Encodings of Three-Dimensional Surface Slant in Cat Visual Areas 17 and 18

Takahisa M. Sanada1 and Izumi Ohzawa1,2
1Graduate School of Engineering Science and 2Graduate School of Frontier Biosciences, Osaka University, Osaka, Japan

Submitted 9 September 2005; accepted in final form 29 December 2005

Sanada, Takahisa M. and Izumi Ohzawa. Encoding of three-dimensional surface slant in cat visual areas 17 and 18. J Neurophysiol 95: 2768–2786, 2006. First published January 4, 2006; doi:10.1152/jn.00955.2005. How are surface orientations of three-dimensional objects and scenes represented in the visual system? We have examined an idea that these surface orientations are encoded by neurons with a variety of tilts in their binocular receptive field (RF) structure. To examine whether neurons in the early visual areas are capable of encoding surface orientations, we have recorded from single neurons extracellularly in areas 17 and 18 of the cat using standard electrophysiological methods. Binocular RF structures are obtained using a binocular version of the reverse correlation technique. About 30% of binocularly responsive neurons have RFs with statistically significant tilts from the frontoparallel plane. The degree of tilts is sufficient for representing the range of surface slants found in typical visual environments. For a subset of neurons having significant RF tilts, the degrees of tilt are correlated with the preferred spatial frequency difference between the two eyes, indicating that a modified disparity energy model can account for the selectivity, at least partially. However, not all cases could be explained by this model, suggesting that multiple mechanisms may be responsible. Therefore an alternative hypothesis is also examined, where the tilt is generated by pooling of multiple disparity detectors whose preferred disparities progressively shift over space. Although there is evidence for extensive spatial pooling, this hypothesis was not satisfactory either, in that the neurons with extensive pooling tended to prefer an untitled surface. Our results suggest that encoding of surface orientations may begin with the binocular neurons in the early visual cortex.

INTRODUCTION

One of the fundamental roles of the visual system is to reconstruct a three-dimensional (3D) model of the external world from a pair of two-dimensional images on the two retinas. Horizontal displacement of the eyes causes small differences between the retinal images. This difference of the retinal images is called binocular disparity and stereopsis is the process of determining depth from binocular disparity. Visual information processing for stereopsis begins in the primary visual cortex and neurons found in this area are known to encode binocular disparities of stimuli for a small area of visual field (Barlow et al. 1967; Ferster 1981; Hubel and Wiesel 1962, 1968; LeVay and Voigt 1988; Nikara et al. 1968; Ohzawa and Freeman 1986a,b; Ohzawa et al. 1990, 1996, 1997).

How does the processing of stereoscopic information proceed once binocular disparity for small localized areas is available? Is it possible next stage of processing that of detecting the rate of change of binocular disparity, i.e., detecting 3D orientations of surfaces in depth? Some recent studies have examined these possibilities and report that a subset of neurons in higher visual areas such as MT, V4, and CIPs encode information regarding slant/tilt of surfaces (Hinkle and Connor 2002; Nguyenkim and DeAngelis 2003; Taira et al. 2000). Response to 3D curvature is also reported in the inferotemporal cortex (IT) (Janssen et al. 1999, 2000 Liu et al. 2004). It is not known, however, whether such surface slant/tilt sensitivity is a unique feature of these higher-order visual areas. Because neurons in these areas receive inputs from primary visual cortex, selectivity for 3D surface slant/tilt may be inherited from the early visual areas. Historically, the role of interocular orientation difference has been examined in some detail (Blakemore et al. 1972; Nelson et al. 1977). More recent work has examined whether a subset of V1 neurons encode surface tilt by orientation disparity based on physiological experiments in the monkey and a computational study (Bridge and Cumming 2001; Bridge et al. 2001). However, possible roles of interocular spatial frequency difference have not been examined physiologically.

As illustrated in Fig. 1A, projection of a slanted surface onto the two retinas produces a spatial frequency difference, such that the eye closer to the nearer end of the slanted surface sees higher spatial frequency than the other eye. Such a difference in spatial frequency across the eyes is a potent cue for perceiving surface slants. With psychophysical experiments, Blakemore and later investigators reported that a difference of spatial frequency across the eyes produces a perception of slant-in-depth (Blakemore 1970; Fiorentini and Maffei 1971; Wilson 1976). Binocular disparity caused by interocular spatial frequency difference is designated dif-frequency disparity (Tyler and Sutter 1979). As expected, the angle of perceived surface slant depends on interocular ratio of spatial frequencies. Despite these psychophysical results, we are not aware of any physiological study that has systematically examined possible roles of dif-frequency disparity for encoding surface slant in the early visual cortex. In this study, we will thus address this question using modern receptive field-mapping techniques.

To provide a framework within which we design our experiments and analyze data, we start with the standard disparity energy model (Ohzawa et al. 1990). With the standard disparity energy model, parameter values for orientation, spatial frequency, position, and size of their monocular receptive field are the same across the eyes. Only the receptive field phase is allowed to be different across the eyes, and this difference in phase determines preferred binocular disparity. One obvious way to incorporate dif-frequency disparity sensitivity is to modify the standard disparity energy model and to allow spatial frequency to be different between left- and right-eye receptive fields of all subunits (S) (Fig. 1B). This idea was...
same spatial frequencies for the two eyes (Hammond and Pomfrett 1991; Read and Cumming 2003).

Considering the predictions illustrated above from a modified version of the disparity energy model, we will first examine the extent to which neurons in areas 17 and 18 of the cat visual cortex exhibit tilted binocular RFs. We will also examine the validity of the dif-frequency disparity energy model by comparing the degree of tilt of binocular RF and monocular spatial frequency-tuning curves for the two eyes.

**Methods**

All animal care and experimental procedures conformed to those established by the National Institutes of Health and were approved by the Osaka University Animal Care and Use Committee.

**Surgical procedure and animal maintenance**

Forty-four adult cats (1.5–4 kg) were prepared for electrophysiological recording as follows. First, subcutaneous injection was given of atropine sulfate (0.017 mg/kg) and hydroxyzine hydrochloride (Atarax, 0.83 mg/kg). Anesthesia was induced and maintained during surgery with isoflurane (2.5–3.5% in O2). Cefotiam hydrochloride (Panospol, 2.8 mg/kg) and dexamethasone sodium phosphate (Decadron, 0.13 mg/kg) were administered. Electrocardiogram (ECG) electrodes and a rectal temperature probe were installed. The rectal temperature probe was coated with lidocaine ointment. ECG and core temperature were monitored using a custom-built PC-based physiological monitoring system. Catheters were inserted into femoral veins of two limbs for infusion of drugs and fluids. A glass tracheal cannula was inserted after tracheostomy. A stereotoxic apparatus was used to secure position the animal’s head. Lidocaine ointment was used at pressure points of ear bars. After securing the animal to the stereotoxic apparatus, anesthesia was switched to thiopental sodium (Ravonal, administered continuously in infusion, 1.0–1.5 mg · kg⁻¹ · h⁻¹). Then, paralysis was induced with an initial dose of gallamine triethiodide and the animal was placed under artificial respiration at the rate of 20–30 strokes/min. The respiration rate and stroke volume were adjusted to maintain the end-tidal CO2 between 3.5 and 4.3%. A CO2 sensor (Datex-Ohmeda) was used to maintain a proper level of respiration. Anesthesia for the rest of recording session was maintained by a combination of 70% N2O-30% O2 and thiopental sodium as noted above. Paralysis was maintained by continuous infusion of Ravonal, gallamine triethiodide (10 mg · kg⁻¹ · h⁻¹) in lactated Ringer solution containing 50% glucose (40 mg · kg⁻¹ · h⁻¹). Body temperature was maintained near 38.3°C with the use of a servo-controlled heating pad. After securing the animal, a craniotomy was performed to access the central representation of the visual area 17 or 18 (Horsley–Clarke P4 L2.5 for recordings of A17, A3 L3 for A18). The dura was carefully removed to allow insertion of microelectrodes. Pupils were dilated with atropine (1%), and nictitating membranes were retracted with phenylephrine hydrochloride (Neoeyesin, 5%). Contact lenses of appropriate power with 4-mm artificial pupil were placed over the corneas.

The area of recording was primarily determined by the coordinate of electrode penetrations, although histological confirmation of recorded areas was conducted for the majority of animals. There is a possibility that a small fraction of neurons, especially from long penetrations, may be classified into a wrong cortical area. However, we did not eliminate those neurons (for which we were not completely certain of the area) from our analyses because they still represent important and valid samples for purposes of this study and there were no obvious areal differences.
Experimental apparatus

Tungsten microelectrodes (5 MΩ, A-M Systems) were used to record spike activities extracellularly. To increase the chance of encountering cells, two electrodes were mounted in parallel in a protective single guide tube and driven by a common microelectrode drive (Narisri). After confirming under a microscope that the electrodes do not penetrate blood vessels on the cortical surface, agar in warm Ringer solution was applied to stabilize and protect the cortex. Then, melted wax was applied over the agar to form a sealed chamber. An oscilloscope and audio speakers were used to monitor raw signals from the microelectrodes. Electrical signals from the microelectrodes were amplified (10,000×) and band-pass filtered (300–5,000 Hz). Then spike sorting was achieved using a custom-built spike sorter (Ohzawa et al. 1996), where each spike was sorted by their waveforms and time stamped with 40-μs resolution.

Visual stimulation and recording procedures

Experiment control functions and generations of visual stimuli were performed using custom-built software. Visual stimuli are generated by a dedicated PC and displayed on a CRT display (Sony GDM-FW900, a resolution of 1,600 × 1,024 pixels, covering the display area of 46.6 × 29.9 cm, 34.3 dot/deg; refresh rate: 76 Hz). A custom-built mirror haploscope was used to present stimuli to left and right eyes separately (Fig. 2A). To preclude projection of stimulus to contralateral eye, a separator was placed between the left and right visual fields. Distance between the screen and the eyes was set to 57 cm, subtending the visual field of 23.3(horizontal) × 29.9(vertical) degrees for each eye. Because we were examining interocular differences in neuronal responses, we carefully set up the haploscope and adjusted distances to the screen to equate the viewing conditions for two eyes as much as possible. The display surface of CRT monitor was carefully set perpendicular to the lines of sights for the subject.

After isolation of spike waveforms from one or more cells, approximate receptive field locations, preferred orientations, and spatial frequencies were determined manually by a mouse-controlled search program. Then, a standard reverse correlation procedure (DeAngelis et al. 1993a,b; Jones and Palmer 1987; Jones et al. 1987) was performed to obtain the accurate position and size of receptive field of two eyes. Subspace reverse correlation (Ringach et al. 1997) was then conducted for the dominant eye to obtain the preferred orientation and spatial frequency. Peaks of orientation and spatial frequency tuning correspond well to the peak values obtained by tests using drifting sinusoidal gratings (Nishimoto et al. 2005). After these preliminary tests, a binocular receptive field map was measured by a binocular reverse correlation procedure (Ohzawa et al. 1990, 1997). To compare interocular spatial frequency difference and tilt of binocular receptive field, spatial frequency, and orientation tests using drifting sinusoidal grating were performed for both left and right eyes, while using optimal values for other stimulus parameters for each cell. However, when using grating stimuli, contrast and temporal frequency were generally set to 50% and 2 Hz, respectively.

Binocular receptive field mapping

Binocular reverse correlation procedure was equivalent to that used by Ohzawa et al. (1997). A pair of one-dimensional bar stimuli was simultaneously presented to left and right eyes by a mirror haploscope setup (Fig. 2A). Twenty stimulus locations of the bar were used to stimulate receptive fields for each eye. This defined 20 × 20-point stimulus grid in the (X_L, X_R) domain (Fig. 2B). Therefore the binocular receptive field was measured by tallying up responses to 1,600 (20 × 20 × 4) different dichoptic pairs of stimuli. The orientation of the bar stimuli was set to the preferred orientation for each eye and for each cell. All possible combinations of left and right eye stimulus positions were included for each left–right permutation of contrast sign (dark–dark, bright–bright, dark–bright, bright–dark). All pairs of positions and combinations of stimulus contrast were presented in a random order, each stimulus lasting for 26 ms (two video frames) or 53 ms (four video frames) without any blank stimulus. Stimulus sequence was reshuffled for each set. A complete stimulus sequence lasted 42 s. Typically 20–40 sequences were used, which took 20–40 min in all. The response map for each contrast subset was calculated by cross-correlating spike trains with stimulus sequences (Fig. 2B). Binocular receptive field is a sum of response maps for matched polarity (bright–bright and dark–dark) conditions minus those for mismatched polarity (dark–bright and bright–dark) conditions. Monocular responses are cancelled by this computation and do not appear in the binocular RF (Ohzawa et al. 1997). We calculated the binocular RF for correlation delays from −100 to +300 ms in 5-ms steps. Because there is no correlation between spike train and stimulus sequence for negative time delays, we defined the response at negative time delays as noise. To obtain the optimal correlation delay, the sum of squared value of all data points in the RF at each correlation delay was obtained for the range of delays, and the peak delay was determined. A binocular receptive field is constructed at this optimal correlation delay. To evaluate the signal-to-noise ratio, we calculated the SD of the response at the optimal correlation delay divided by the average SD for negative correlation delays (−100 to −5 ms in 5-ms steps). We rejected data when the total spikes are <1,000 impulses and the peak response at the optimal delay did not exceed the mean of response at negative correlation delays +10SD.

Spatial frequency-tuning test

Left and right spatial frequency tunings were obtained by using drifting sinusoidal gratings in a separate test. Orientations of grating stimuli were
fixed at the optimal value for each eye because preferred orientations were typically different by 5–15° for the two eyes, probably arising from cyclorotation of the eye after paralysis (Nelson et al. 1977). The gratings were presented in a random order and each presentation lasted for 4 s interspersed with 1 s of interstimulus intervals. Mean firing rates were calculated at each spatial frequency. One-dimensional Gaussian functions were fitted to each spatial frequency tuning. Preferred spatial frequencies were obtained by the peak position of the fitted Gaussian function.

RESULTS

Binocular tests were conducted for a total of 271 neurons that were recorded from both areas 17 and 18 of 44 cats. Of these, binocular RFs could be obtained with sufficient signal-to-noise ratio (see METHODS) for 177 neurons. These neurons are further classified into two groups. Sixty-four neurons are classified as separable type and 113 neurons are classified as inseparable.

Binocular RFs for representative examples from the separable and inseparable types are illustrated in Fig. 3. Figure 3A depicts a binocular RF for a simple cell recorded in area 18. The binocular RF appears to be described reasonably well by a product of left and right monocular receptive field profiles for simple cells, as reported by Anzai et al. (1999a). Correlation between the standard simple/complex RF types and separability of binocular RF is high, but these classifications are not identical. This issue, including the basis for our choice of using the separability, will be described later (see following text). An exemplar complex cell recorded in area 18 is illustrated in Fig. 3C. The binocular RF showed a horizontally elongated structure like that in previous studies (Anzai et al. 1999b; Ohzawa et al. 1990, 1997). Complex cells tend to exhibit binocular RFs that are not left–right separable. Such inseparable receptive fields are well described by a disparity energy model where the sum of output of quadrature pairs of separable RFs constructs an inseparable RF (Anzai et al. 1999a,b; Ohzawa et al. 1990, 1997). On closer examination of this binocular RF, we noticed a small amount of tilt in the binocular RF from the frontoparallel axis in the clockwise direction. We wished to determine whether these small tilts are reliable properties of the neurons or arise from experimental noise or variability. Note that a small degree of tilt in the $(X_L, X_R)$ domain translates into a substantially larger surface slant in real object space in front of the animal. This is because, under realistic viewing conditions, the lines of sight from the two eyes to a fixation point crosses with a much more acute angle than the 90° angle for the $(X_L, X_R)$ domain. For example, given a viewing distance of 57 cm and interpupillary distance of 3 cm, a tilt of 5° in the $(X_L, X_R)$ domain is equal to the surface slant that is 73.3° from the frontoparallel plane (see APPENDIX). Therefore even a small visible tilt in the $(X_L, X_R)$ domain may have a large perceptual significance.

To estimate quantitatively the tilt of binocular RFs, we analyzed binocular RFs in the frequency domain. Frequency analysis is highly effective for evaluating the orientation of binocular RF without regard to specific local features of the RF, phase, or position. It also uses the entire set of RF data. Representative Fourier spectra of binocular RFs are shown in the right column. Figure 3, B and D shows Fourier spectra of the binocular RFs shown in Fig. 3, A and C, respectively. The axes (along oblique edges) of the domain are now left and right frequencies. The spectrum for separable binocular RF (Fig. 3B) has four peaks, whereas that for the inseparable neuron (Fig. 3D) shows a pair of strong peaks.

**FIG. 3.** Binocular receptive fields and their Fourier spectra are shown for simple (A, B) and complex (C, D) cells, respectively. Binocular RF of simple cells tended to be separable in the $(X_L, X_R)$ domain with 4 peaks in the spectrum, whereas those for complex cells tended to be inseparable and with 2 peaks in the frequency domain. There is a small but apparent tilt ($\theta$) of the binocular RF in C.
Alternatively, the same frequency domain may be referenced by a pair of orthogonal axes, along the vertical and horizontal directions corresponding to the diagonals of the square domain (Fig. 3B). These dimensions are defined as the disparity frequency and the frontoparallel frequency for vertical and horizontal axes, respectively (see Appendix). Interestingly, the four quadrants of the domain may be assigned to either disparity frequency tuning or frontoparallel frequency tuning. Top and bottom quadrants represent tuning for disparity, as indicated by two spectral peaks in Fig. 3D. The locations of the peaks in these domains allow extraction of such parameters as the optimal disparity frequency and binocular RF tilt. Left and right quadrants, on the other hand, will have substantial peaks only for separable neurons, and represent spatial frequency tuning of combined input from the two eyes. Therefore the peaks in these quadrants define the optimal frontoparallel frequency.

The process of determining binocular RF parameters in the frequency domain is illustrated further in Fig. 4. Based on the observation that substantial peaks are present in the left and right quadrants only for separable RFs, we define an index of separability of receptive field in the \( X_L \times X_R \) domain, the binocular separability index (BSI), as follows

\[
\text{BSI} = 1 - \frac{R_D - R_F}{R_D + R_F}
\]

where \( R_D \) is the peak response amplitude in the bottom quadrant, \( R_F \) is the response in the left quadrant along the cross section parallel to the left frequency axis going through the peak in the bottom quadrant, and taken at the same right frequency (inset of Fig. 4A). The left rather than the right quadrant is selected arbitrarily because the profiles in the left and right quadrants are symmetrical about the origin. The value of BSI ranges from 0 to 1. Based on the disparity energy model, simple cells will exhibit high BSI and complex cells will show BSI close to 0. Therefore neurons with BSI >0.73 are defined as the separable type, and otherwise, the inseparable type. The cutoff criterion for the BSI (0.73) gave the most consistent agreement with our visual inspection; neurons that have BSI values >0.73 have visually separable profiles for binocular RF and vice versa.

Using the same spectral profile of the binocular RF as above, the “tilt” of the binocular RF (\( \theta \)) may be defined as the angular deviation of the spectral peak from the disparity frequency axis, connecting the top and bottom corners of Fig. 4C. If the spectral peak is exactly on the disparity frequency axis, original binocular RF has zero tilt. A nonzero \( \theta \) indicates a corresponding tilt of binocular RF.

To estimate peak frequencies with greater accuracy, we interpolated the Fourier spectrum by cubic spline before evaluating the binocular RF tilt. A spatial frequency step of 0.005 (cycles/deg) is used as the resolution of interpolation for all neurons. To determine the step size for interpolation, we calculated percentage errors for binocular RF tilts for various resolutions of interpolation by simulations. Fourier transforms are performed on simulated binocular RF data obtained from model binocular complex cells with various interocular spatial frequency ratios (\( f_{\text{disparity}} / f_R = 0.66 \) to 1.5), and various disparity frequencies (\( f_{\text{disparity}} = 0.07 \) to 0.5; see Appendix). The data array is set to the same size as that in our experiments (20 x 20 grid). Then, interpolations are tested for various final resolutions (0.005 to 0.1 cycle/deg). On average, a sufficiently small error level (0.96 ± 0.03% error) for binocular RF tilt is obtained with the interpolation resolution of 0.005 cycle/deg. The percentage error increased to 13.43 ± 0.17% at the 0.1 cycle/deg resolution.

Because there is always a spectral peak in the bottom quadrant regardless of binocular RF separability, the calcula-
tions outlined above are applicable both to separable and to inseparable type of neurons. Note that cross sections going through the spectral peak that are parallel to the left and right frequency axes depict monocular spatial frequency-tuning curves, as estimated from the binocular RF data. These tuning curves are illustrated at the bottom left and right insets of Fig. 4C. The “tilt” angle of the binocular RF (θ) may be determined from the peak coordinate of the binocular RF (f_{0L}, f_{0R}), as follows

$$\tan (45 + \theta) = \frac{f_{0L}}{f_{0R}}$$

(2)

The line that goes through the spectral peak and the origin is defined as the cardinal disparity axis for the neuron.

Binocular RF tilt θ is transformed into disparity gradient, which is more commonly used to quantify slants of oriented surfaces in 3D. Disparity gradient represents surface slant independent of viewing distance. It is usually defined as

$$\Delta d = \frac{d_A - d_B}{\gamma}$$

(3)

where \(d_A\) and \(d_B\) are binocular disparities for two observed objects and \(\gamma\) is the angular separation between the directions for the two objects as viewed from the cyclopean eye, i.e., the midpoint between the two eyes (Burt and Julesz 1980). Therefore, a slant in actual space can be represented as disparity gradient, which may take on a value between \(-2.0\) and \(2.0\). Disparity gradients at these limiting values indicate the cases where two objects lie on a common line of sight for one eye. It was reported that absolute value of disparity gradient for two dots must be \(<1\)–\(2\) for binocular fusion depending on exact dot parameters (Burt and Julesz 1980; Prazdny 1985; Trivedi and Lloyd 1985). For this reason, we would expect most neurons to be encoding disparity gradient within these limits, if neural encoding of surface slants is constructed in an efficient manner. Note that the disparity gradient in Eq. 3 defines a property of the stimulus configuration. What we wish to estimate here instead is a property of a binocular neuron, i.e., its preferred disparity gradient given the cell’s binocular RF profile. This may be obtained from the binocular RF tilt θ, as described in the following equation

$$\Delta d = 2 \times \frac{1 - \tan (45 - \theta)}{1 + \tan (45 - \theta)}$$

(4)

To intuitively grasp the relationship between these metrics of surface slants, consider the following realistic example. When the binocular RF tilt θ is \(10^\circ\), the disparity gradient is 0.35, which corresponds to about \(80^\circ\) of physical surface slant at 57 cm of viewing distance. Using the disparity gradient as defined above, we will quantify and summarize RF slant for all neurons below.

As illustrated in Figs. 3 and 4, simple and complex cells tended to show different binocular RF profiles, binocularly separable and inseparable, respectively. However, simple/complex and separable/inseparable classifications are not the same. There are simple cells that are classified as inseparable, and vice versa. For the reasons outlined below, we will use the separable/inseparable type classification throughout the paper. However, before we set out to perform all the analyses based on this classification, we should examine the correlation between the two classification methods.

Note that an ideal complex cell based exactly on the disparity energy model will have a BSI of exactly 0 (Anzai et al. 1999b; Ohzawa et al. 1990, 1997). On the other hand, ideal binocular simple cells that linearly sum left and right eye input will have a BSI of 1 (Anzai et al. 1999a; Ohzawa et al. 1990, 1996). The actual population of neurons we have recorded exhibited substantial deviations from the ideal cases as illustrated in Fig. 5.

First, for the simple/complex classification, we use the standard criteria based on the F1/F0 ratio, the ratio of the amplitude modulation (AM) in response to an optimal drifting sinusoidal grating stimulus to the average firing rate for the same response (Li et al. 2003; Skottun et al. 1991). Relationships between left and right F1/F0 ratios are plotted in Fig. 5A. The ratios were evaluated at the optimal spatial frequency for each eye. Circle and triangle symbols indicate data recorded from areas 17 and 18, respectively. The correlation of F1/F0 ratios for the left and right eyes is highly significant (for area 17, \(r = 0.9, P < 0.001, N = 66\); for area 18, \(r = 0.83, P < 0.01\)). However, for some neurons, classified types were mismatched between the eyes. B: scatterplot is shown of F1/F0 ratio vs. binocular separability index (BSI). These 2 parameters showed a significant correlation (Pearson’s \(r = 0.76\) and 0.78 for left and right eyes, respectively, \(P < 0.001, n = 135\)). F1/F0 ratios were obtained from responses to drifting sinusoidal gratings of optimal spatial frequency for each eye. Open and filled symbols depict data from left and right eyes, respectively. Therefore each cell has 2 symbols for F1/F0 ratios, connected by a line segment for indicating paired data. C: F1/F0 ratios show a binodal distribution. Filled and open bars indicate right and left eyes, respectively. D: distribution of BSI is shown. Majority of neurons have inseparable binocular RF (as indicated by BSI <0.73), many of which are classified as simple based on the F1/F0 ratio.
However, there are several neurons with a large mismatch in the F1/F0 ratios between the eyes. That is, some neurons had highly modulated responses to sinusoidal drifting gratings for one eye, but practically no modulation was observed for the opposite eye. The relationship between F1/F0 ratio and BSI is illustrated in Fig. 5B. Open and filled symbols depict data for the left and right eyes, respectively. Each cell has two symbols for F1/F0 ratios (for the two eyes), connected by a line segment for indicating paired data. Although these two parameters show significant correlations ($r = 0.76$ and $0.78$ for left and right, respectively; $P < 0.001$, $n = 135$), there are many cases where the predictions of ideal model cases break down. For example, neurons with BSI values close to zero had a wide variety of F1/F0 ratios, indicating that binocularly inseparable RFs may be observed commonly in both simple and complex cells. Figure 5, C and D indicates the distributions for left and right F1/F0 ratios and the distribution of BSI, respectively. Filled and open bars in Fig. 5C indicate data for the right and left eyes, respectively. The F1/F0 ratios show a bimodal distribution as reported previously (Li et al. 2003; Mechler and Ringach 2002). Note also that BSI is derived directly from the data from a key binocular measurement in this study, whereas F1/F0 ratios are obtained from monocular tests and therefore are expected to be less directly linked to binocular properties. There have also been questions about a multitude of factors that influence F1/F0 ratios (Mata and Ringach 2005). Considering further that the use of the classical criteria in simple/complex classification can sometimes result in discrepant types between the eyes, the use of binocular separability of the RF offers a better classification method overall for the purposes of this study.

Recall that one of the purposes of this study is to examine whether the apparent tilt of binocular RF profile is based on the difference in the optimal spatial frequencies across the eyes (Fig. 1). The question is addressed in the next several figures based on results of binocular RF and spatial frequency-tuning measurements from both binocularly separable and inseparable neurons. Data from representative examples of binocularly separable neurons are illustrated in Fig. 6. Binocular RF profiles are shown in the left column. In the middle column, monocular Fourier spectra derived from the binocular RF are shown as solid and dashed curves for the left and right eyes, respectively. These are cross sections through the peak of the Fourier spectrum as illustrated in Fig. 4C, taken parallel to the left and right frequency axes. Actual spatial-frequency-tuning curves obtained by drifting sinusoidal grating stimuli are illustrated in the right column. The predicted tuning curves in the middle column and those in the right column should be comparable directly under certain linearity assumptions (DeAngelis et al. 1993a,b). Open and filled symbols depict responses for the left and right eyes, respectively. Error bars represent the SE. A horizontal dashed line indicates the spontaneous firing rate. Gaussian functions were fitted to the tuning curves. Vertical thin lines indicate the peaks of fitted Gaussian function for left and right eyes, respectively. Preferred spatial frequencies are significantly different between the eyes (bootstrap test, $P < 0.05$). Spatial frequency ratio is 0.71. B: data from another separable binocular RF is shown for a simple cell in the same format as A. Binocular RF is significantly tilted from frontoparallel plane (binocular RF tilt $= -6.3$; $P < 0.05$, bootstrap test). Predicted disparity gradient is $-0.22$. Spatial frequency ratio is 0.69. C: additional example of separable binocular RFs is shown (binocular RF tilt $= -8.5$). Predicted disparity gradients and spatial frequency ratio are $-0.3$ and 0.7, respectively.

FIG. 6. Examples are shown of 3 separable type neurons that have different frequency tunings across the eyes. A, left: separable binocular RF of a simple cell is depicted. A, middle: left and right Fourier spectra of the binocular RF are shown as solid and dashed curves, respectively. Although the tilt of binocular RF is small (binocular RF tilt $= -7.0$), it is statistically significant ($P < 0.05$, bootstrap test). Predicted disparity gradient is $-0.25$. A, right: spatial frequency-tuning curves obtained by drifting sinusoidal grating stimuli are illustrated. Open and filled symbols depict responses for the left and right eyes, respectively. Error bars depict SEs. Horizontal dashed line indicates the spontaneous firing rate. Gaussian functions were fitted to the tuning curves. Vertical thin lines indicate the peaks of fitted Gaussian function for left and right eyes, respectively. Preferred spatial frequencies are significantly different between the eyes (bootstrap test, $P < 0.05$). Spatial frequency ratio is 0.71. B: data from another separable binocular RF is shown for a simple cell in the same format as A. Binocular RF is significantly tilted from frontoparallel plane (binocular RF tilt $= -6.3$; $P < 0.05$, bootstrap test). Predicted disparity gradient is $-0.22$. Spatial frequency ratio is 0.69. C: additional example of separable binocular RFs is shown (binocular RF tilt $= -8.5$). Predicted disparity gradients and spatial frequency ratio are $-0.3$ and 0.7, respectively.

For the cell presented in Fig. 6A, tilt of the binocular RF (as measured by the displacement of spectral peaks illustrated in Fig. 4C) is statistically significant (tilt $= -7°$, $P < 0.05$, bootstrap test). The bootstrap test for estimating significance of tilt is conducted as follows. Binocular RF mapping consists of trials, each of which contains a randomized sequence of complete permutations of left and right stimuli (Ohzawa et al. 1990, 1997). From spike data for a total of $N$ (typically 40) trials, $N$ trials are randomly drawn while allowing duplications, from which a new binocular RF is constructed. For each neuron, this process was repeated 1,000 times to obtain the estimates of
variability in the RF measurements (Efron 1982; Efron and Tibshirani 1993). When the mean tilt of the distribution of resampled binocular RFs was deviated from zero by more than 1.96 SD, the RF tilt was judged to be significant. With this criterion, the probability of RF tilt being on the opposite side of zero is <5%.

The binocular RF tilt determined as above is automatically reflected as a difference in the predicted spatial frequency-tuning curves shown in Fig. 6A (middle). The predicted disparity gradient for this cell is 0.25, as calculated by Eq. 4. A similar statistically significant difference in the optimal spatial frequencies for the two eyes is also observed for the actual tuning curves measured by drifting sinusoidal gratings (Fig. 6A, right; P < 0.05, bootstrap test) in that the optimal spatial frequency for the right eye (vertical dashed line) is higher than that for the left eye (vertical solid line). Therefore for this neuron, there is a good correspondence between the tilt of the binocular RF (measured by reverse correlation) and the interocular difference between the optimal spatial frequencies (measured by drifting gratings).

Similar additional data from two separable binocular RFs are shown in Fig. 6, B and C in the same format as that of Fig. 6A. For these two cells (both of which were simple), tilt angles $\theta$ of binocular RFs were significantly different from zero ($P < 0.05$, bootstrap test). The tilt angles of binocular RF for Fig. 6, B and C are $-0.3$ and $-8.5^\circ$, which correspond to predicted preferred disparity gradients of $-0.22$ and $-0.3$, respectively. Again, for these additional cells, the actual spatial frequency-tuning curves measured by drifting gratings (right column) also show statistically significant difference between the eyes ($P < 0.05$, bootstrap test). The ratios of optimal spatial frequencies (left/right) are 0.71, 0.69, and 0.70 for cells in Fig. 6, A–C, respectively. Again, the direction of the difference in predicted spatial frequency-tuning profiles (middle column) corresponds well to that for the measured data (right column) for each neuron. Therefore these results for binocularly separable neurons indicate that the tilt angles of their binocular RFs and their predicted disparity gradients correspond well with the left–right differences of optimal spatial frequencies measured by monocularly presented drifting gratings.

Data from representative examples of inseparable neurons are illustrated in Fig. 7 in the same format as that of Fig. 6. Spatial frequency-tuning curves are not available for $B$ and $D$ either because spikes for one of the cells appeared after the initial tuning tests were already completed or, for the case of $D$, data for the frequency-tuning test did not show significantly modulated responses as a function of spatial frequency ($P > 0.05$, ANOVA). Binocular RFs shown in Fig. 7, pairs $A$ and $B$, $C$ and $D$ are from neurons that were recorded simultaneously. The neuron shown in $A$ is an example for which the binocular RF was significantly tilted from the frontoparallel plane ($P < 0.05$, bootstrap test). In fact, all of the examples except for that in Fig. 7C had statistically significant tilt for their binocular RFs. Note that the neuron illustrated in $B$ had a statistically significant tilt in the opposite direction from the other member of the pair shown in $A$. The opposite tilt directions for the pairs of neurons clearly indicate that the tilts of binocular RFs do not arise from optical factors such as errors in the eye-display distances or magnification differences between the eyes. Because there are significant differences in the degree of tilt among simultaneously recorded neurons, these variations must be neural in origin.

Another pair of simultaneously recorded neurons also exhibited a clear difference in the tilts of binocular RFs. Although the neuron depicted in Fig. 7C did not have a significant RF tilt, the other member of the pair had its RF tilted significantly from
the frontoparallel plane. Tilt angles for $D$, $E$, and $F$ are 10.2, $-3.8$, and $6.8^\circ$, which correspond to 0.36, $-0.13$, and 0.24 as disparity gradients, respectively.

As with binocularly separable neurons presented in Fig. 6, independent measurements of spatial frequency-tuning curves are also conducted using drifting sinusoidal gratings. Preferred spatial frequencies, shown as vertical solid and dashed thin lines in the right column, differ significantly across the eyes for Fig. 7, $A$, $E$, and $F$ ($P < 0.05$, bootstrap test), but not for Fig. 7C ($P > 0.05$, bootstrap test). This is consistent with the lack of significant tilt of binocular RF for this neuron. Therefore for all cases shown in Fig. 7, $A$, $C$, $E$, and $F$, directions of interocular spatial frequency difference correspond well to the frequency difference of binocular RFs.

Paired recordings are also possible between neurons of different binocular separability. Such an example is shown in Fig. 8. Binocular RFs shown in Fig. 8, $A$ and $B$ are separable and inseparable RF, respectively. For both neurons, binocular RFs exhibit significant tilts from the frontoparallel plane ($P < 0.05$, bootstrap test). Furthermore, the tilts are in opposite directions between the two neurons. The tilt angles for cells in Fig. 8, $A$ and $B$ are 3.2 and $-3.7^\circ$, with the corresponding preferred disparity gradients of 0.11 and $-0.13$, respectively. Actual spatial frequency-tuning curves measured with drifting gratings are shown in the right column. As expected from the binocular RF tilts, the interocular difference in the preferred spatial frequencies are opposite for the two neurons. The left preferred spatial frequency is significantly higher than the right frequency (frequency ratio = 1.23, $P < 0.05$, bootstrap test) for $A$; the difference is significant and opposite (frequency ratio = 0.87, $P < 0.05$, bootstrap test) for $B$. Taken together with the results from the previous figure, both separable and inseparable binocular RFs show a variety of tilts that are consistent with the interocular difference in the monocularly measured preferred spatial frequencies. Therefore the notion of the basis of 3D surface tilt representation, as illustrated in Fig. 1, appears quite likely based on these examples.

What is the range of binocular RF tilts observed for cells in areas 17 and 18? Distributions of disparity gradients of both separable and inseparable cells are illustrated in Fig. 9. Most neurons had disparity gradients in the range of $-0.5$ to 0.5. The SDs of the mean were 0.19 and 0.14 for separable and inseparable types, respectively. Black bars indicate cells whose binocular RFs are tilted significantly from the frontoparallel plane, whereas white bars indicate those with nonsignificant tilt ($P < 0.05$, bootstrap test). About 30% of neurons exhibited significant tilts (28%, $18/64$ for separable; 33%, $37/113$ for inseparable). Therefore the distributions of disparity gradients...
in areas 17 and 18 are capable of supporting slant-in-depth encoding.

Although paired recordings of multiple neurons are ideal for demonstrating variations of binocular RF tilts (see Figs. 7 and 8), such recordings are not always possible. The majority of the data in our sample must be analyzed as individual binocular RF recordings. Therefore we have analyzed the effects of potential artifactual sources that may contribute to apparent tilts of measured binocular RF profiles. One such possibility is a difference in viewing distances between left and right eyes that may be caused by positioning errors of the CRT monitor and the mirrors used in the haploscope setup (Fig. 2A). Another possibility is a magnification difference between left and right eyes that may result from improper corrections for refractive errors. Both of these optical factors produce apparent differences in the spatial frequency content as imaged on the retina for the two eyes. Contributions of viewing distance errors to disparity gradients are illustrated in Fig. 10. We are confident that our positioning error of optical elements in the setup is well within 5 cm. Given this assumption, what is the limit of erroneous change in the disparity gradient? Figure 10A shows that a 5-cm distance error translates into a disparity gradient of about 0.1. Distributions of disparity gradients for both separable and inseparable binocular RFs, and that of viewing distance errors, are illustrated in Fig. 10B. The SD (σ) of the error distribution is set such that 1.96σ = 0.1. Statistical tests for data and artifactual distributions are carried out by the F-test. Distribution of preferred disparity gradients is significantly wider than that of the error distribution (test for equal variance, \( F = 13.4, P < 0.001 \) for separable type; \( F = 7.42, P < 0.001 \) for inseparable type). Similarly, we also calculated possible contributions of interocular magnification differences.

A 3% magnification difference between the two eyes (assuming the power of the cat’s eye of 78D, and |error in refractive correction in diopters| < 2D) will result in a disparity gradient of ±0.03 (Hughes 1979). The SD for this distribution is so small that we can essentially ignore the effect of refractive errors. Even considering the simultaneous contributions of the two factors, the variations of binocular RF tilts observed in our data cannot be accounted for by these artifactual sources (test for equal variance, \( F = 6.7, P < 0.001 \) for separable type; \( F = 3.71, P < 0.001 \) for inseparable type). These results suggest that the tilts of binocular RFs and spatial frequency differences are intrinsic neuronal characteristics and are able to carry signals regarding 3D orientations of surfaces in visual scenes.

**Relationship between disparity gradient and spatial frequency ratio**

In representative examples shown in Figs. 6–8, the spatial frequency differences across the eyes were generally qualitatively correlated with the tilt of binocular RFs. How does this correlation hold for the entire population of neurons? In general, how do other binocular tuning characteristics correlate with monocular tuning properties? Figure 11 summarizes the results relevant for addressing these questions.

First, left and right preferred spatial frequencies, as obtained from tests using drifting sinusoidal gratings, are compared in Fig. 11, A and B for separable and inseparable binocular RF, respectively. Peak spatial frequencies are obtained from the peak of fitted Gaussian functions (Eq. 5). The identity relationship and 1-octave difference between the eyes are illustrated as solid and dotted lines, respectively. Circles and triangles indicate cells recorded from areas 17 and 18, respectively. Preferred spatial frequencies for the left and right eyes are well correlated (Pearson’s \( r = 0.96, n = 45 \) for separable type; \( r = 0.96, n = 90 \) for inseparable type, \( P < 0.05 \)). Differences of left and right spatial frequencies were within the range of +1 to −1 octave regardless of separability.

To examine the correlation between the interocular frequency difference and the tilt of binocular RF, the ratios of preferred spatial frequencies were computed as follows and compared with the disparity gradients. The frequency ratio is given by

\[
\text{Frequency Ratio} = \frac{f_L}{f_R}
\]

where \( f_L \) and \( f_R \) are left and right preferred spatial frequencies from measurements with drifting gratings. The results of comparisons are illustrated in Fig. 11, C and D for separable and inseparable cells, respectively. Cells recorded from areas 17 and 18 are plotted as circles and triangles, respectively. Black and gray symbols indicate cells that exhibited significant and

\[
F = 13.4, P < 0.001 \quad \text{separable:} \quad F = 7.42, P < 0.001
\]
Figure 11. Goodness of predictions of binocular RF tilts (expressed as disparity gradients) based on the ratios of left and right optimal spatial frequencies are examined. A and B: illustration of correlations of left and right optimal spatial frequencies for separable and inseparable binocular RFs, respectively. Optimal frequencies are obtained from the peak of fitted Gaussian functions. Dotted lines indicate 1 octave difference in the optimal frequencies for the 2 eyes. C and D: disparity gradients from separable and inseparable binocular RFs are plotted against the frequency ratios obtained from A and B. Frequency ratios are defined as \( f_L/f_R \) where \( f_L \) and \( f_R \), respectively, are left and right optimal spatial frequencies obtained from drifting sinusoidal grating tests. For inseparable neurons shown in D, there is a significant correlation (all data: \( r = 0.26, n = 90, P < 0.05, \) Spearman’s correlation coefficient). Correlation analysis on the subset of neurons that exhibited significant tilt (see Fig. 9B) showed somewhat improved correlation coefficient (\( r = 0.5, n = 29, P < 0.01 \)). Correlation for separable type was not significant. Black and gray symbols indicate data from neurons that exhibited significant and nonsignificant tilts of binocular RFs. Circles and triangles depict cells recorded from areas 17 and 18, respectively. Solid line represents prediction of the dif-frequency disparity energy model. There are neurons with a clear and statistically significant interocular spatial frequency difference, and yet possess clearly frontoparallel binocular RF, and vice versa. Therefore in sections further below, we will examine possibilities of additional factors that may contribute to tilts of binocular RFs.

An additional point was examined in relation to predicting binocular properties from monocular tuning characteristics. Figure 11, E and F shows the relationship between binocular disparity frequency and monocular preferred spatial frequency. The disparity energy model predicts identity between the two frequencies. Regarding this question, Ohzawa et al. (1997) reported the discrepancy between the model prediction and the data. They reported that the disparity frequency tended to be lower than the monocular spatial frequency as measured by drifting gratings stimuli. Because their analyses were performed only for complex cells, it is not clear at which stage of binocular processing this discrepancy occurs. Based on a new set of data and a more robust analysis method, we have addressed this issue. In our analysis, we use Fourier analysis both for separable and inseparable RFs to obtain disparity frequencies. For the monocular preferred spatial frequency, the average of left and right preferred spatial frequencies (from data in Fig. 11, A and B) are used. Cells recorded from areas 17 and 18 are plotted as circles and triangles. The scatterplot for the inseparable binocular RFs showed a discrepancy between the disparity frequency and the spatial frequency, in that the disparity frequency tends to be lower than the monocularly measured preferred spatial frequency (Fig. 11F). Deviations of actual data from the identity line tended to be larger for high spatial frequencies (slope = 0.71). Because most inseparable binocular RFs are from complex cells (Fig. 5B), our data show a trend similar to that reported in previous work (Ohzawa et al. 1997). In contrast, separable binocular RFs show a much better fit with the identity relationship between the two frequencies. The slope of separable RF is close to 1 (slope = 0.92) (Fig. 11E). These results probably suggest that separable cells sum monocular inputs through linear processing, whereas neurons with inseparable RFs have substantial nonlinearities in their processing. The source of the deviation must therefore lie...
between the linear subunits of complex cells and the final complex cell stage if we assume the hierarchical organization similar to that in the disparity energy model.

Aspect ratio of binocular receptive field

Although the dif-frequency version of the disparity energy model accounts for the trend in the data as we have seen in the previous section, we wondered whether there are additional mechanisms by which tilted binocular RFs are constructed. Another possibility we examine is a hierarchical organization as illustrated in Fig. 12. A tilt in the binocular RF profile may be generated if the outputs of multiple disparity energy units are combined, where each unit is tuned to a specific disparity without tilt and its preferred disparity progressively shifts as a function of its frontoparallel position (Fig. 12A). Such a hierarchical pooling produces a binocular RF, shown in Fig. 12B. This neuron (Fig. 12B) will have a highly elongated and tilted binocular RF. The angle of tilt depends on the rate at which subunits’ preferred disparities shift with the frontoparallel position. Such an organization predicts a substantial elongation of the binocular RF in the frontoparallel dimension. The degree of pooling may be quantified by an aspect ratio of binocular RF. If the hierarchical organization underlies slant sensitivity of binocular neurons, there should be a correlation between the tilts of binocular RF and their aspect ratios.

To obtain structural parameters of binocular RFs such as the RF sizes and aspect ratios, we conducted frequency analysis (Fig. 12C). Although RF sizes may be obtained by direct measurements in the spatial domain (Fig. 12B), we have found that estimating RF size in the frequency domain (Fig. 12C) is more robust. The procedure for computing spectral data was described earlier (Fig. 4), except that a two-dimensional Gaussian function is fitted to the amplitude spectrum, and its SDs are used. Binocular RF sizes, 2a and 2b for frontoparallel and disparity directions, respectively, are calculated as the inverse of SDs of fitted spectral amplitude profiles

$$a = \frac{1}{\sigma_a}, \quad b = \frac{1}{\sigma_b}$$

where $\sigma_a$ and $\sigma_b$ are the SDs of the fitted Gaussian in the disparity and frontoparallel frequencies, respectively. The aspect ratio of a binocular RF is defined by the ratio of SDs as

$$\text{Aspect ratio} = \frac{a}{b}$$

The aspect ratio of $<1$ indicates that the envelope of binocular RF is elongated along the disparity axis. If it is $>1$, the envelope of binocular RF is elongated along the frontoparallel axis. Therefore if there are neurons with the hierarchical organization as illustrated in Fig. 12, A and B, we would expect aspect ratios of RFs for those neurons to be substantially $>1$. The disparity energy model with no such pooling predicts the aspect ratio equal to 1.

Distributions of aspect ratios for separable and inseparable RFs are shown in Fig. 13, A and B. For most neurons, aspect ratios were $>1$ for inseparable RFs (Fig. 13B). Mean aspect ratios are 1.15 and 1.67 for separable and inseparable RFs, respectively. The result for inseparable RFs, the majority of which are complex cells, indicates a substantial degree of spatial pooling, deviating substantially from prediction of the disparity energy model. The relationship between the aspect ratio and the disparity gradient is presented in Fig. 13, C and D. If the hierarchical organization hypothesis (Fig. 12A) is correct as a basis for slant selectivity, neurons with highly elongated receptive fields should possess a wide range of disparity gradients. In contrast, neurons with aspect ratios close to 1 should show a narrow distribution for disparity gradients near zero. However, our data show the opposite: Disparity gradients tended to be highly variable for neurons with low aspect ratios, but were relatively small for those with high aspect ratios for
inseparable RFs (separable: \( n = 45, P = 0.09 \), Mann–Whitney \( U \) test; inseparable: \( n = 90, P < 0.05 \), Mann–Whitney \( U \) test).

It may be argued that the increased range of disparity gradients for neurons with small aspect ratios may simply reflect poorer reliability for estimating tilts for these cells. For example, orientations of ellipses may be determined more reliably for highly elongated ellipses than for nearly circular ones, given a constant level of noise or measurement errors. To examine this factor, we show the relationship between the confidence (SD, i.e., the length of error bars in Fig. 13, \( C \) and \( D \)) for estimates of the disparity gradient and the aspect ratio in Fig. 13, \( E \) and \( F \). The scales of the vertical axes are equal for Fig. 13, \( C \) and \( D \) for comparison. Although there is a tendency for cells with smaller aspect ratios to have longer error bars (\( n = 18, P < 0.01 \), Pearson’s \( r = -0.77 \); black symbols in Fig. 13E), the error-bar length is much smaller than the mean value of disparity gradient. It is nearly a constant fraction (20%) of the absolute value of the disparity gradient for neurons with significant tilts (black symbols). Therefore it is probably accurate to say that the variability of disparity gradient is almost independent of the aspect ratio. Therefore these results indicate that highly tilted binocular RFs are not constructed by the spatial pooling process as shown in Fig. 12. Nevertheless, 21.6% (8/37) of significantly tilted inseparable cells have substantially elongated binocular RFs (aspect ratio \( \geq 2 \)), indicating that pooling may play at least some role in constructing slightly tilted binocular RFs.

The fact that the model of Fig. 12 is rejected should not be interpreted to mean that pooling is not important. Rather, it may play an important role in slant discrimination. It is possible that the role of pooling for constructing RF with a high aspect ratio is to create neurons that can signal near-zero surface slants with greater accuracy, allowing fine slant discriminations for surfaces near frontoparallel. Our results are certainly consistent with such a possibility.

Having defined the aspect ratio of binocular RF, we return to the question of the relationship between the interocular spatial frequency difference and binocular RF tilt. Do neurons with untitled binocular RF with a clear interocular spatial frequency difference have elongated RFs (with high aspect ratios)? These neurons cannot be explained by either of the models we have examined so far. However, one possibility we have not considered is the opposite of the model in Fig. 12 where the spatial pooling is performed over highly tilted subunits but along the exact frontoparallel direction. That is, although individual unpoled units possess tilted binocular RFs, the spatial pooling produces a counteracting effect, thereby canceling the tilts of pooled members. We therefore examined the aspect ratios of representative neurons of this type. Four neurons in the rightmost part of the scatterplot in Fig. 11D have been selected. These neurons have a frequency ratio \( >1.5 \). Three of the four neurons had large aspect ratios of 1.91-2.53, and one of them had an aspect ratio of 1.08. The results are not conclusive, but there is a tendency for these neurons to have highly elongated binocular RFs.

Does the aspect ratio relate to other parameters of binocular RF? Figure 14 illustrates relationships among the depth-domain aspect ratio, RF sizes, and preferred spatial frequencies. The relationships between the depth-domain aspect ratio and RF sizes are illustrated in Fig. 14, \( A \) and \( B \). Sizes of binocular RFs are defined both in the disparity and frontoparallel directions as illustrated in Fig. 12. There is no correlation between the aspect ratio and the RF size in the frontoparallel direction (Fig. 14A) (separable: \( r = 0.14, P > 0.05, n = 45 \); inseparable: \( r = -0.02, P > 0.05, n = 90 \)). In contrast, there is a significant negative correlation between the aspect ratio and the RF size in the disparity direction (Fig. 14B) (separable: \( r = -0.1, P > 0.05, n = 45 \); inseparable: \( r = -0.48, P < 0.001, n = 90 \)). These results indicate that RFs with high aspect ratios tended to have narrow absolute RF sizes in the depth dimension.

It is known that monocular RF sizes are inversely correlated with the preferred spatial frequency (DeAngelis et al. 1993b; De Valois et al. 1982). Because the RF size in the frontoparallel direction is essentially the average of monocular RF sizes, it is expected to show similar correlation with the preferred spatial frequency. How, then, is the RF size in the depth dimension related to the preferred spatial frequency? Figure 14, \( C \) and \( D \) depicts relationships between sizes of binocular RF and the preferred spatial frequency. As expected, RF sizes in the frontoparallel direction are inversely correlated with the preferred spatial frequency, as shown in Fig. 14C (\( C: r = \)).
is not uniform in the binocular domain. The pooling occurs more for neurons tuned to high spatial frequencies and tends to occur only along the frontoparallel direction but not in the disparity direction.

**Relationship between orientation and disparity gradient**

Orientation bias was found for encoding of binocular disparity in that neurons with dissimilar RF profiles (RF phases) between the two eyes tended to prefer near-vertical orientations (DeAngelis et al. 1991, 1995; Ohzawa et al. 1996). Is there a similar orientation bias for the neural representation of slant-in-depth? The relationship between the preferred orientation and disparity gradient is illustrated in Fig. 15. The preferred orientation was evaluated by the peak of fitted Gaussian function to the orientation tuning data measured by drifting sinusoidal gratings. The average orientation for the two eyes was used. Each preferred orientation is represented as an angle from the horizontal. Black and gray symbols depict binocular RFs tilted significantly and nonsignificantly from the frontoparallel plane. Circles and triangles indicate cells recorded from areas 17 and 18, respectively. If the slant-in-depth encoding depends on the preferred orientation of RFs, there should be a positive correlation between these parameters, although no correlations are observed between the two parameters (Pearson’s $r = 0.09, n = 168, P > 0.05$).

**Relationship between SF/DF ratio and RF structures**

As illustrated in Fig. 11F, there is a discrepancy between the preferred spatial frequency and the disparity frequency for inseparable RFs as originally reported by Ohzawa et al. (1997) and confirmed in the present study. That the discrepancy was found for inseparable binocular RFs, but not for separable ones, suggests that the discrepancy originates from a stage that pools the output of multiple simple-type subunits of complex disparity energy units. Moreover, nonlinearities in these pooling processes may be the source of the discrepancy. If this is
the case, there may be a correlation between the size of discrepancy and the degree of pooling quantified by the aspect ratio. To examine this, we compared the aspect ratio with the size of discrepancy, which is quantified by the ratio of the preferred spatial frequency to the disparity frequency (SF/DF ratio). Because the preferred spatial frequency tended to be higher than the disparity frequency (Fig. 11F), the SF/DF ratio is >1 for most neurons.

Figure 16A illustrates the relationship between the aspect ratio and the SF/DF ratio. Filled and open symbols depict separable and inseparable RFs, respectively. Circles and triangles indicate cells recorded from areas 17 and 18, respectively. There is a positive significant correlation between the two parameters for inseparable RFs but not for separable RFs (separable: \( r = 0.09, P > 0.05, n = 45 \); inseparable: \( r = 0.55, P < 0.001, n = 90 \)). Figure 16, B and C illustrates relationships between the RF size in the frontoparallel direction and the SF/DF ratio, and between RF size in the disparity direction and the SF/DF ratio, respectively. No significant correlations are observed from these scatters (\( B: r = 0.006, P > 0.05, n = 45 \) for separable; \( r = 0.08, P > 0.05, n = 90 \) for inseparable; \( C: r = -0.03, P > 0.05, n = 45 \) for separable; \( r = -0.2, P > 0.05, n = 90 \) for inseparable). The significant correlation between the SF/DF ratio and the aspect ratio suggests that the pooling process is responsible for the discrepancy between the monocular preferred spatial frequency and the binocular disparity frequency.

DISCUSSION

We approached the possible mechanisms for encoding of the 3D surface slant by analyzing characteristics of binocular RFs of neurons in early visual cortex. One of our key findings is that the neurons in the early visual cortex possess a variety of binocular RF tilts that are sufficient for encoding the range of physical surface slants that occur under normal viewing conditions. Therefore our results suggest a possibility that surface slant encoding as reported by previous studies in higher-order visual areas such as MT, CIP, V4, and IT (Hinkle and Connor 2002; Janssen et al. 1999, 2000; Liu et al. 2004; Nguyenkim and DeAngelis 2003; Taira et al. 2000) may originate at least partially in the early visual areas. In this section, we discuss potential problems in our experimental procedures and interpretations of our data, as well as relationships to previous findings in the literature.

Selectivity of neural responses for 3D surface slant was estimated by the tilt of binocular RF from the frontoparallel plane. Because the degrees of tilts are small, approximately 10° at most, we had to show that these tilts do not arise from noise, variabilities of experimental calibrations, or other nonneural factors. We have presented three pieces of evidence to establish that observed variations of binocular RF tilts are real and neural in origin. First, bootstrap tests were performed to show that measured binocular RFs possess significant nonzero tilt in the presence of neural response variability. Second, we have shown that pairs of neurons recorded simultaneously had significantly different RF tilts. If the RF tilts are attributable to optical factors such as errors in distance or magnification adjustments for the two eyes, the degree and direction of tilt should be similar for the pair of neurons. Our examples show tilts of binocular RFs that are in opposite directions between the pair of cells (Figs. 7 and 8). Third, we have determined that the contribution of artificial sources of errors is much smaller than the variance we observe in the data. Distributions of actual RF tilts are significantly wider than that for artifacts (Fig. 10). Based on these pieces of evidence, we conclude that there are true variations in the tilt of binocular RFs that are neural in origin.

The range of distribution of disparity gradients, which were converted from the tilts of binocular RF maps, is ±0.5 both for separable and inseparable binocular RFs (Fig. 9). These ranges are capable of representing actual surface slants in the real world of >70° assuming the 57-cm viewing distance. The range of disparity gradient representation we have found (for the cat) is also similar to that for neurons in area MT of the monkey (Nguyenkim and DeAngelis 2003).

Although the possible ranges of slant encoding are similar across visual areas, there is a critical difference between the slant representation in the early visual cortex and higher-order areas. Neurons in extrastriate areas are known to have selectivities to surface orientations that are invariant with respect to positional disparity (Hinkle and Connor 2002; Nguyenkim and DeAngelis 2003; Taira et al. 2000). Although we have not tested explicitly for the disparity invariance of slant-in-depth selectivity, it is clear that a model based on the tilted binocular RF by itself cannot be position-disparity invariant. Therefore neurons in areas 17 and 18 are likely to be highly sensitive to positional disparity, but they also carry additional information on surface slant because the RF model predicts the maximum

![Fig. 16. Relationships are shown of binocular RF characteristics and ratio of spatial frequency to disparity frequency (SF/DF ratio). Because the disparity energy model predicts SF/DF = 1, the ratio reflects discrepancy between the data and predictions by the model. A: relationship between aspect ratio and SF/DF ratio is illustrated. Significant correlation (separable: \( r = 0.09, P > 0.1, n = 45 \); inseparable: \( r = 0.55, P < 0.001, n = 90 \)) is observed for inseparable RFs. B and C: relationships between the SF/DF ratio and the binocular RF sizes in frontoparallel and disparity directions are illustrated, respectively. Circles and triangles indicate data for neurons recorded from areas 17 and 18, respectively. Open and filled symbols depict separable and inseparable RFs, respectively.](http://jn.physiology.org/cover ftp://ftp.jn.org)
firing for the neuron when both the disparity and the surface slant match the binocular RF.

Dif-frequency organization for slant-in-depth encoding

We have examined a modified disparity energy model based on the interocular spatial frequency difference (dif-frequency). When viewing a slanted 3D surface, the spatial frequency contents for the corresponding areas of the surface are different between the eyes (Blakemore 1970; Fiorentini and Maffei 1971; Tyler and Sutter 1979; Wilson 1976). It was known that for some neurons in early visual cortex, the preferred spatial frequencies for the two eyes were not always the same (Hammond and Pompfrett 1991; Read and Cumming 2003). However, no examination of corresponding predicted tilt in the binocular RF (Fig. 1), which would be more direct evidence for surface slant representation, was available. Comparison of the preferred disparity gradient and the ratio of optimal spatial frequencies for the two eyes as measured by drifting gratings shows a significant correlation between these two parameters. However, the correlation was significant only for neurons with inseparable RFs. In general, the correlation was not as good as we initially expected. In fact, only about one third of the neurons show responses consistent with the theoretical prediction. Other neurons are distributed outside the range of prediction for the dif-frequency disparity energy model. Thus we have explored additional alternative possibilities.

Hierarchical organization for slant-in-depth encoding

Another obvious possibility for generating slant-in-depth selectivity is by spatially pooling multiple neurons with progressive shifts of their preferred disparities (Fig. 12A). To examine this possibility, we analyzed the aspect ratio of binocular RFs. The prediction based on the pooling model of Fig. 12 was not fulfilled despite the evidence for extensive spatial pooling. On the contrary, the relationship between the aspect ratio and the disparity gradients correlated in the opposite direction from our expectation. The neurons with little pooling (aspect ratio near 1) tended to have a variety of preferred disparity gradients, whereas RFs of those with substantial pooling were not tilted (Fig. 13D). However, the opposite result does not necessary rule out possible roles for the spatial pooling. For example, we should also note the possibility that spatial pooling actively generates neurons with high aspect ratios and tuned to near-frontoparallel surface slants to enhance slant discrimination performance for near-frontoparallel surfaces. Such a possibility is consistent with our findings.

In addition, our findings may have a possible bias in the way disparity gradients and lateral spatial extents are negatively correlated. If we assume equal average physical spatial extents for depth and frontoparallel directions for a large number of objects in the physical world, slanted surfaces on average should occupy a narrower frontoparallel extent than that of nonslanted surfaces. It would be of interest to examine stereoimage statistics of natural scenes to determine the exact form of such a correlation.

Two kinds of dif-frequency models

Psychophysically, Halpern et al. (1996) reported that dif-frequency organization per se does not provide a robust slant-in-depth signal. Such a result appears to contradict the premise of the dif-frequency notion. However, we must note that there are two distinct levels of dif-frequency models. One is the strong form of the dif-frequency model that was examined and ruled out by Halpern et al. This model is based on the notion that a spatial frequency difference as such (without consistent local binocular correlations) is sufficient to signal surface slant. The other, weaker form of dif-frequency model, which we have examined, is an extension to the disparity energy model where the spatial frequency difference provides additional information regarding slant on top of local disparity information. Neurons with tilted binocular RFs will respond maximally when both the local binocular disparity and the surface slant simultaneously match the RF parameters. Therefore such a neuron is tuned to both the disparity and the interocular frequency difference.

Our findings on the effects of the interocular frequency difference is highly analogous to those reported by Bridge and Cumming (2001) with respect to the interocular orientation difference. They have found that monkey V1 neurons show responses to interocular orientation difference in a predictable manner based on the “dif-orientation” disparity energy model, and that the neural responses depend on both the binocular disparity and orientation difference. This is exactly what we find for spatial frequency. They have also found that the V1 neurons are not tuned for the relative orientation difference. Tuning for the relative orientation difference means that the optimal orientation difference is invariant regardless of the absolute orientations of the stimuli. Similarly, the model based on tilted binocular RF predicts no tuning for the relative spatial frequency difference. As with the disparity invariance of surface orientation tunings found in higher-order visual areas, tunings for the relative orientation or spatial frequency difference may be found in those cortical areas.

Is there an orientation bias for slant-in-depth encoding?

We investigated the possible orientation bias for slant-in-depth encoding because such an orientation bias has been found for phase-disparity encoding (DeAngelis et al. 1991, 1995). As apparent from Fig. 15, there is no orientation bias in the distributions of disparity gradients. Perhaps, this difference may be explained by the fact that the ratio of spatial frequencies across the eyes is independent of the orientation. In other words, neurons with any preferred orientation, except those tuned to the exact horizontal, can make equal contributions for signaling a given surface slant (i.e., by signaling a given spatial frequency ratio). The situation is quite different for the RF phase disparity because the key parameter—the horizontal disparity (the primary determinant of depth)—is dependent on the orientation. To produce neurons tuned to a given horizontal disparity, the required RF phase difference is smaller for neurons tuned to orientations closer to horizontal (Ohzawa et al. 1996). Therefore although there is no need for neurons having large phase difference at near horizontal orientations, neurons tuned to any orientation are equally important and useful for signaling slant information. Admittedly, this is
Discrepancies between the monocular and the disparity-tuning properties

The discrepancies between the optimal spatial and disparity frequencies (Fig. 11F) are similar to those reported previously (Ohzawa et al. 1997; Read and Cumming 2003). However, we have found substantial discrepancies only for neurons with inseparable RFs but not those with separable RFs. These results suggest that the discrepancy originates from some form of nonlinearity in the pooling process that underlies a hierarchical chain of processing where outputs of units with separable RFs are used to construct neurons with inseparable RFs. This notion is strengthened by the results presented in Fig. 16 in that neurons with larger aspect ratios (thus more pooling) tended to have a greater degree of discrepancy. Unfortunately, from our study it is not possible to determine details of where exactly the presumed nonlinearity lies. For example, it is still not known whether neurons with large aspect ratios receive input from complex cells organized as a disparity energy unit (with aspect ratio = 1) or if they directly collect input from neurons with separable RFs without the intermediate units. Further studies will be needed to address these questions.

In conclusion, neurons in areas 17 and 18 appear to encode slant-in-depth of 3D surfaces by having a variety of tilts in their binocular RFs. There are sufficient variations in the RF tilt angles for representing the range of 3D surface slants that occur in the real world. However, there may be multiple mechanisms by which tilted binocular RFs are generated. RF tilts for a subset of neurons could be accounted for by the dif-frequency model. However, neither the dif-frequency model nor the hierarchical pooling model could completely explain the entire data. It is possible that these neurons in the early visual areas contribute to surface slant selectivity of neurons in higher-order visual areas.

APPENDIX

The relationship between disparity gradient and slant in the actual world is described by Blakemore (1970). Here we extend the description of surface slant into the frequency domain, and derive the relationships among disparity gradient, tilt angle of binocular receptive field, disparity frequency, frontoparallel frequency, and monocular spatial frequencies.

When two objects, A and B, are separated in 3D space as shown in Fig. A1, disparity gradient for the line connecting the two objects (\(\Delta d\)) is defined as the difference of binocular disparities for the two objects, \((d_A - d_B)\), divided by their spatial separation in cyclopean space (\(\gamma\)). Thus the disparity gradient is defined as

\[
\Delta d = \frac{d_A - d_B}{\gamma}
\]  
(A1)

\(d_A - d_B = \alpha - \beta \quad \gamma = \frac{\alpha + \beta}{2}
\]  
(A2)

where visual angles \(\alpha\) and \(\beta\) indicate the separation between the objects in the monocular retinal space for the left and right eyes, respectively. \(\alpha\) and \(\beta\) are calculated from the viewing distance and the separation of eyes as

\[
\tan \alpha = \frac{a}{b + c} \quad \tan \beta = \frac{a}{b - c} \quad e = \frac{\alpha \tan \phi}{2}
\]  
(A3)

where \(\alpha\) indicates the separation between left and right eyes, \(b\) is the distance to the fixation point from the subject, and \(c\) is the distance between the fixation point and the objects. \(\phi\) is the slant angle in the real-world space.

Similarly, slant of binocular RF may be expressed in terms of disparity gradient. When the size of binocular RF for the left and right eyes are \(\alpha\) and \(\beta\), the binocular RF size is described as a function of slant of binocular RF, \(\theta\)

\[
\alpha = \beta \tan (45 - \theta)
\]  
(A4)

Therefore by substituting \(\alpha\) and \(\beta\) into Eqs. A1 and A2, we obtain the disparity gradient as

\[
\Delta d = \frac{2[1 - \tan (45 - \theta)]}{1 + \tan (45 - \theta)}
\]  
(A5)

If a slanted surface contains \(n\) cycles of a grating, the spatial frequencies as viewed by the left and right eyes are written as

\[
f_l = \frac{n}{\alpha} \quad f_r = \frac{n}{\beta}
\]  
(A6)

Therefore the spatial frequency ratio is

\[
\frac{f_l}{f_r} = \frac{\beta}{\alpha} = \tan (45 + \theta)
\]  
(A7)

The disparity gradient is rewritten in terms of the spatial frequency ratio as

\[
\Delta d = \frac{2(f_{\text{ratio}}) - 1}{f_{\text{ratio}} + 1}
\]  
(A8)

The theoretical curve shown by solid line in Fig. 11, C and D is obtained from this equation.

Disparity frequency and frontoparallel frequency of tilted binocular RF of complex cell are derived as follows. We begin with a model of complex cells based on a generalized disparity energy model (Ohzawa et al. 1990, 1997; Qian and Mikaelian 2000) where the left and right spatial frequencies may be different. According to this model, a complex cell receives input from quadrature pairs of simple cells. Members of the quadrature pairs may be modeled as having left and right monocular RFs that are even \((W_{\text{even}})\) and odd \((W_{\text{odd}})\) symmetric

\[
W_{\text{even}} = e^{-i \omega x_0} \cos (2\pi f_s x), \quad W_{\text{odd}} = e^{-i \omega x_0} \cos (2\pi f_s x + \varphi)
\]  
(A9)
profiles the complex cell is the sum of squared sums of the left and right RFs from frontoparallel space, spanning 2° disparity. Therefore left and right positions are written as $x_L$, $x_R$. We now rewrite Figure 1, C and E is derived from this equation.

To present the binocular RF data, we remove the contributions of monocular terms by taking the difference of binocular RFs (measured with contrast-matched and mismatched stimuli) as described by Ohzawa et al. (1997), thereby extracting the pure binocular interaction component. The last term of Eq. A11, $2(W_{\text{Leven}} W_{\text{Reven}} + W_{\text{Lodd}} W_{\text{Rodd}})$ is the binocular interaction component. Therefore

\[ R_{\text{interaction}} = 2(W_{\text{Leven}} W_{\text{Reven}} + W_{\text{Lodd}} W_{\text{Rodd}}) \]

By substituting Eq. A14 into Eq. A12, $R_{\text{interaction}}$ is expressed as

\[ R_{\text{interaction}} = 2e^{-i\frac{\omega}{2}} \cos\left\{2\pi f_{\text{L}} x_L + 2\pi f_{\text{R}} x_R + \phi\right\} \]

Therefore based on this equation, frequencies in the frontoparallel and disparity dimensions are

\[ f_{\text{frontoparallel}} = f_L - f_R \]

\[ f_{\text{disparity}} = f_R + f_L \]

We thank laboratory members H. Tanaka, S. Nishimoto, R. Kimura, K. Sasaki, M. Fukui, M. Iida, M. Arai, T. Ninomiya, and T. Ishida, who participated in recording sessions.

ACKNOWLEDGMENTS

This work was supported by Ministry of Education, Culture, Sports, Science and Technology Grant 15029230 and the Project on Neuroinformatics Research in Vision through special coordination funds for promoting science and technology and by Japan Society for the Promotion of Science Grant 13308048.

REFERENCES


