fig. 1, right) were conducted using programs written in MATLAB (Mathworks, Natick, MA). Although the model also contains a pathway for the luminance contrast signal (Fig. 1, left), this pathway does not respond to the contrast-envelope stimuli we have used. Therefore, only the contrast-envelope pathway needed to be simulated. Stimuli were one-dimensional and defined using 320 pixels. The spatial resolution of the simulation was such that 10 pixels correspond to 1° of visual angle. Therefore, the stimulus width corresponds to 32°. For contrast-envelope stimuli used in the simulation, the envelope spatial frequency was 0.1 c/deg (period, 100 pixels), and the carrier spatial frequency was 1.0 c/deg (period, 10 pixels).

Both the first- and second-stage filters were modeled as one-dimensional Gabor functions, which are defined by the following:

$$f(x) = A \exp(- (x - \mu)^2 / 2\sigma^2) \cos(2\pi f(x - \mu) + \rho),$$

where  $A$ ,  $\mu$ ,  $\sigma$ ,  $f$ , and  $\rho$  are the amplitude, center position, width (SD), spatial frequency, and phase, respectively. Each location along the  $x$ -axis contains the first-stage filters (i.e., Gabor functions of 0, 90, 180, and 270° phases), although only a filter with zero phase is depicted in Fig. 1 for clarity. The spatial frequency was 1.0 c/deg for the first-stage filter and 0.1 c/deg for the second-stage filter, which are near the average optimal carrier and envelope spatial frequencies of area 18 neurons, respectively. The width was 0.444° for the first-stage filters and 4.44° for the second-stage filters so that they had 1.3 octave bandwidths.

Each first-stage neuron derives its output by half-wave rectifying and squaring (“half-squaring” in short) the signal from the first-stage filter. The half-wave rectification represents the fact that the spike discharge rate cannot signal negative values without spontaneous discharge, which is minimal for early visual cortical neurons. The sum of signals from the two quadrature pairs of the first-stage neurons (i.e., neurons with Gabor filters of 0, 90, 180, and 270° phases) provides a signal proportional to stimulus energy at each location (Adelson and Bergen, 1985; Ohzawa et al., 1990).

The second-stage filter computes a weighted sum of outputs of the first-stage neurons. The output of this filter determines the contrast-envelope signal of the second-stage cell. Plus and minus signs drawn inside the second-stage filter indicate the positive and negative weights for the input from the first-stage cells, respectively. The second-stage filter signal is then half-wave rectified and squared.

To calculate responses to each drifting contrast-envelope stimulus, we calculated responses of the model neuron for progressively increasing envelope phases (in 6° steps). The F1 component (the amplitude of a sine function fitted to the responses for one cycle of phase change) was taken as the response amplitude for each stimulus. The carrier phase was not varied (stationary). To obtain a disparity-tuning curve, the envelope interocular phase was changed in 6° steps from 0 to 360°. The peak envelope interocular phase (peak disparity) is one at which the largest F1 response is observed.

## Results

### Cell population

We recorded from 151 neurons that responded to luminance or contrast-envelope stimuli according to our criteria. Of these, nearly all neurons ( $n = 148$ ) responded to luminance stimuli, whereas 70 responded to contrast-envelope stimuli. Forty-five percent of luminance-responsive neurons (67 of 148) responded to contrast-envelope stimuli. However, taking our sampling strategy into account (see Materials and Methods), we do not consider that this number represents the actual proportion of envelope-responsive neurons in this area. Our analysis is mostly confined to the 70 neurons that responded to the contrast-envelope stimuli.

Of these 70 neurons, 13 were classified as simple and 57 were classified as complex, applying standard harmonic analyses for responses to optimal contrast-envelope stimuli. We classified neurons as simple if their responses exhibited a high degree of temporal modulation at the stimulus temporal frequency ( $F1/F0 > 1$ ) (Skottun et al., 1991). For 38 neurons, only the envelope was drifted at 2 Hz; the carrier was stationary. For 32 neurons, both the carrier and envelope were drifted at the same temporal frequency (2 Hz) in an attempt to obtain more robust responses. However, the carrier motion is unlikely to have caused the F1 component of responses, because (1) the average  $F1/F0$  ratio was not different between the 32 cells (0.51; SD, 0.41) and the other 38 cells (0.57; SD, 0.51) for which stationary carrier was used, (2) the ratio of simple cells was not different for the two groups (5 of 32 and 8 of 38), and (3) when one simple neuron was tested using drifting contrast-envelope stimuli with both stationary and drifting carriers, the  $F1/F0$  ratio was invariant (1.27 for the stationary carrier vs 1.28 for the moving carrier).

### Binocular interaction for the contrast-envelope stimuli

Are cortical neurons capable of signaling stereoscopic depth defined in contrast cues? Specifically, are neurons tuned for binocular disparity defined solely in the contrast envelope? The question is addressed by dichoptically presenting contrast-envelope stimuli and shifting their interocular envelope phase over 0–360° (Ohzawa and Freeman, 1986a). The interocular carrier phase was kept constant.

Figure 3 shows representative results of interocular phase tuning measurements for the contrast-envelope stimuli in area 18. Figure 3, *A* and *B*, show responses of two neurons in the form of peristimulus time histograms (PSTHs). These two neurons were recorded simultaneously from different electrodes. Initial monocular tests (data not shown) found that each of these neurons responded to luminance and contrast-envelope stimuli of similar spatial frequency and orientation (Zhou and Baker, 1993; Marschall and Baker, 1998a). The spatial frequency of the carrier of the contrast-envelope stimuli (1.35 c/deg) was outside the luminance frequency pass-band for the neuron (high-cut frequency, 0.4 c/deg), so that they only stimulated the nonlinear processing pathway (Fig. 1, right). Both of these neurons were driven by each of left- and right-eye stimulation and had little spontaneous discharge (Fig. 3, top three rows). The bottom 12 rows show binocular responses for each interocular phase. Two important features are observed. First, for both neurons, responses were modulated by the interocular phase. In Figure 3*A*, strong responses were observed for phases  $\sim 150^\circ$ , whereas responses to opposite phases (near  $330^\circ$ ) were small. For the cell in Figure 3*B*, modulation by phase was weaker, but responses to  $60^\circ$  were clearly stronger than those to opposite phases. Second, these two neurons were tuned to different interocular phases. These points are better illustrated in Figure 3*C*, in which the mean discharge rates for these responses are plotted as a function of the interocular phase. Data for neurons shown in Figure 3, *A* and *B*, are indicated by open squares and filled circles, respectively. The peaks of these tuning curves are at markedly different phases. This difference indicates that they are tuned to different retinal disparities of contrast-envelope stimuli, although we do not know exactly what these disparities actually are in absolute terms because of the lack of accurate eye position information in paralyzed preparation (Ohzawa and Freeman, 1986a). It is unlikely that this peak difference reflects variations of temporal RF difference between eyes, because it is known that time courses of responses for the two eyes are quite similar for luminance stimuli (Ohzawa et al., 1996). Data for another pair of two simultaneously recorded neurons (also from different electrodes) are shown in Figure 3*D–F*. Binocular responses were clearly modulated by the interocular phase, and the optimal phases were very different, similar to those for the previous pair of neurons.

To examine the extent to which envelope-responsive neurons signal envelope-defined depth information, we quantified the depth of modulation by the interocular phase. For the 70 neurons that responded to the contrast-envelope stimuli, we calculated the modulation depth index of phase tuning curves for these stimuli (Fig. 3*G*). The higher the index, the more modulated the responses are as a function of phase. A substantial proportion of these neurons shows phase-specific binocular responses, having the index  $> 0.3$  (Ohzawa and Freeman, 1986a,b; Smith et al., 1997a). For 36% of our neurons ( $n = 25$  of 70) (Fig. 3*G*, filled portion of bars), the amplitude of the fitted sinusoid was significantly different from zero ( $F$  test,  $p < 0.05$ ; see Materials and Methods). Nine were classified as simple cells, whereas 16 were classified as complex cells. Note that the break between the statistically significant phase tunings (Fig. 3*G*, filled bars) and the nonsignificant phase tunings (open bars) was approximately at the modulation depth of 0.3, justifying the selection of the criterion in the previous studies. Most of the remaining neurons responded equally to the two eyes, but binocular interaction was not tuned to the interocular envelope phase. Such non-phase-specific neurons are also known to exist for luminance-defined stimuli in area 17 of the cat (Ohzawa and Freeman, 1986a,b) and

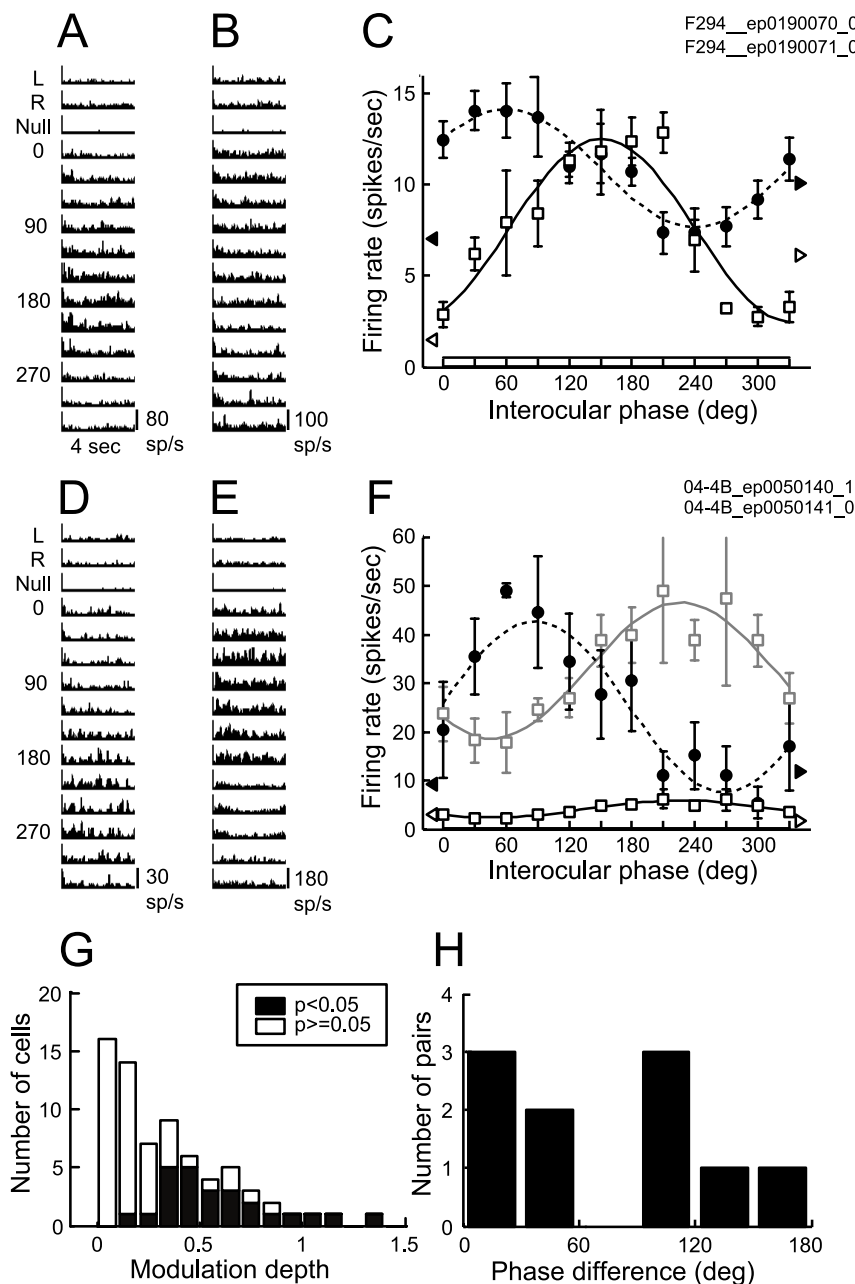
V1 of the monkey (Smith et al., 1997a). These results indicate that a substantial proportion of envelope-responsive neurons is capable of signaling depth based on binocular disparities of the contrast-envelopes in visual stimuli.

It is important that the two pairs of simultaneously recorded neurons are tuned to different disparities. This is because it makes it possible for representing a range of different disparities of the envelope, reflecting visual stimuli present in the three-dimensional structures of the external world. Because it is not possible to compare disparity-tuning curves recorded at different times in a paralyzed preparation [unless such methods as a reference cell technique in Hubel and Wiesel (1970) or Ferster (1981) are used], we have analyzed phase tuning curves from 10 pairs of simultaneously recorded neurons. For these neurons, tunings for envelope disparity were statistically significant. Six of 10 pairs are from two sets of three neurons that were recorded simultaneously. The remaining four pairs were obtained by simultaneous recordings from two neurons. Although a relatively small number of paired recordings were available (because of compounded difficulty of encountering pairs of neurons, both of which are binocular and envelope sensitive). Figure 3*H* clearly shows that a substantial fraction of neurons is tuned to different interocular phases. The result indicates that these neurons are tuned to a wide range of disparities of the envelope to allow representation of depth variations.

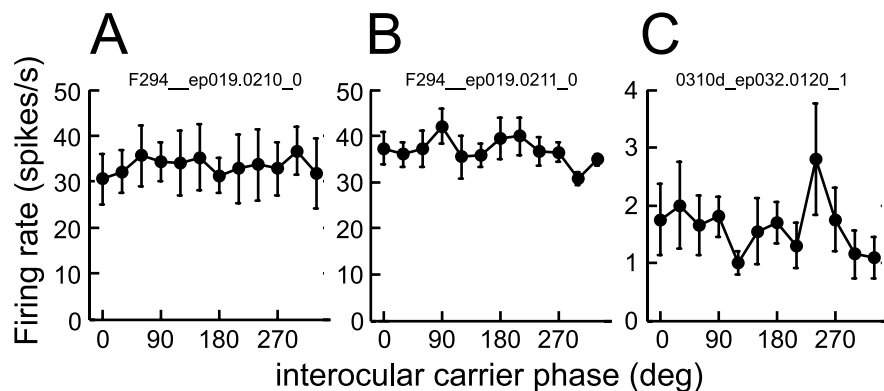
#### No sensitivity for the interocular carrier phase

In the experiments described so far, we had always set the carrier phases of the contrast-envelope stimuli to arbitrary but constant values, implicitly assuming that the carrier interocular phases do not affect the interocular phase tunings of the neurons for the envelope. To examine whether this assumption is correct or not, we recorded responses of envelope-responsive neurons while varying the carrier interocular phase in 30° steps over one cycle. The interocular phase of the envelope was fixed at the optimal value.

Results are shown in Figure 4. Only three disparity-selective neurons for contrast-envelope stimuli were tested for this experiment, but all of these neurons showed essentially no sensitivity for the carrier interocular phase. Modulation depths of the neurons in Figure 4*A–C* were 0.02, 0.05, and 0.06, respectively. Neurons in Figure 4, *A* and *B*, are the same



**Figure 3.** Interocular phase tunings of area 18 neurons for contrast-envelope stimuli. *A–C*, Data for a pair of neurons recorded simultaneously from different electrodes. *A*, *B*, PSTHs for monocular (left, first row; right, second row), null (third row), and interocular (bottom 12 rows) phase stimuli of the two neurons. The discharge rates and stimulus duration (4 s) are indicated at the bottom rows. *C*, Mean discharge rates are shown against the interocular phase, which is linearly related to binocular disparity of stimuli. Response amplitudes computed from PSTHs shown in *A* and *B* are indicated by open squares and filled circles, respectively. Data are fitted with sinusoidal curves. Error bars indicate SE. Left- and right-eye monocular responses are indicated by triangles at the left and right margins, respectively (open triangle, neuron in *A*; filled triangle, neuron in *B*). Horizontal lines near the phase axis indicate the spontaneous firing rate. The spatial frequency of the envelope and carrier was 0.13 and 1.35 c/deg, respectively, whereas their orientations were both 65°. The envelope was drifted at 2 Hz for these cells and all other neurons in this study. The carrier was stationary for these neurons. *D–F*, Data for another pair of simultaneously recorded cells (from different electrodes), shown in the same format as for the first pair. In *F*, data for the neuron shown in *D* (open squares) are scaled up by a factor of 8 for comparison and plotted in gray, together with the original curve and data near the horizontal axis. The spatial frequencies of the envelope and the carrier were 0.08 and 1.15 c/deg, whereas their orientations were 60 and 120°, respectively. The carrier was stationary for these neurons. *G*, Distribution of modulation depth of the disparity-tuning curves. Modulation depth is defined as the ratio of the amplitude of F1 to that of the F0 component of the tuning curves (see Materials and Methods). The mean modulation ratio and its SD were 0.35 and 0.29, respectively ( $n = 70$ ). Filled bars indicate a subset of neurons for which significant modulations by the interocular phase were observed ( $n = 25$ ). *H*, Distribution of differences of optimal interocular phases for simultaneously recorded cells ( $n = 10$ ). L, Left; R, right; sp/s, spikes per second.



**Figure 4.** Insensitivity of envelope-responsive neurons to the carrier interocular phase. *A–C*, Responses of three neurons are shown as a function of the carrier interocular phase (varied in 30° steps over one cycle), whereas the interocular envelope phase was fixed at the optimal value. The neurons in *A* and *B* are the same as those shown in Figure 3, *A* and *B*, respectively. The orientation, spatial frequency, and temporal frequency of the envelope and the carrier are described in the legend to Figure 3. The difference of the response magnitude of these neurons between this figure and Fig. 3 is simply attributable to the fact that the data were obtained at different times. For the neuron in *C*, the envelope and the carrier spatial frequency were 0.18 and 1.2 c/deg, respectively. Envelope and carrier orientations were 115 and 210°, respectively. The envelope and carrier were both drifted at 2 Hz. Spontaneous firing rates were 0 for all the three neurons.

ones presented in Figure 3, *A* and *B*. These results justify our assumption that the carrier interocular phases do not affect disparity tuning of neurons for the contrast envelope. Implications of these results for binocular models to process the contrast envelope are described below.

#### Site of binocular convergence for the envelope pathway

What forms of binocular models of the contrast-envelope pathway underlie the results presented so far? Because there are two stages of processing in the contrast-envelope pathway, first- and second-stage cells, we can consider two types of models in which binocular convergence occurs at different stages, as illustrated in Figure 5. In Figure 5*A*, binocular convergence occurs at first-stage cells. This model is motivated by reports that suggest that the first-stage cells reside in area 17, taking into account the high spatial frequency tuning and orientation tuning for carriers of contrast-envelope stimuli (Mareschal and Baker, 1998a, 1999). Because the majority of the area 17 neurons are binocular, this naturally leads to the first-stage convergence model. On the other hand, several psychophysical studies have proposed another model in which binocular convergence occurs after second-stage filters, as shown in Figure 5*B* (Wilcox and Hess, 1996; McKee et al., 2004). Which model is more likely? To address this question, we simulated the behavior of each model and compared them with binocular responses of area 18 neurons.

The basic architecture of the models is described in Materials and Methods (see Simulations). In the first-stage convergence model (Fig. 5*A*), signals from the left- and right-eye filters are linearly summed, half-wave rectified, and squared at the first-stage neurons, so as to produce simple-cell-like outputs. These individual first-stage cells are tuned to a given interocular carrier phase. However, based on the results in Figure 4, it is necessary to destroy the sensitivity to the interocular carrier phase. Therefore, the model is constructed such that there are 16 overlapping arrays of first-stage cells covering the same visual field. The 16 arrays represent the first-stage cells, the left–right filters of which have the monocular and interocular phases of 0, 90, 180, and 270°, in various combinations as illustrated in Figure 5*C*. (In real neural circuits, the situation is probably closer to many first-stage cells with random monocular and interocular phases at each location.

The 16-array case is an idealized quadrature implementation for modeling.) Figure 5*A* illustrates only the array with zero monocular phase and zero interocular phase, and the remaining 15 arrays are not shown for clarity. When these signals from different arrays are pooled together at each location, the output becomes insensitive to the carrier phase monocularly and binocularly. The second-stage filters pool the signals from the first-stage cells to achieve the carrier-phase insensitivity and weight them with their filter shapes. The filter output determines the responses of the second-stage cell, subject to the half-squaring.

In the second-stage binocular convergence model (Fig. 5*B*), the first-stage cells are monocular, and the binocular convergence occurs after the second-stage filter. Here again, the figure illustrates only the even-symmetric (0-phase) first-stage filters, although there are four arrays of the

first-stage filters having filter phases of 0, 90, 180, and 270° as in the model in Figure 5*A*. Each monocular second-stage filter pools the signals from these four arrays of first-stage neurons, thereby achieving the carrier-phase insensitivity.

For both models, responses of the second-stage neurons show temporal modulation at the frequency of the drifting envelope, and we did find such “simple-type” neurons in our sample. We also found many “complex-type” neurons that did not show such temporally modulated envelope responses (Fig. 3). These results are in agreement with previous findings (Zhou and Baker, 1993). Note that such unmodulated responses may simply be constructed by applying the energy model organization to the second-stage cells. Therefore, we only simulate behaviors of a single simple-type second-stage cell as illustrated in Figure 5.

Figure 6 shows results of a simulation for the models in response to binocular contrast-envelope stimuli. Spatial frequencies of the carrier and the envelope of these stimuli are 1.0 and 0.1 c/deg, which are approximately the same as the average values used for physiological experiments. The interocular carrier phase was 0°. The second-stage filter phase was set to be 0° (even symmetric). Figure 6*A* shows results of simulations for the first-stage convergence model (Fig. 5*A*). The three panels plot tuning curves of simulated responses of the second-stage cells as functions of the envelope interocular phase, when left and right first-stage filters of the first-stage cells had position disparities of  $-4.0$ ,  $0$ , and  $4.0^\circ$  in visual angle, respectively. Peak envelope interocular phases in the three panels of Figure 6*A* are observed at  $-144$ ,  $0$ , and  $144^\circ$ , respectively. These optimal envelope phases correspond exactly to the position disparities given to the left and right first-stage filters (i.e.,  $-4$ ,  $0$ , and  $4^\circ$  in visual angle). We confirmed that identical results are obtained when we use different interocular carrier phases for the stimuli (data not shown). Results are also the same when the second-stage filter had an odd-symmetric shape (phase,  $90^\circ$ ). This is not surprising because these responses were calculated for drifting envelope stimuli, thereby making tuning curves independent of the phase of the second-stage filters. Forms of interocular phase tuning curves were not affected, regardless of whether we used the sinusoidal carrier or noise carrier, the patterns of which are correlated or

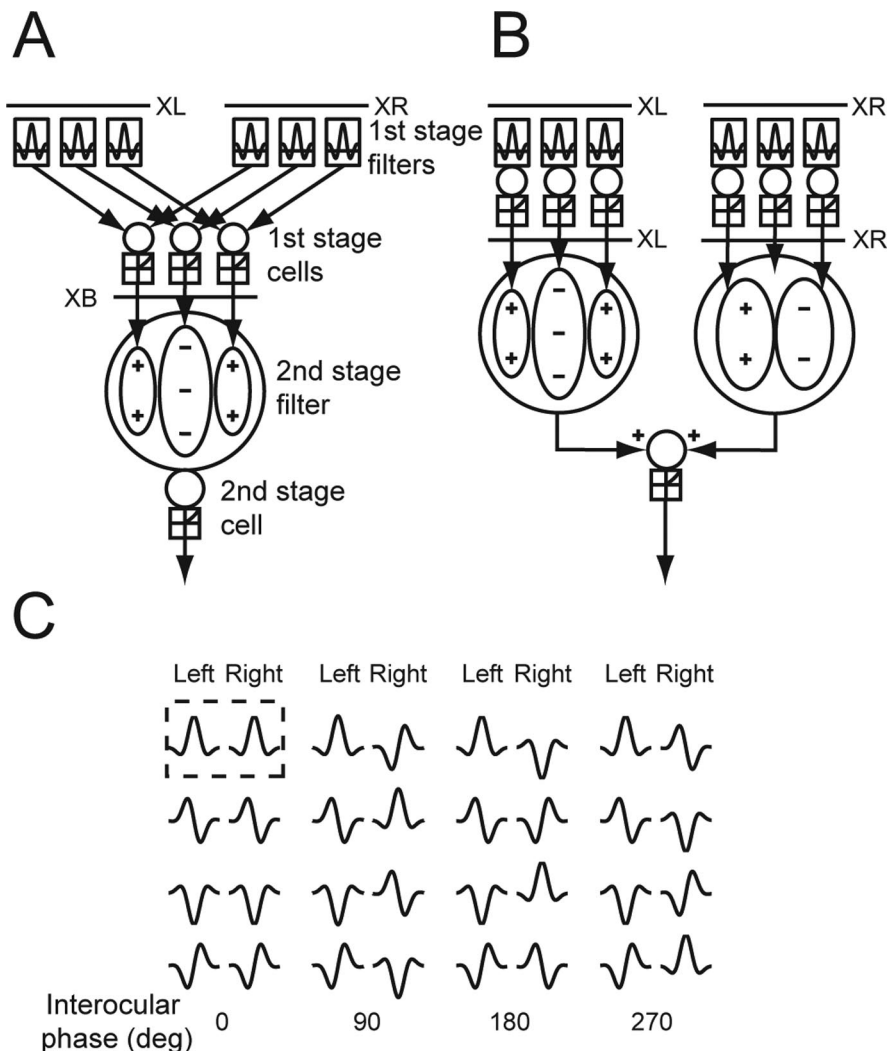
uncorrelated between eyes. This is because the first-stage filters are not sensitive to monocular or interocular carrier phases.

Simulations for the second-stage convergence model are shown in Figure 6B, in which the interocular phases of the second-stage filters are  $-144$ ,  $0$ , and  $144^\circ$ , respectively. In terms of position disparities, these values correspond to  $-4$ ,  $0$ , and  $4^\circ$  of visual angle, respectively. Therefore, for the second-stage convergence model (Fig. 5B), neurons are tuned to the interocular envelope phases that are identical to the interocular phase of the second-stage filters. The carrier interocular phase does not change tuning curves at all (data not shown), which is consistent with the results in Figure 4.

One may wonder whether disparities of contrast-envelope stimuli can be unambiguously calculated in the two models, especially when the stimuli are nonperiodic as in a complex natural environment. We confirmed by simulation that the first-stage convergence model (Fig. 5A) and second-stage convergence model (Fig. 5B) have highly similar disparity-tuning curves for noise contrast-envelope stimuli and similar bandpass properties for spatial frequency, and that unambiguous representations of disparities for these stimuli are equally possible in either of the two models by pooling the output of multiple neurons (Fleet et al., 1996) (data not shown).

Although the results of the simulations above indicate that both models can encode various envelope disparities, which is more physiologically plausible? There is a difficulty for the first-stage convergence model (Fig. 5A) in that the preferred envelope disparity is entirely determined by the position disparities of the first-stage filters. When we consider that binocular first-stage filters tuned to high spatial frequency are likely to be in area 17 and position disparities of neurons there are mostly  $<1^\circ$  (Anzai et al., 1999), the idea that this model codes various envelope disparities of up to several degrees seems highly unlikely. For example, given a typical envelope spatial frequency of  $0.1$  c/deg and a phase disparity limit of  $90^\circ$  (Fig. 3H), the maximum optimal disparity is expected to be  $2.5^\circ$ , which exceeds the upper bound in the physiological data.

Given spatially large second-order filters and physiological plausibility of correspondingly large disparities encoded either by phase or position disparities at this stage, our results are more consistent with the model in Figure 5B. However, for a subset of neurons tuned to small envelope disparities, we cannot rule out the possibility that the model in Figure 5A is used. If one considers a simplicity of the organization as a factor, the second-stage

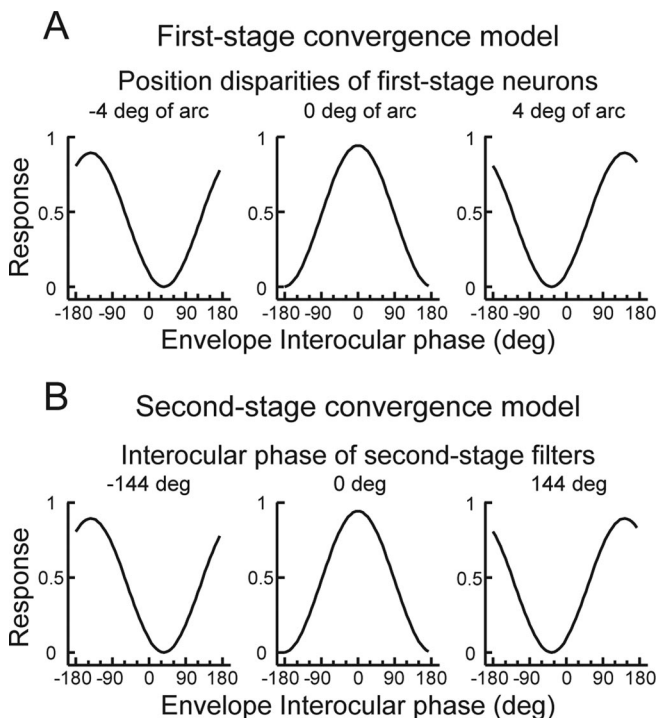


**Figure 5.** Possible models of the binocular pathway for processing envelope stimuli. (The luminance pathway is not shown for clarity.) **A**, First-stage convergence model. Signals from left- and right-eye first-stage filters are linearly combined at the first-stage neurons. These signals are then integrated at a second-stage neuron via the second-stage filter. Signals of second-stage cells are then half-wave rectified and squared. First-stage and second-stage filters are both modeled by Gabor functions. The model contains 16 overlapping arrays of identical first-stage cells. That is, within a given array, the first-stage filters have the same monocular and interocular phases selected from four possible phases:  $0$ ,  $90$ ,  $180$ , or  $270^\circ$ , resulting in a total of 16 pairings as shown in **C**. Only the array for zero monocular and interocular phases is shown for clarity (a pair inside a dashed rectangle in **C**). **B**, Second-stage convergence model. Signals from left and right first-stage filters are monocularly processed through the first-stage neurons and carried to left and right monocular second-stage filters, respectively. The model contains four overlapping arrays of first-stage cells, although only the zero-phase filter array is shown. Outputs of the second-stage filters for the two eyes are then combined at the second-stage neurons. These signals are then half-wave rectified and squared. This neuron can be tuned to various disparities, depending on the relationship of shapes of the second-stage filters between the eyes. Note that a non-zero envelope disparity preference, as illustrated, is generated by the phase difference based on the phase model (Anzai et al., 1999), although the similar envelope disparity tunings may be produced by the position model as well. **C**, 16 pairings of 4 possible monocular and interocular phases for first-stage cells are shown. XL,  $x$ -axis for left eye; XR,  $x$ -axis for right eye; XB, binocular  $x$ -axis.

convergence model in Figure 5B is more likely than the conditionally separated models required by the alternative.

#### Cue-invariant binocular interaction

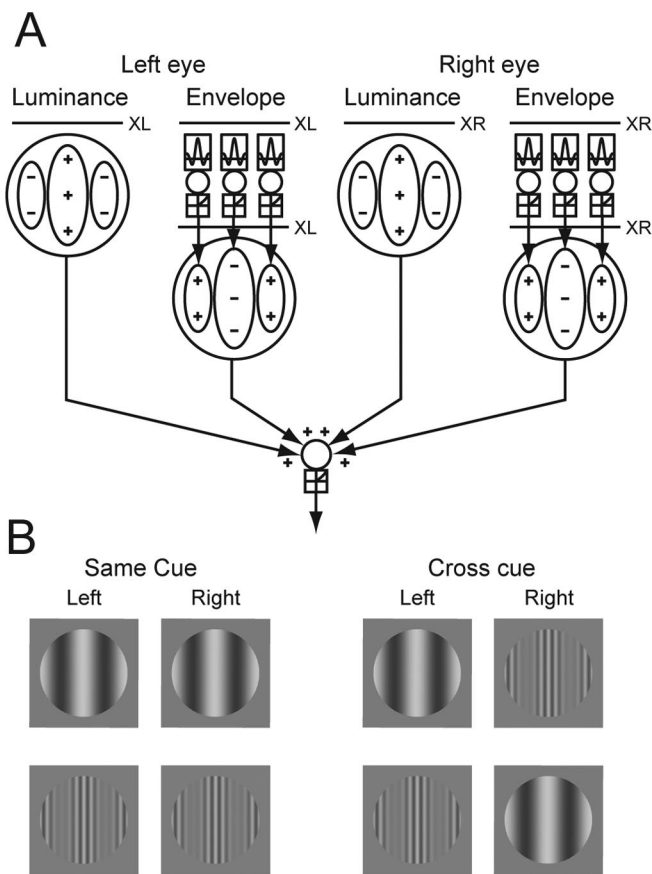
In monocular experiments, area 18 neurons that respond to contrast-envelope stimuli have been shown to respond to luminance stimuli as well (Zhou and Baker, 1993). If a neuron is tuned to disparities for both stimuli, are the optimal disparities matched for the different cues? In other words, we have examined whether there is cue invariance for binocular disparity tuning between



**Figure 6.** Results of simulations of the binocular version of the filter-rectify-filter models are shown. Spatial frequencies of the carrier and envelope of the stimuli are 1 and 0.1 c/deg, which are approximately the same as the average values used for the physiological experiments. **A**, Results of simulations for the first-stage convergence model. The left, middle, and right panels show results when left and right first-stage filters of the first-stage cells had position disparities of  $-4.0$ ,  $0$ , and  $4.0^\circ$  of visual angle, respectively. **B**, Results of simulations for the second-stage convergence model. The left, middle, and right panels show results when the interocular phase of the second-stage filters is  $-144$ ,  $0$ , and  $144^\circ$ , respectively.

contrast and luminance cues. Such cue invariance is generally thought to be a desirable characteristic and has been demonstrated for other properties (Albright, 1992; Sary et al., 1993; Zhou and Baker, 1993; Mareschal and Baker, 1998a; O’Keefe and Movshon, 1998; Tanaka et al., 2001). In addition to the question of cue invariance, we hope to construct binocular descriptions of luminance and contrast-envelope processing pathways in an integrated manner. The simplest model constructed on the basis of the results so far is a single-point linear convergence model shown in Figure 7A. If this model is true, neurons that show binocular interaction for both luminance and contrast-envelope stimuli should also show a similar interaction even if the eyes receive different types of stimuli. To test this notion, we presented binocular luminance stimuli (Fig. 7B, top left pair), binocular contrast-envelope stimuli (bottom left pair), and cross-cue stimuli (right two pairs) in which the left and right eyes see different types of stimuli, and responses of neurons were examined as a function of the interocular phase.

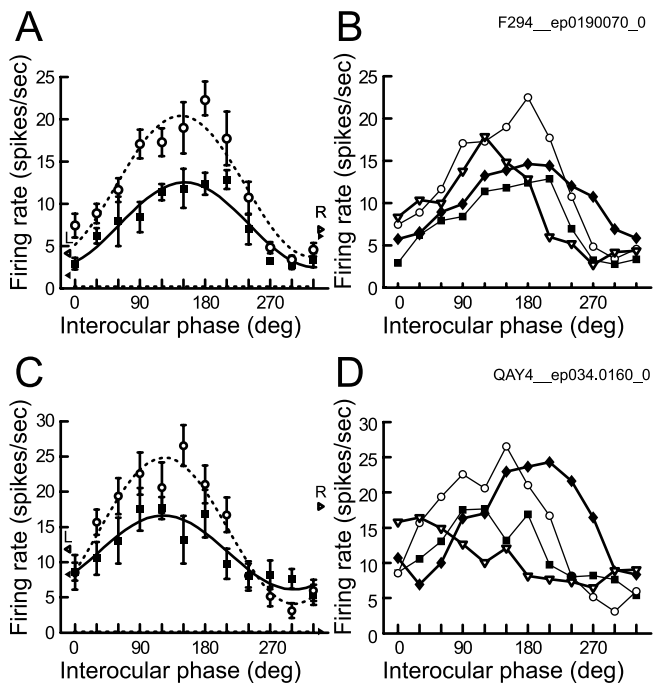
Figure 8 shows results of the cue-invariance tests for two neurons. In Figure 8A, phase tuning curves of a neuron for the luminance and contrast-envelope stimuli are depicted with open circles and filled squares, respectively. There is a marked similarity in the form of the two tuning curves in that both tuning curves peaked at very similar disparities (phases), although the response amplitude for the luminance stimuli was substantially higher. Similar results were obtained for another neuron as shown in Figure 8C. The similarity of the optimal interocular phases indicates that the cue invariance is present for these neurons.



**Figure 7.** A binocular convergence model incorporating the luminance and envelope processing pathways (**A**) and stimuli to test this model (**B**). **A**, Signals from the luminance and envelope processing pathways from the two eyes converge linearly at a single neuron. These signals are then half-wave rectified and squared. **B**, Four types of dichoptic stimuli to test this single-point linear convergence model. The left two columns show conventional stereo stimuli (same-cue stimuli). The top and second rows are luminance and contrast-envelope stimuli, respectively. The right two columns show those in which the two eyes receive different types of stimuli (cross-cue stimuli). For each binocular stimulus, the interocular phase was varied in  $30^\circ$  steps over one cycle. These binocular stimuli, together with monocular and blank (null) stimuli, were interleaved randomly in a single run. XL, x-axis for left eye; XR, x-axis for right eye.

What happens when different cue types are combined through the two eyes? Responses to cross-cue stimuli are shown in Figure 8, B and D, for the neurons in Figure 8, A and C, respectively. Phase tuning curves for these stimuli are indicated by thick curves [open triangle, luminance–envelope pairing (i.e., luminance stimuli for left-eye and contrast-envelope stimuli for the right eye); filled diamond, envelop-luminance pairing]. For comparison, tuning curves for the matched-cue stimuli (luminance stimuli for the two eyes or contrast-envelope stimuli for the two eyes) are shown again in this panel using thin lines. Even for the cross-cue stimuli, both neurons showed responses that are clearly modulated by the interocular phase. However, optimal phases for the cross-cue stimuli were different from those for the same-cue stimuli. As described below, such shifts of tuning curves are not applicable for all neurons. However, because these neurons are among those that showed the most robust responses and modulated tuning curves with statistically significant shifts (sequential *F* test;  $p < 0.05$ ), it is worth considering why these shifts occur. Applying the harmonic analysis of PSTHs obtained for drifting contrast-envelope and luminance gratings of several different temporal frequencies, Mareschal and Baker (1998b) have shown that the latency for the contrast-envelope pathway is





**Figure 8.** Cue-invariant phase tuning of neurons in area 18. **A–D**, Interocular phase tuning of two cells in cat area 18 to luminance, contrast-envelope, and cross-cue stimuli. **A**, Mean discharge rates of a cell to binocular luminance and contrast-envelope stimuli are plotted as functions of the interocular phase with open circles and filled squares, respectively. Fitted sinusoids are indicated by dashed and solid curves for luminance and contrast-envelope stimuli, respectively. Left- and right-eye monocular responses are indicated by triangles at the left and right margins, respectively (open triangles, luminance; filled triangles, contrast-envelope stimuli). Note that the optimal interocular phases for the two stimuli are approximately the same, indicating cue invariance. Error bars indicate the SEM. A dashed line almost overlapping the horizontal axis depicts the level of spontaneous discharge. **B**, Responses to cross-cue stimuli for the neuron in **A**. The inverted triangles indicate responses for the conditions in which luminance and contrast-envelope stimuli are presented for the left and right eye, respectively. Filled diamonds indicate responses to the other cross-cue stimuli pairing (see Fig. 7*B*). Tuning curves for the same-cue stimuli are shown again with thin lines. Note the clear modulations of these tuning curves. **C, D**, Data for another neuron, shown in the same format as for the first neuron. The spatial frequency of the luminance stimuli and envelope of the contrast-envelope stimuli was 0.13 c/deg for the cell in **A** and 0.08 c/deg for the cell in **C**. The spatial frequency for the carrier was 1.35 c/deg for both cells. For both neurons, the envelope was drifted at 2 Hz, and the carrier was stationary.

longer than that for the luminance pathway by ~100–500 ms. Because the peak interocular phase for drifting stimuli depends not only on the spatial difference of RFs, but also on differences of latencies between the eyes, the envelope–luminance latency difference may explain the phase shifts observed in Figure 8, *B* and *D*. This idea predicts that the tuning curves for the luminance–envelope pairing (open triangles) and the envelope–luminance pairing (filled diamonds) are shifted in the opposite directions from those for the same-cue pairings, respectively. This is what is actually observed in Figure 8. These phase shifts approximately correspond to a latency difference of 40–90 ms.

Figure 9 shows the population summary of the results. In Figure 9*A*, the observation, as mentioned above, that the luminance stimuli are more effective than the contrast-envelope stimuli is examined for the population data ( $n = 70$ ). In this figure, the maximum binocular response to the luminance stimuli for each cell is plotted against that to the contrast-envelope stimuli. As shown by the fact that most of the data points lie above the diagonal, these neurons generally responded much more strongly to the luminance stimuli (mean, 29.7 spikes/s; SD, 21.1 spikes/s) than the envelope stimuli (mean, 11.4 spikes/s; SD, 12.1 spikes/s).

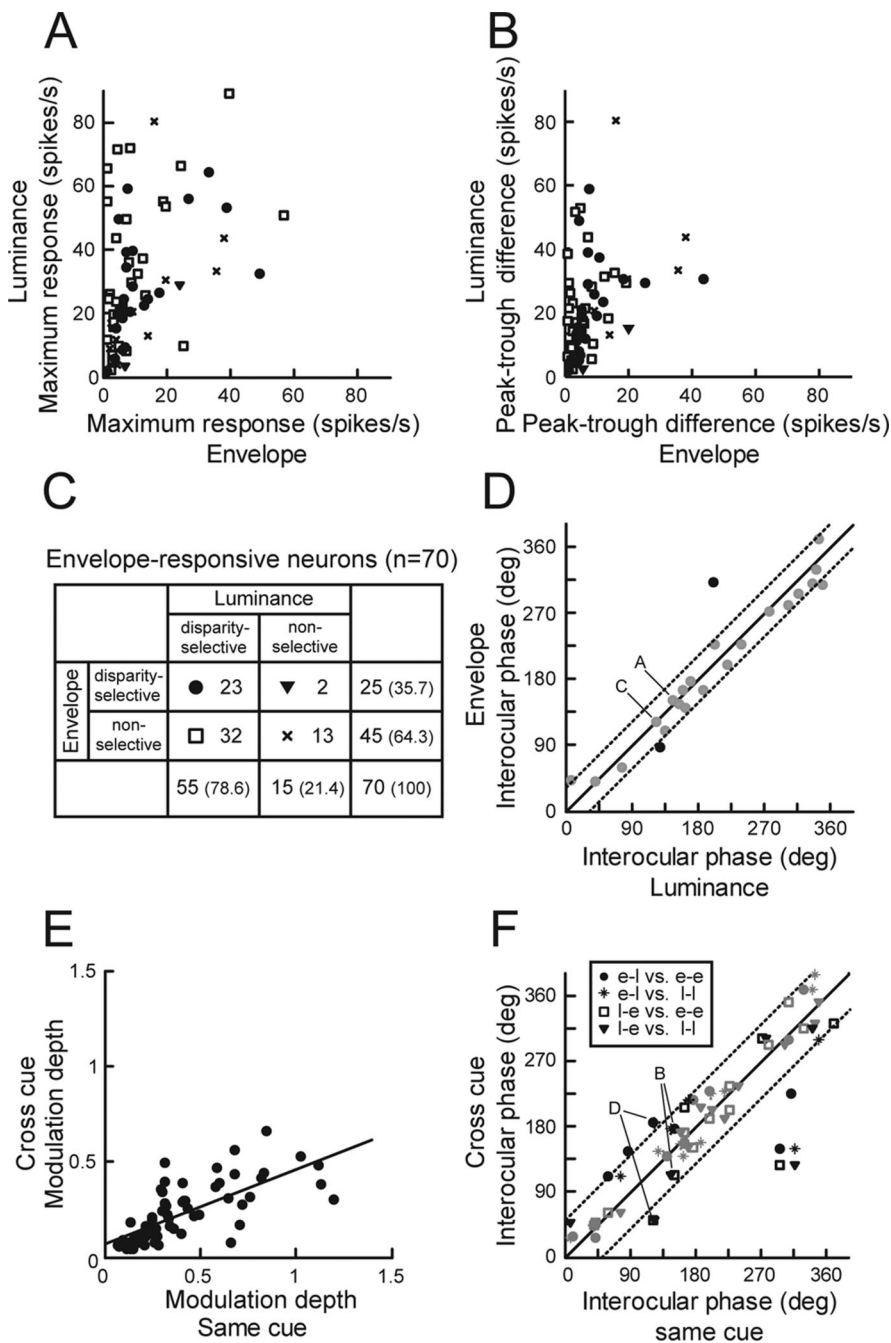
A similar difference in terms of modulations of responses by binocular disparity (interocular phase) is also observed between the luminance and contrast-envelope stimuli. The peak-to-trough response amplitude difference of disparity-tuning curves for the luminance stimuli (18.4 spikes/s; SD, 13.5 spikes/s) was also larger than that for the contrast-envelope stimuli (6.97 spikes/s; SD, 6.96 spikes/s) in most cases (Fig. 9*B*).

The difference in the effectiveness of luminance and contrast-envelope stimuli seems to influence the degree of disparity selectivity. As shown in Figure 9*C*, of the 70 envelope-responsive neurons, 55 (78.6%) were disparity selective for luminance stimuli. This proportion was not significantly different from that for the unselected population of area 18 neurons (70.9%; 107 of 151 neurons;  $p > 0.05$ ,  $\chi^2$  test). Of these 55 neurons, only 23 were disparity selective to the envelope stimuli, and the remaining 32 neurons were not disparity selective for the envelope stimuli. These 32 neurons signal disparity via luminance only. Therefore, they cannot be described as cue invariant, but at least no conflicting information about disparity would be signaled with different cues.

What is the proportion of neurons that exhibit cue invariance for the luminance and contrast-envelope stimuli as shown in Figure 8, *A* and *C*? In Figure 9*D*, the peak phase for the luminance stimuli is plotted against that for the contrast-envelope stimuli for 23 neurons that were disparity selective for both luminance and contrast-envelope stimuli. All but four neurons were within  $\pm 33^\circ$  (1 SD) about the diagonal (dashed lines). Moreover, all but two neurons do not reveal statistically significant shifts of optimal phases (sequential *F* test). This indicates cue invariance in that neurons are tuned to highly similar binocular disparities, regardless of whether contrast or luminance cues are used.

Figure 9*E* plots the average modulation depth of two phase tuning curves for the cross-cue stimuli against that for the two phase tuning curves for the same-cue stimuli. Notice that the two indices are correlated ( $n = 66$ ;  $r = 0.72$ ;  $p < 0.0001$ ), although the degree of the modulation for the cross-cue stimuli was generally weaker. This modulation for the cross-cue stimuli is consistent with the simple linear convergence model. The modulated responses for the cross-cue stimuli are also expected by a model in which binocular convergence occurs separately for envelope and luminance pathways first at intermediate neurons, and then rectified outputs from each pathway are combined. However, if this is true, there should be many neurons responsive to contrast-envelope stimuli only without sensitivity to luminance cues. We (Fig. 9*A–C*) and others (Zhou and Baker, 1993; Mareschal and Baker, 1998a) have not found neurons that clearly show such characteristics.

Do disparity tunings for dichoptic combinations of different cues (cross-cue stimuli) change from those for the matched-cue conditions, as shown in Figure 8, *B* and *D*? Of 23 neurons that were disparity selective for both binocular luminance and envelope stimuli, 20 also showed disparity selectivity for at least either of the cross-cue stimuli. For each of these neurons, we plotted the peak interocular phase for the same-cue conditions (binocular luminance stimuli and binocular contrast-envelope stimuli) against that for the two cross-cue conditions (luminance–envelope pairing and envelope–luminance pairing), if paired tuning curves (one for the same cue and the other for the cross-cue) are both disparity selective (Fig. 9*F*). Because each neuron has one to four data points (depending on the significance of tuning), a total of 66 points are plotted in this figure. The dashed lines are at  $\pm 52^\circ$  about the diagonal representing 1 SD limit of the distribution. Statistical tests revealed that 24 of these points (36%) had signif-



**Figure 9.** Comparison of responses of envelope-responsive neurons for luminance and contrast-envelope stimuli. **A**, Peak binocular response of each neuron to binocular luminance stimuli is plotted against that for the contrast-envelope stimuli. Solid circles indicate neurons that were disparity selective for both stimuli ( $n = 23$ ). Open squares and solid triangles indicate those that were disparity selective for only luminance ( $n = 32$ ) and contrast-envelope ( $n = 2$ ) stimuli, respectively. The x symbols indicate those that were disparity selective for neither stimulus ( $n = 13$ ). See **C** for illustration of the meaning of these markers. **B**, Peak-to-trough difference of disparity-tuning curves is compared for luminance and contrast-envelope stimuli. The convention of the markers is the same as that in **A**. **C**, Classification of the data population. The table shows the classification of 70 envelope-responsive neurons into four groups according to their disparity selectivity for luminance and contrast-envelope stimuli. The sum of these numbers are shown in the right column and the bottom row, respectively. The proportion (percentage) of these sums of the total population ( $n = 70$ ) is indicated by the number in parentheses. The markers used in Figure 9, **A** and **B**, indicating these groups are also shown within the corresponding cells. **D**, Optimal interocular phases for contrast-envelope stimuli are plotted against those for luminance stimuli ( $n = 23$ ). Most neurons are within a narrow range about the diagonal, indicating cue invariance. Black symbols indicate cells with a statistically significant shift of optimal phases between the two conditions. Dashed lines indicate a 1 SD limit of  $\pm 33^\circ$ . Data for the neurons in Figure 8, **A** and **C**, are indicated by letters. **E**, Modulation depth for the cross-cue stimuli. The average modulation depth of two phase tuning curves for the cross-cue condition is plotted against that of the two phase tuning curves for the same-cue condition. Notice that the two indices are correlated ( $r = 0.72$ ;  $p < 0.0001$ ;  $n = 66$ ), although the modulation depth for the cross-cue stimuli was generally weaker. The oblique line indicates a regression line (slope, 0.38). **F**, Optimal interocular phases for each of the same-cue stimuli (binocular luminance or contrast-envelope stimuli)

icant shifts of the disparity-tuning curves (sequential  $F$  test, black symbols). Note also that data for the same-cue versus cross-cue comparison (Fig. 9F) are more broadly scattered about the diagonal than those for the same-cue comparison (Fig. 9D), as indicated by the narrower 1 SD limit (dashed lines) for the same-cue condition. The variance of the peak phase difference between the same-cue and cross-cue condition (variance, 2669;  $n = 66$ ) was statistically significantly larger than that between the two same-cue conditions [variance, 1000;  $n = 23$ ;  $F$  test;  $F$  value, 2.67;  $p < 0.01$ ;  $df = (65,22)$ ]. This indicates a greater degree of cue invariance for the same-cue condition. In Figure 8, **B** and **D**, peaks of the phase tuning curves for the two cross-cue conditions were shifted in opposite directions about those for the same-cue conditions. This finding was not always true for other neurons when examined individually, because no systematic separation of different symbol types is particularly apparent in Figure 9F by casual inspections. Nevertheless, when we sorted the population data into two groups according to the predicted directions of peak phase shifts [circles and asterisks ( $n = 30$ ) vs squares and inverted triangles ( $n = 32$ )], we found that the mean peak phase shifts were statistically significantly different between these two groups [ $13.5^\circ$  ( $n = 30$ ) and  $-6.5^\circ$  ( $n = 32$ );  $p < 0.05$ ,  $t$  test]. (The four outliers in the bottom right part of Fig. 9F were excluded from this analysis, because their phase shifts exceeded  $135^\circ$ , which makes the determination of the shift direction ambiguous.) This indicates that one reason for the larger scatter for the cross-cue condition may be differences in latency between luminance and contrast-envelope cues (Mareschal and Baker, 1998b). Together, these results suggest that the disparity tunings for the cross-cue condition tended to change from those for the same-cue conditions.

Because contrast-envelope stimuli are

←

are plotted against those for the cross-cue stimuli, for 20 neurons that were disparity selective for both of the same-cue stimuli and at least either of the two cross-cue stimuli. Data for each cell are represented by up to four symbols (depending on significance), because there are four possible pairings of same-cue and cross-cue stimuli (see inset). The inset shows the symbol legend depicting left–right pairings of luminance (l) and contrast-envelope (e) stimuli ( $n = 16, 16, 17$ , and 17 from top to bottom). Filled symbols indicate data with a statistically significant shift of optimal disparities between the two conditions. Dashed lines representing the 1 SD limit are at  $\pm 52^\circ$ . Data for the neurons in Figure 8, **B** and **D**, are indicated by letters.

generally weaker than the luminance stimuli for driving neurons, one may argue that the differences of optimal disparities between luminance and cross-cue conditions are attributable to changes in latencies that are dependent on the stimulus strength. However, this is unlikely, because it has been shown that, using binocular luminance stimuli, the optimal interocular phases are hardly affected by the stimulus contrast. This is true even under conditions in which the stimulus contrasts are mismatched by a factor of 10 interocularly (Smith et al., 1997b; Truchard et al., 2000). Therefore, the shifts of optimal disparities between the same-cue and cross-cue condition seem to be attributable to differences in cues, rather than differences in effective strengths of stimuli.

## Discussion

We have shown that cat area 18 neurons respond to second-order boundaries (contrast envelope) with clear dependence on their interocular phase difference. In addition, these neurons show similar disparity-tuning curves for luminance and second-order cues. This cue invariance may be accounted for by linear convergence of monocular luminance and contrast-envelope processing pathways. Modulation of responses by the interocular phase for “cross-cue stimuli” further supports this model. We discuss below possible implications of these results in the context of previous studies.

### Relationship to other physiological studies

To the best of our knowledge, there has not been a physiological study that systematically investigated stereoscopic coding of contrast-envelope stimuli. Among possibly related studies (Bakin et al., 2000; Cumming and Parker, 2000), Bakin et al. (2000) reported that ~40% of disparity-selective neurons in V2 responded to edge disparities of repetitive bars extending into the surround of their classical RFs. Because the purpose of their study was to examine modulatory contextual effects on the disparity selectivity from stimuli far outside the classical RF, they did not try to constrain the stimuli such that their spatial frequency content were outside the luminance pass-band of the neurons. Therefore, it is not clear whether the V2 neurons signal disparities for boundaries defined solely in contrast cues, and direct comparisons with our results are not possible.

In our experiments, responses to contrast-envelope stimuli are not attributable to responses to luminance boundaries, because our stimuli contain no luminance energy at all within the luminance spatial frequency pass-band of the neurons (see Materials and Methods). Moreover, these neurons are often narrowly tuned to the spatial frequency of the carrier (Fig. 2). Therefore, our results are based on pure envelope responses generated through neural nonlinear mechanisms (Fig. 1).

### Comparisons with findings from psychophysical and computational studies

One important role of contrast-envelope cues is that they provide depth information over a wide range of binocular disparities. Wilcox and Hess (1995) have shown that, using Gabor patches, the upper disparity limit for stereopsis is determined by disparities of the contrast envelope rather than the carrier parameters. The limit was linearly related to the envelope size and always much above the Panum fusional area, the largest disparity below which stimuli are binocularly fused and appear as single. If envelope-sensitive neurons are closely related to this depth perception, they should be able to code a wide range of disparities as a population. They should also reveal size-disparity correlation,

in that neurons with larger RFs encode a larger range of disparities. Unfortunately, periodic stimuli used in our study do not allow direct measurements of the disparity coding range of the cells. However, the cue invariance in disparity selectivity suggests that response properties found for luminance stimuli are likely to apply to contrast-envelope stimuli also. Therefore, given the scatter of optimal disparities for luminance stimuli, which reaches 5° in cat area 18 (Ferster, 1981) and the size-disparity correlation found for luminance stimuli (Ohzawa et al., 1997; Prince et al., 2002), envelope-sensitive neurons should have desired properties for mediating depth perception based on the contrast envelope.

Achieving stereopsis in a complex visual environment has been an intensively studied topic. As pointed out by Julesz (1971) and Marr and Poggio (1979), depth signals given by local luminance features are often false or ambiguous in complex scenes in which there are many possibilities of local binocular matches. Several mechanisms have been suggested to reduce these false matches and compute depth correctly, such as coarse-to-fine processing (Marr and Poggio, 1979) and pooling of disparity sensor activities over positions, multiple spatial frequencies, and orientations (Fleet et al., 1996; Qian and Zhu, 1997). For textured surfaces, coarse-scale disparity information is effectively obtained using the contrast envelope (Wilcox and Hess, 1995, 1997; Schor et al., 1998). Therefore, the notion of coarse-to-fine and multichannel pooling probably should be extended to include the additional disparity information originating from the envelope-processing pathway.

Disparities at the surface edges are also suggested to be a strong cue for stereo-matching (McKee and Mitchison, 1988), especially for surfaces with repetitive texture elements. McKee et al. (2004) recently found that depth judgment performance highly depends on depth signals based on disparities of the contrast envelope, but not on disparities derived from luminance cues at the surface edge. These results indicate that disparity information from the contrast envelope is important in depth perception.

The proposed model for the binocular contrast-envelope processing pathway (Fig. 5B) possesses a structure in which nonlinear rectification occurs before binocular convergence. This structure has a superficial similarity to that found in the model proposed by Read et al. (2002), which explains reduced responses to anti-correlated random-dot stereograms. In their model, second-stage binocular simple neurons receive input from first-stage monocular simple neurons. As with our model, the binocular neurons studied by Read et al. (2002; their Fig. 6) can be tuned to non-zero envelope disparities if retinal positions of RFs of the left and right monocular neurons are different. However, the critical difference is that their model does not include a second-stage filter that sums activities of the first-stage filters. Without such a second-stage filter, no selectivity to contrast envelopes can be generated. In our model, the second-stage neurons implement filtering by collecting excitatory and inhibitory inputs from many first-stage filters over a wide range of spatial positions, thereby generating envelope selectivity. Therefore, there is a fundamental difference in these two models.

We found that disparity tunings for cross-cue stimuli were clearly present (Fig. 9E), although there is greater variability in the disparity values neurons can signal for these stimuli (Fig. 9F). Consistent with these neuronal properties, Edwards et al. (2000) showed that human subjects can transiently perceive depth based on disparities in the cross-cue stimuli. But, this perception depends on the contrast of luminance stimuli for one eye. Perhaps, good stereo-performance requires effective strength of signals

from the two eyes to be balanced, as is shown for the binocular luminance stimuli (Halpern and Blake, 1988; Legge and Gu, 1989).

### Organization of envelope processing pathways

What is the neural basis for the first-stage filters (Fig. 1)? Mareschal and Baker (1998a, 1999) found that area 18 neurons are selective to carrier orientation and suggested that the first-stage neurons reside within the cortex. Because most of area 18 neurons are tuned to very low spatial frequencies ( $<0.25$  c/deg) and not to the frequency range of the carrier (0.5–2 c/deg), a likely candidate would be area 17 neurons that can be tuned to as high as 2 c/deg (Movshon et al., 1978). Considering that the majority of these cortical neurons are binocular (Ohzawa and Freeman, 1986a,b; Smith et al., 1997a), it seems that the model with binocular first-stage neurons should be a reasonable organization. However, our results support the model in which the first-stage neurons are monocular (Fig. 5B). Of course, there is still a possibility that first-stage neurons are a monocular subset of area 17 cortical neurons. However, the idea that only monocular neurons are selectively used for constructing envelope-sensitive neurons in area 18 seems questionable.

An alternative hypothesis is that first-stage neurons are in subcortical pathways, possibly in the lateral geniculate nucleus (LGN). Because LGN neurons are essentially monocular, this hypothesis is consistent with the monocular first-stage neurons. LGN cells are sensitive to as high a spatial frequency as area 17 neurons (Derrington and Fuchs, 1979). Furthermore, studies on the guinea pig showed that retinal Y cells respond to the contrast-envelope stimuli (Demb et al., 2001). These suggest that direct LGN input into area 18 (Ferster, 1990) may be a basis for responses to contrast-envelope stimuli in area 18. Although some neurons exhibit a clear tuning for carrier orientation, the majority of envelope responses are weakly tuned or essentially untuned for carrier orientation (Mareschal and Baker, 1999). Therefore, the subcortical origin for the carrier signal is a likely possibility for some, if not all, cortical neurons. But, this is not consistent with the psychophysical studies that suggest the oriented first-stage filters (Langley et al., 1996; Wilcox and Hess, 1996). Additional studies are needed to resolve these issues.

### Single or parallel pathways for different cues?

A cue-invariant disparity tuning requires neurons to have precisely phase-aligned RFs for luminance and contrast-envelope processing pathways. If four pathways (for two cues and two eyes) are completely separate before they arrive at area 18 cells, as shown in Fig. 7, it appears difficult for the neurons to have such an alignment. A possible modification, for easier alignment, is to construct the model such that the contrast-envelope pathway converges with the luminance pathway after the first-stage filter stage but before the second-stage filter stage. This allows the two pathways processed by common (second-stage) filters. Note that this is quite different from the single pathway model based on the point-wise nonlinearity (see Materials and Methods), because this model has the filter-rectify-filter organization. Note also that, because the second-stage filter is linear, this model is still within a framework of linear convergence of separate luminance and envelope pathways, a parallel-pathway model, although the two pathways overlap greatly. A future study is required to examine whether such schemes are the case.

### References

Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A* 2:284–299.

- Albright TD (1992) Form-cue invariant motion processing in primate visual cortex. *Science* 255:1141–1143.
- Anzai A, Ohzawa I, Freeman RD (1999) Neural mechanisms for encoding binocular disparity: receptive field position versus phase. *J Neurophysiol* 82:874–890.
- Bakin JS, Nakayama K, Gilbert CD (2000) Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. *J Neurosci* 20:8188–8198.
- Cavanagh P, Mather G (1989) Motion: the long and short of it. *Spat Vis* 4:103–129.
- Chubb C, Sperling G (1988) Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception. *J Opt Soc Am A* 5:1986–2007.
- Cumming BG, Parker AJ (2000) Local disparity not perceived depth is signaled by binocular neurons in cortical area V1 of the macaque. *J Neurosci* 20:4758–4767.
- Demb JB, Zaghoul K, Sterling P (2001) Cellular basis for the response to second-order motion cues in Y retinal ganglion cells. *Neuron* 32:711–721.
- Derrington AM, Fuchs AF (1979) Spatial and temporal properties of X and Y cells in the cat lateral geniculate nucleus. *J Physiol (Lond)* 293:347–364.
- Draper NR, Smith H (1998) *Applied regression analysis*, Ed 3. New York: Wiley.
- Edwards M, Pope DR, Schor CM (1999) Orientation tuning of the transient-stereopsis system. *Vision Res* 39:2717–2727.
- Edwards M, Pope DR, Schor CM (2000) First- and second-order processing in transient stereopsis. *Vision Res* 40:2645–2651.
- Ferster D (1981) A comparison of binocular depth mechanisms in areas 17 and 18 of the cat visual cortex. *J Physiol (Lond)* 311:623–655.
- Ferster D (1990) X- and Y-mediated current sources in areas 17 and 18 of cat visual cortex. *Vis Neurosci* 4:135–145.
- Fleet DJ, Wagner H, Heeger DJ (1996) Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision Res* 36:1839–1857.
- Freeman RD, Ohzawa I (1988) Monocularly deprived cats: binocular tests of cortical cells reveal functional connections from the deprived eye. *J Neurosci* 8:2491–2506.
- Halpern DL, Blake RR (1988) How contrast affects stereoacuity. *Perception* 17:483–495.
- Hess RF, Wilcox LM (1994) Linear and non-linear filtering in stereopsis. *Vision Res* 34:2431–2438.
- Hubel DH, Wiesel TN (1970) Stereoscopic vision in macaque monkey. Cells sensitive to binocular depth in area 18 of the macaque monkey cortex. *Nature* 225:41–42.
- Julesz B (1971) *Foundations of cyclopean perception*. Chicago: The University of Chicago.
- Langley K, Fleet DJ, Hibbard PB (1996) Linear filtering precedes nonlinear processing in early vision. *Curr Biol* 6:891–896.
- Langley K, Fleet DJ, Hibbard PB (1999) Stereopsis from contrast envelopes. *Vision Res* 39:2313–2324.
- Legge GE, Gu YC (1989) Stereopsis and contrast. *Vision Res* 29:989–1004.
- Mareschal I, Baker Jr CL (1998a) A cortical locus for the processing of contrast-defined contours. *Nat Neurosci* 1:150–154.
- Mareschal I, Baker Jr CL (1998b) Temporal and spatial response to second-order stimuli in cat area 18. *J Neurophysiol* 80:2811–2823.
- Mareschal I, Baker Jr CL (1999) Cortical processing of second-order motion. *Vis Neurosci* 16:527–540.
- Marr D, Poggio T (1979) A computational theory of human stereo vision. *Proc R Soc Lond B Biol Sci* 204:301–328.
- McKee SP, Mitchison GJ (1988) The role of retinal correspondence in stereoscopic matching. *Vision Res* 28:1001–1012.
- McKee SP, Verghese P, Farell B (2004) What is the depth of a sinusoidal grating? *J Vis* 4:524–538.
- Movshon JA, Thompson ID, Tolhurst DJ (1978) Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J Physiol (Lond)* 283:53–77.
- Nishimoto S, Arai M, Ohzawa I (2005) Accuracy of subspace mapping of spatiotemporal frequency domain visual receptive fields. *J Neurophysiol* 93:3524–3536.
- Ohzawa I, Freeman RD (1986a) The binocular organization of simple cells in the cat's visual cortex. *J Neurophysiol* 56:221–242.
- Ohzawa I, Freeman RD (1986b) The binocular organization of complex cells in the cat's visual cortex. *J Neurophysiol* 56:243–259.

- Ohzawa I, DeAngelis GC, Freeman RD (1990) Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science* 249:1037–1041.
- Ohzawa I, DeAngelis GC, Freeman RD (1996) Encoding of binocular disparity by simple cells in the cat's visual cortex. *J Neurophysiol* 75:1779–1805.
- Ohzawa I, DeAngelis GC, Freeman RD (1997) Encoding of binocular disparity by complex cells in the cat's visual cortex. *J Neurophysiol* 77:2879–2909.
- O'Keefe LP, Movshon JA (1998) Processing of first- and second-order motion signals by neurons in area MT of the macaque monkey. *Vis Neurosci* 15:305–317.
- Prince SJ, Cumming BG, Parker AJ (2002) Range and mechanism of encoding of horizontal disparity in macaque V1. *J Neurophysiol* 87:209–221.
- Qian N, Zhu Y (1997) Physiological computation of binocular disparity. *Vision Res* 37:1811–1827.
- Read JC, Parker AJ, Cumming BG (2002) A simple model accounts for the response of disparity-tuned V1 neurons to anticorrelated images. *Vis Neurosci* 19:735–753.
- Sary G, Vogels R, Orban GA (1993) Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science* 260:995–997.
- Schor CM, Edwards M, Pope DR (1998) Spatial-frequency and contrast tuning of the transient-stereopsis system. *Vision Res* 38:3057–3068.
- Skottun BC, De Valois RL, Grosf DH, Movshon JA, Albrecht DG, Bonds AB (1991) Classifying simple and complex cells on the basis of response modulation. *Vision Res* 31:1079–1086.
- Smith AT (1994) The detection of second-order motion. In: *Visual detection of motion* (Smith AT, Snowden RJ, eds), pp 145–176. London: Academic.
- Smith III EL, Chino YM, Ni J, Ridder III WH, Crawford ML (1997a) Binocular spatial phase tuning characteristics of neurons in the macaque striate cortex. *J Neurophysiol* 78:351–365.
- Smith III EL, Chino Y, Ni J, Cheng H (1997b) Binocular combination of contrast signals by striate cortical neurons in the monkey. *J Neurophysiol* 78:366–382.
- Tanaka H, Uka T, Yoshiyama K, Kato M, Fujita I (2001) Processing of shape defined by disparity in monkey inferior temporal cortex. *J Neurophysiol* 85:735–744.
- Thomas OM, Cumming BG, Parker AJ (2002) A specialization for relative disparity in V2. *Nat Neurosci* 5:472–478.
- Truchard AM, Ohzawa I, Freeman RD (2000) Contrast gain control in the visual cortex: monocular versus binocular mechanisms. *J Neurosci* 20:3017–3032.
- Wilcox LM, Hess RF (1995) Dmax for stereopsis depends on size, not spatial frequency content. *Vision Res* 35:1061–1069.
- Wilcox LM, Hess RF (1996) Is the site of non-linear filtering in stereopsis before or after binocular combination? *Vision Res* 36:391–399.
- Wilcox LM, Hess RF (1997) Scale selection for second-order (non-linear) stereopsis. *Vision Res* 37:2981–2992.
- Zhou YX, Baker Jr CL (1993) A processing stream in mammalian visual cortex neurons for non-Fourier responses. *Science* 261:98–101.