on some of the 34 papers and 45 posters presented at the conference, which was held at York University in Toronto on 22 – 26 June 1993.

It hardly needs to be mentioned that one of the factors responsible for the rapid growth of ideas about binocular stereopsis and optic flow has been interaction between those investigating 3-D vision from the different perspectives of physiology, psychophysics, and computation/machine vision, and hence we invited researchers from each of these areas. Important new evidence about the biological basis of stereopsis and optic flow was reported in several papers. For example, Trotter showed that the responses of disparity cells in monkey visual cortex can be modulated by the distance to the display while DeAngelis et al provided further evidence that disparity selectivity may be based on phase rather than position differences. The report by Orban et al of monkey MST cells selective to the different optic flow components of div, curi, and def provoked a vigorous debate (especially from those working in computational vision) about the relationship between physiological and computational levels of explanation. The use of optic flow information in judging the direction of heading was represented in the psychophysical work of Warren and Saunders and Regan et al and the computational investigations of Campani et al and Hildreth.

The largest group of papers and posters at the meeting was devoted to the psychophysics of disparity processing. The perception of 3-D surface properties was a key theme, represented in the papers of Tyler and Kontsevich, McKee et al, Gillam and Blackburn, and Howard. In our own paper, we presented clear evidence for the role of vertical disparities in the scaling of 2-D size and 3-D depth. The appropriate metric for representing 3-D space has also become a lively issue as is clear from the papers of Koenderink et al and Todd et al (both this issue). In the real world, disparity and optic flow cues do not exist in isolation, and hence it is not surprising that there has been a growth of interest over the last ten years in the interactions between depth cues. At the conference, Stevens dramatically demonstrated some of those interactions on a specially constructed 12 feet × 8 feet stereo screen while Frisby et al argued the merits of viewing the world through spectacles containing multiple pinholes in order to minimise accommodation cues. Ono and Mapp set the stage with that of van de Grind et al show that the laws which capture and describe binocular judgements of visual direction are still not fully understood. Eklundh's computational studies using active 'head' and Mayhew's use of neural network architectures for the control of camera position are unfortunately not represented in these issues. Overall, the conference clearly demonstrated the value of an interdisciplinary approach to understanding 3-D vision. Hopefully, the next fifteen years will allow us to distinguish the Helmholtzian wood from the trees.

The conference was generously supported by NATO as an Advanced Research Workshop, NSERC (Canada), and The Institute for Space and Terrestrial Science at York University. We are also extremely grateful to Suzanne McKee who acted as Editor for the papers published in these special issues of Perception and who tirelessly pursued the delinquent authors and referees.

Brian Rogers
Department of Experimental Psychology, South Parks Road, Oxford OX1 3UD

References


Rock I, 1984 Perception (New York: Scientific American Library)

Neuronal mechanisms underlying stereopsis: how do simple cells in the visual cortex encode binocular disparity?

Gregory C DeAngelis, Izumi Ohzawa, Ralph D Freeman

Group in Vision Science, School of Optometry, University of California, Berkeley, CA 94720-2020, USA

Based on paper presented at the Conference in Binocular Stereopsis and Optic Flow, Toronto, Canada. 22 – 26 June 1993

Abstract. Binocular neurons in the visual cortex are thought to form the neural substrate for stereoscopic depth perception. How are the receptive fields of these binocular neurons organized to encode the retinal position disparities that arise from binocular parallax? The conventional notion is that the two receptive fields of a binocular neuron have identical shapes, but are spatially offset from the point of retinal correspondence (zero disparity). We consider an alternative disparity-encoding scheme, in which the two receptive fields may differ in shape (or phase), but are centered at corresponding retinal locations. Using a reverse-correlation technique to obtain detailed spatiotemporal receptive-field maps, we provide support for the latter scheme. Specifically, we show that receptive-field profiles for the left and right eyes are matched for cells that are tuned to horizontal orientations of image contours. However, for neurons tuned to vertical orientations, the left and right receptive fields are predominantly dissimilar in shape. These results show that the striate cortex possesses a specialized mechanism for processing vertical contours, which carries the horizontal-disparity information needed for stereopsis. Thus, in a major modification to the traditional notion of the neural basis of stereopsis, we propose that binocular simple cells encode horizontal disparities in terms of phase at multiple spatial scales. Implications of this scheme are discussed with respect to the size–disparity correlation observed in psychophysical studies.

1 Introduction

The image formed upon the retina of each eye is a two-dimensional representation of the three-dimensional world. One of the major tasks of the visual system is to reconstruct the third dimension of a scene (ie depth) from this pair of two-dimensional images. Although there are several monocular cues to depth, including size, perspective, and relative motion, animals with frontally located eyes have access to a potent additional cue. Because horizontal separation of the eyes gives a slightly different view to each eye (binocular parallax), there are small variations in position, mainly along the horizontal dimension, between corresponding features in the two retinal images (see figure 1). This difference in position between corresponding features on the two retinas is called binocular disparity. The visual system uses horizontal binocular disparities to gauge relative depth (Wheatstone 1838), as evidenced by the vivid sensation of depth that one experiences when viewing a random-dot stereogram (see Julesz 1971).

It is generally thought that binocularly driven neurons in the primary visual cortex form the first stage for the processing of binocular disparity, because cells at lower levels in the visual pathway (ie the retina and lateral geniculate nucleus) are not driven by stimulation through both eyes (eg Hubel and Wiesel 1961, 1962). However, the specific neural mechanisms that underlie depth discrimination are not well understood. The goal of our research is to understand how the receptive fields (RFs) of binocular cortical neurons are organized to process the disparity information used in stereopsis. Because simple cells appear to receive most of the direct thalamic input to

1 Author to whom correspondence and requests for reprints should be addressed.
the striate cortex (eg Hubel and Wiesel 1962; LeVay and Gilbert 1976; Gilbert 1977; for reviews, see Gilbert and Wiesel 1981; Gilbert 1983; Martin 1984), it is generally assumed (eg Maske et al 1984; Nomura et al 1990) that simple cells are responsible for the initial encoding of binocular disparity. Complex cells appear to be ideally suited to perform higher-level processing of the disparity signals encoded by simple cells (Ohzawa et al 1990; see section 4.5).

In this study, we examine the mechanism through which binocular disparities are encoded by simple cells in the cat's striate cortex. Traditionally, it has been assumed that horizontal disparities are encoded by a group of neurons with varying relative displacements, or incongruities, between RFs for the left and right eyes (Barlow et al 1967; Nikara et al 1968; Joshua and Bishop 1970; von der Heydt et al 1978; Maske et al 1984). The two RFs of these binocular neurons are assumed to be identical in shape, as suggested by previous studies (Hubel and Wiesel 1962; Maske et al 1984). This traditional model for disparity encoding is illustrated schematically in figure 2a, where the left and right RFs (solid curves) of a hypothetical simple cell have the same luminance-sensitivity profile, but are located at noncorresponding (incongruous) points on the two retinæ. As shown in figure 2b, the preferred disparity for such a cell is determined by how much the two RFs are displaced from the point of retinal correspondence.

On the basis of the traditional model shown in figure 2a, several groups of investigators (Barlow et al 1967; Nikara et al 1968; Joshua and Bishop 1970; Hubel and Wiesel 1973; von der Heydt et al 1978) have searched for horizontal–vertical differences in the distribution of positional disparities (incongruities) between the left and right RFs of cortical cells. This difference is expected if the visual system encodes disparity information efficiently. Because the eyes are separated laterally, binocular parallax produces a much larger range of horizontal than of vertical disparities.

![Figure 1](image1.png)

**Figure 1.** Basic geometry of the disparity-encoding problem. In this illustration, an observer is fixating on the point shown as a filled circle (top). The images of this fixation point fall upon corresponding points on the two retinæ. By definition, the binocular position disparity between these corresponding points is assigned a value of zero. If the observer maintains his or her fixation point, then the images of an object which is nearer to the subject (indicated by an open circle) will now fall on noncorresponding points on the two retinæ (ie a crossed, disparity). In this example, the right retinal image of the nearby object is displaced laterally from the image of the fixation point. For a binocular neuron to be tuned to the crossed disparity associated with the nearby object, it must have receptive fields (RFs) in the left and right eyes that are arranged so that their peak sensitivities (shown a thick patch at the back of the eyes) occur at the correct noncorresponding retinal coordinates.

Thus, only a small range of vertical disparity values near zero needs to be encoded. In the original study of RF position disparities, Barlow et al (1967) reported that the range of horizontal incongruities is three times as large as the range of vertical incongruities. This finding suggested that the visual cortex is specialized for the processing of horizontal binocular disparities, since these carry the information useful for depth discrimination. However, subsequent studies (Joshua and Bishop 1970; Hubel and Wiesel 1973; von der Heydt et al 1978) have not confirmed the findings of Barlow et al (1967). Perhaps the differences between studies can be attributed to the confounding factor of RF eccentricity and the problems associated with monitoring eye position (see Ferster 1981; LeVay and Voigt 1988). Also, in most of these studies, quantitative methods were not used to map RFs; hence, estimates of the center location of RFs may have been subject to errors.

An alternative to the traditional model of disparity encoding is possible if one discards the assumption that the left and right RFs of binocular cells must have identical structure (Freeman and Ohzawa 1990). This alternative scheme is illustrated in figure 2c. Suppose that the RF centers, as defined by the peaks of the RF envelopes (dashed curves), are located at corresponding retinal points and that the left and right RF profiles differ in shape (as indicated by a 90° phase difference in figure 2c).
In this case, the optimal disparity of the cell is determined by the difference in the shape (or phases) of the RF profiles, and by the size (or spatial frequency) of the fields (see figure 2d). This encoding scheme links the disparity selectivity of each neuron to its spatial frequency selectivity (see figure 12 and section 4.2), in such a way that low-frequency cells encode coarse disparities and high-frequency cells encode fine disparities (see Marr and Poggio 1979 for a similar idea).

To determine which mechanism for the encoding of disparity actually operates in the visual cortex, it is first necessary to compare RF profiles for the left and right eyes. If these profiles are always closely matched, then the traditional model (figure 2a) would be favorred. On the other hand, if the left and right RF profiles differ for some cells, then the alternative scheme (figure 2c) for disparity encoding would be plausible. In previous studies (Hubel and Wiesel 1962; Maske et al 1984) it was reported that RFS for the left and right eyes are closely matched in structure, but the methods used in these studies were either quantitative or inadequate for drawing firm conclusions (see section 4.1). In this study, we reexamine the degree of similarity between the left and right RFS of binocular simple cells. To make quantitative comparisons between the two eyes, detailed RF maps are obtained by using a reverse-correlation technique (Jones and Palmer 1987a; DeAngelis et al 1993a). By fitting these RF profiles with Gabor functions, a quantitative index of the similarity of the two RFs is obtained.

We find that cells tuned to stimulus orientations near horizontal tend to have closely matched RFs for the left and right eyes. In contrast, cells tuned to orientations near vertical may have very differently shaped RFs for the two eyes. These findings are consistent with the idea (figure 2c; see also Nomura et al 1990) that horizontal disparities are encoded by differences in the shapes (or phases) of the left and right RFs. Thus, the striate cortex possesses a specialized mechanism for processing vertical contours, which provide the horizontal-disparity information needed for estimating depth through stereopsis. A preliminary account of some of these findings has been published previously (DeAngelis et al 1991). We also examine how the temporal dynamics of RF structure (see DeAngelis et al 1993a, 1993b) relate to a phase-based scheme for disparity encoding. For cells that have space-time inseparable RFs (McLean and Palmer 1989; Reid et al 1991; DeAngelis et al 1993a), we show that changes in RF phase as a function of time are matched for the two eyes, so that the interocular phase difference remains constant over time. Predictions of the phase-encoding scheme are discussed with regard to psychophysical results.

2 Methods
All experiments were performed with adult cats reared in a normal environment. During data collection, the animal was anesthetized and paralyzed, in order to prevent eye movements. Control of eye movements is necessary in order to obtain detailed RF profiles, a procedure which is time-consuming. Detailed descriptions of the experimental apparatus and procedures have been given in recent reports (DeAngelis et al 1992, 1993a). Hence, only a brief description of the relevant details is given below.

2.1 Surgical procedures
Each animal was prepared for single-unit recordings from the primary visual cortex (area 17) by means of standard surgical procedures (see DeAngelis et al 1993a). During surgery, the cat was anesthetized with halothane gas (2.5%-3% in O2). A rectal temperature probe was inserted, electrocardiographic (ECG) and electroencephalographic (EEG) electrodes secured, and a femoral vein catheterized. A tracheostomy was performed and a tracheal tube inserted. The animal was then secured in a stereotaxic apparatus by means of ear bars and a mouth bar. A small section of skull and dura (~5 mm in diameter centered on Horsley–Clarke coordinates P4 L2) was then removed to allow insertion of a pair of tungsten-in-glass microelectrodes (Levick 1972) into the visual cortex. After the electrodes had been lowered to just above the cortical surface, agar and melted wax were used to fill the hole and create a sealed chamber. The cat was then paralyzed with gallamine triethiodide (Flaxedil), which was continuously infused at a rate of 10 mg kg⁻¹ hr⁻¹, along with 1 mg kg⁻¹ hr⁻¹ of sodium thiopental (Surital) as a supplementary anaesthetic. Artificial ventilation was carried out with a gas mixture of 70% N₂O, 29% O₂, and 1% CO₂. Temperature, heart rate, EEG, ECG, expired CO₂ level, and intratracheal pressure were monitored continuously, and an appropriate level of anesthesia maintained. Pupils were dilated with atropine (1%), nictitating membranes were retracted with Neosynephrine (10%), and corrective (+2D) contact lenses with 4 mm artificial pupils were positioned on each cornea.

Experiments typically lasted for four days. At the end of an experiment, the animal was administered an overdose of pentobarbital sodium (Nembutal). After the experiment, cortical tissue was prepared for histological examination. Electrode tracks were reconstructed, and cortical laminae identified. This analysis confirmed that all cells were recorded from area 17, and that cells were sampled from all laminae. The majority of simple cells were recorded from layers 4 and 6.

2.2 Experimental apparatus and data collection
The apparatus used for conducting these experiments is shown schematically in figure 3. The animal faced a large rear-projection screen, on which bars of light of variable dimension can be swept manually via a joystick. A pair of beam splitters (70% reflectance) allowed the cat to view stimuli on the tangent screen, as well as patterns that were presented on either of two video displays, one for each eye. The video displays (Mitsubishi Electronics; mean luminance 45 cd m⁻²; screen subtense 28 deg x 22 deg) have a resolution of 1024 pixels x 804 pixels and are refreshed at 76 Hz. A variety of visual stimuli can be generated on these displays by a dedicated computer which employs two high-resolution graphics boards (Imagrib). A second computer controls the visual stimulator through a serial port and coordinates data acquisition, real-time analysis, and data display.

Figure 3. Schematic illustration of the experimental apparatus used in this study. Responses of cortical neurons were recorded extracellularly while an anesthetized, paralyzed cat viewed visual stimuli that were presented on a large tangent screen or on either (or both) of a pair of video displays. See text for additional details.
The tungsten-in-glass microelectrodes were advanced through the cortex until neural activity was observed. Action potentials from a single neuron (or small group of neurons) were isolated with the use of a spike discriminator, and the time of occurrence of each action potential was recorded with 1 ms resolution and stored on disk. These signals were also used by the controlling computer to display responses in real time.

2.3 Experimental protocol

The optic disks were projected onto the large tangent screen by means of a reversible ophthalmoscope. The positions of the arcuate centers were then estimated from the positions of the optic disks. Once the action potential of a single cell was isolated, the RF was initially explored with a bar of light that was moved manually. The location of the RF (relative to the positions of the optic disks) was marked on plotting paper by means of a large beam splitter behind the rear-projection screen. Ocular dominance and direction selectivity were also estimated at this time.

Before starting quantitative measurements, an interactive "search" program (see DeAngelis et al. 1993a) was used to make preliminary observations of the stimulus selectivity of each cell. In this procedure, the cell was stimulated with the use of round grating patches whose orientation, spatial frequency, position, and size were manually controlled with a pointing device. Once the optimal orientation and spatial frequency had been estimated, a small (1 - 2 deg) patch of grating was positioned carefully to give the largest response. This position was taken as the location of the center of the RF, and all stimuli used subsequently were centered on this point. For binocular cells, this search procedure was performed for both the dominant and the nondominant eye.

After the search procedure, quantitative measurements of the orientation and spatial-frequency tuning of the cell were obtained by presenting sequences of drifting sinusoidal gratings in which one stimulus parameter (orientation) was varied. In some of these experiments, temporal-frequency-tuning curves were also obtained. Typically, large field (10 - 20 deg diameter) grating stimuli were used for these tests; optimally sized grating patches were used if the cell exhibited end-inhibition or side-inhibition (DeAngelis et al. 1994). To construct tuning curves, gratings were presented for 4 s each in blocks of randomly interleaved trials, during which peristimulus time histograms of the responses were accumulated. Each stimulus was typically presented 4 - 6 times, and successive stimuli were separated by a period of 2 - 3 s during which the animal viewed blank screens of the same mean luminance as the gratings. For binocular cells, tuning curves were constructed for each eye by interleaving left-eye and right-eye stimuli during this process. After presentation of the complete set of stimuli, the magnitude of the accumulated response to each different stimulus was evaluated by Fourier analysis of the peristimulus time histograms. Because the responses of simple cells to drifting sinusoidal gratings are modulated at the temporal frequency of the stimulus, the ratio of the amplitude of the first harmonic to the mean firing rate can be used to classify simple cells (see Skottun et al. 1991 for review).

2.4 Receptive-field mapping (reverse-correlation analysis)

After measuring orientation and spatial-frequency selectivity with gratings, we obtain a detailed spatiotemporal RF profile for each eye. This is done by using a reverse-correlation technique (DeBoer and Kuyper 1968; Jones and Palmer 1987a). A detailed description of our reverse-correlation algorithm can be found in a recent publication (DeAngelis et al. 1993a). Here, we present an abbreviated account, with an emphasis on the important concepts.

The RF (for each eye) of any visual neuron is inherently a function of three independent variables: two dimensions of space (X and Y, retinal coordinates) and time (T). For many cells, spatial and temporal aspects of RF structure cannot be considered independent (see McLean and Palmer 1989; Reid et al. 1991; DeAngelis et al. 1993a, 1993b); thus, we must characterize RFs in the joint space-time domain. Traditionally, mapping of RFs has been achieved by repeatedly presenting a stimulus and recording the response of the neuron over some period of time after the stimulus has been presented (e.g. Movshon et al. 1978; Palmer and Davis 1981). However, this procedure is too slow to be used for mapping RFs, with high resolution, in the three-dimensional space-time domain. An alternative approach (generally known as the white-noise method; see Marmarelis and Marmarelis 1978) is to stimulate the neuron with a continuous, random sequence of very brief stimuli and to store the ongoing response of the neuron. By cross correlating the response of the neuron with the stimulus sequence, one can characterize the relationship between stimulus and response. If the neuron behaves linearly, which is approximately true for simple cells (DeAngelis et al. 1993b), this cross-correlation procedure yields the spatiotemporal impulse response of the neuron (see DeBoer and Kuyper 1968). For simple cells, the spatiotemporal RF profile (impulse response) can be used to predict, with reasonable accuracy, the responses to a variety of stimuli (DeAngelis et al. 1993b; Jones and Palmer 1987b).

An efficient algorithm for computing the cross correlation between stimulus and response is known as reverse (or triggered) correlation (DeBoer and Kuyper 1968). Our reverse-correlation algorithm is shown diagrammatically in figure 4. The visual stimulus is a pseudorandom sequence of small bright and dark rectangular bars that are presented on a stimulus grid which typically has 20 x 20 spatial locations. This stimulus grid is centered over the RF, and its size is adjusted to cover the entire RF. Both the stimulus grid and the small bar stimuli are oriented to match the preferred orientation of each cell (as determined from the orientation-tuning curve). The bar stimuli are typically about 1.5 deg x 0.5 deg in size and are presented for a duration of 40 ms, although these parameters are adjusted according to the spatial and temporal resolution of the response of each cell (see DeAngelis et al. 1993a for details). We generally attempt to make the stimuli as small as possible, in both space and time, provided that they still elicit a reasonable response from the neuron. Bright and dark bars have effective luminances, as seen by the cat through beam splitters (figure 3), of 19.5 and 0.5 cd m^-2, respectively, on a background of 10 cd m^-2. Thus, the bright and dark bars have opposite but equal contrasts.

The reverse-correlation algorithm operates as follows. Individual bar stimuli are presented one at a time, in rapid succession, on the stimulus grid. For each successive 40 ms presentation, both the location of the bar on the stimulus grid and the contrast of the bar (bright or dark) are chosen randomly. Each time an action potential occurs, we correlate it with the recent history of the stimulus. Figure 4 illustrates this process for four different values (T_1 - T_4) of the reverse-correlation delay. Analysis for each value of T will yield a 'snapshot' of the RF at a different point in time. Figure 4 shows how one action potential is assigned to a spatial RF map at each value of T. For example, for T_1 = 20 ms, the spike is assigned to a dark bar that was presented on the right-hand side of the stimulus grid. To record this event, we increment a two-dimensional histogram bin at the coordinate corresponding to the position of the center of the stimulus. Analogously, for T_2 = 80 ms, the spike is assigned to a bright bar, and the appropriate histogram bin is incremented. Separate histograms are collected for the responses to bright and dark bars at each correlation delay. This assignment process is repeated for each action potential that is elicited by the stimulus sequence. If there is coupling between the stimulus and response at a particular correlation delay (T_i), then a pattern will emerge in the bright and dark response histograms; otherwise, the histograms will show no structure (i.e. they will be flat).

For simple cells, regions responsive to bright and dark stimuli are complementary (nonoverlapping); hence, for these cells, we construct a composite RF profile by
subtracting the dark-bar responses from the bright-bar responses (Jones and Palmer 1987a; DeAngelis et al. 1993a). The composite spatiotemporal RF profile for a typical simple cell is shown in the bottom row of figure 4. Note that the spatial profile is weak at \( T_1 = 20 \) ms, becomes stronger at \( T_2 = 80 \) ms and \( T_3 = 140 \) ms, and becomes weak again at \( T_3 = 200 \) ms (i.e. the strength of the RF waxes and wanes over time). Outside of this approximate time range, the RF profile has no discernible structure. Note also that the spatial arrangement of the bright-excitatory and dark-excitatory subregions (solid and dashed contours, respectively) changes drastically as a function of time (subregions appear to move across the RF). Recent work suggests that this spatiotemporal behavior is the basis of direction (Reid et al. 1991; DeAngelis et al. 1993b) and speed (McLean and Palmer 1989) selectivity.

For a stimulus grid having \( 20 \times 20 \) locations, one stimulus sequence consists of 800 different bar stimuli (400 grid locations \( \times \) 2 bar contrasts), and lasts about 30 s. Usually, the stimulus sequence was repeated 30–50 times in order to obtain enough spikes for a smooth RF profile. Thus, the time required for a complete measurement of the RF profile was in the range of 15–30 min. The presentation order of stimuli was randomized each time the stimulus sequence was repeated, so that a particular temporal sequence of bright and dark bars was never duplicated.

3 Results

For this study, data have been obtained for sixty-five binocular simple cells recorded from the striate cortices of eighteen adult cats. Other cells from these cats were recorded for different studies. For each cell, spatiotemporal receptive field profiles were obtained by stimulating through the left and right eyes (in separate trials).

3.1 Comparison of receptive-field profiles for the two eyes

Figure 5 shows RF profiles, as obtained by means of the reverse-correlation technique, for a binocular simple cell. We first compare the spatial structure of RFs for the two eyes, at one particular point in time \( T_3 = 50 \) ms. Figures 5b and 5c show two-dimensional (2-D) spatial \((X-Y)\) RF profiles for the left and right eyes, respectively. Solid contours in the \( X-Y \) profiles delimit bright-excitatory (or ON) subregions, whereas dashed contours indicate dark-excitatory (or OFF) subregions. Notice that these spatial profiles are quite similar in structure, each having a dark-excitatory subregion to the left of a bright-excitatory region. To simplify the comparison between RFs for the two eyes, one-dimensional (1-D) spatial profiles (filled circles in figures 5d and 5e) are constructed by integrating the 2-D \((X-Y)\) profiles along the \(Y\) axis, which is always parallel to the preferred orientation of the cell.

To quantify differences in shape between 1-D RF profiles for the left and right eyes, each profile is fit with a Gabor function, \(G(x)\), given by

\[
G(x) = k \exp \left( -\frac{2(x-x_o)^2}{w^2} \right) \cos(2\pi f_{\text{prof}}(x-x_o)+\Phi),
\]

where \(k, x_o, w, f_{\text{prof}}\), and \(\Phi\) are free parameters. A Gabor function, which is simply a sinusoid modulated over space by a Gaussian envelope (Gabor 1946), is chosen because it has a simple mathematical form with convenient parameters, and also because it provides a good fit to RF profiles of simple cells (e.g. Marcolja 1980).

(1) In this paper, we generally use the terms bright-excitatory and dark-excitatory, rather than ON and OFF (e.g. Hubel and Wiesel 1962), to describe the subregions of simple cells. These two descriptions are not quite equivalent. A particular spatial location within the RF may be bright-excitatory or dark-excitatory at different times during the response (i.e. the temporal impulse response may be biphasic, see figure 5f and 5g). It is the temporal sequence of bright and dark excitation that determines whether the cell will give an ON or OFF response to a flashed bar stimulus. If, for example, the temporal response (at a given position) has a bright-excitatory phase followed by a dark-excitatory phase, the simple cell will give an ON response to a bright bar (and an OFF response to a dark bar). Conversely, a subregion that exhibits dark excitation followed by bright excitation will give an OFF response to a bright bar (ON response to a dark bar).
Field and Tolhurst 1986; Jones and Palmer 1987b). The solid curves in figures 5d and 5e are Gabor functions that best fit the 1-D RF profiles. To compare RF shapes for the left and right eyes, we can use the phase, $\Phi$, of the sinusoidal component relative to the center of the Gaussian envelope (see equation 1). The value of $\Phi$ reflects the order and relative strengths of the bright-excitatory and dark-inhibitory RF subregions. The difference in phase, $\delta\Phi$, between RF profiles for the two eyes is, thus, a measure of their similarity. For comparison across cells, the absolute value of phase difference, $|\delta\Phi|$, is used to quantify differences in shape between RF profiles for the two eyes. Values of $|\delta\Phi|$ close to 0° indicate that the left and right profiles are closely matched. Values of $|\delta\Phi|$ near 180° indicate that one profile is nearly inverted with respect to the other. For the simple cell of figure 5, $|\delta\Phi| = 16^\circ$, indicating that the left and right RF profiles are fairly similar in shape.

It should be noted that $\delta\Phi$ summarizes differences in shape between the left and right RFs only if other RF parameters, namely the preferred spatial frequency ($f_{opt}$) and RF width ($w$), are well matched for the two eyes. This is, indeed, the case for the cells that we have studied (see figures 5–7; quantitative interocular comparisons of RF parameters other than phase will be presented in a forthcoming paper). Thus, $|\delta\Phi|$ can be used to quantify differences in shape between the two RFs. Note that the preferred orientations of the two eyes do typically differ by 5°–10° (see figures 5a and 6a). However, these orientation disparities are due, at least partly, to cyclorotation of the eyes after paralysis and are not likely to be used by the cat for detecting tilt in depth (Nelson et al. 1977).

It must also be noted here that the shape, or phase, of the receptive field for each eye can change with time (see figure 4, bottom). Thus, it is possible that the phase difference between RFs for the two eyes may also vary with time. By constraining 1-D spatial RF profiles at finely spaced intervals of the reverse correlation delay, $T$, we can examine how the spatial profile changes over time. These changes are best summarized as an $X$–$T$ profile (see figures 5f and 5g; see also DeAngelis et al. 1993a, 1993b). It will be shown below (figures 9 and 10) that the difference in phase between RFs for the left and right eyes generally remains constant over time, even though the individual phases of the two RFs may change drastically. Hence, for the remainder of this and the next section, we shall only deal with 1-D RF profiles obtained at the optimal correlation delay (indicated by horizontal lines in figures 5f and 5g).

Figure 6 shows data for another binocular simple cell. Figures 6b and 6c show 2-D spatial RF profiles obtained by stimulating the left and right eyes, respectively. Notice that the pattern of bright-excitatory and dark-inhibitory subregions is markedly different for the two eyes. This can be seen more clearly in figures 6d and 6e, where filled circles show 1-D RF profiles that are obtained by integrating the X–Y data. Solid curves represent the Gabor functions that best fit these 1-D profiles. The phase difference, $|\delta\Phi|$, between the left and right profiles is $123^\circ$, reflecting the fact that the two profiles have dissimilar shapes (in this case the two fields are approximately mirror images). For this cell, the 1-D profiles were obtained at a correlation delay of 68 ms. Complete X–T profiles for this binocular simple cell are shown in figures 6f and 6g.

Figure 7a shows 1-D RF profiles (filled circles), as well as Gabor fits (solid curves), for three more binocular simple cells, all having preferred orientations close to vertical. For each of these cells, 1-D RF profiles are obtained at the optimal reverse-correlation delay, as shown in figures 5 and 6 (the optimal delay is that which yields a 1-D RF profile having maximal area under the curve). Optimal delays vary from cell to cell (range 40–132 ms) but the optimal delays for the two eyes are closely matched (median interocular difference = 6 ms) for each cell. Note the similarity of the left and right RF profiles for the cell with $|\delta\Phi| = 13^\circ$. In contrast, almost completely opposite profiles are seen for the cell with $|\delta\Phi| = 159^\circ$, and a somewhat smaller mismatch is observed for the cell with $|\delta\Phi| = 113^\circ$.

Figure 7b is a summary of the results obtained for sixty-five binocular simple cells. Note that the vast majority of cells with preferred orientations near horizontal have values of $|\delta\Phi|$ near 0°. In fact, 79% (15/19) of cells tuned to orientations within 20° of horizontal have $|\delta\Phi|$ values smaller than 30°. The left and right RF profiles are well matched for these cells. In contrast, cells tuned to orientations within 20° of vertical exhibit a wide variety of phase differences between RF profiles for the two eyes. These cells are capable of encoding a range of binocular disparities through
differences in RF phase between the two eyes, as shown below (see figures 11 and 12). Thus, the data of figure 7b show clearly that there is a cortical specialization for the processing of vertical contours, which carry horizontal-disparity information.

![Image](image_url)

**Figure 6.** Comparison of left and right RF profiles for another binocular simple cell. The format of this figure is identical to that of figure 5. (b) and (c) 2-D spatial RF profiles for the left and right eyes, respectively. Note the alternating arrangement of bright-excitatory and dark-excitatory subregions, which is typical of simple cells. (d) and (e) 1-D spatial profiles (filled circles) that are obtained by integrating (over $Y$) the $X-Y$ profiles shown in (b) and (c). Solid curves in (d) and (e) are the best-fitting Gabor functions [see equation (1)]. For this cell, the phase difference between the left and right RFs is $[\Phi] = 123^\circ$. (f) and (g) $X$-$Y$ profiles for the left and right eyes, respectively.

### 3.2 Dependence of phase differences upon other stimulus parameters

It is clear from figure 7b that the distribution of phase differences, $[\Phi]$, between RF profiles for the left and right eyes depends upon the preferred orientations of a population of cells. This finding is consistent with the disparity-encoding scheme presented in figure 2c, because cells tuned to orientations near vertical exhibit a broad range of phase differences. In order to evaluate further this hypothesis, and to rule out the potentially confounding effects of other stimulus parameters, it is necessary to consider how the distribution of phase differences depends upon stimulus parameters other than orientation.

![Image](image_url)

**Figure 7.** A summary of phase differences between RF profiles for the left and right eyes. (a) 1-D RF profiles (filled circles) and fitted Gabor functions (solid curves) are shown for three simple cells whose preferred orientations are $75^\circ$, $85^\circ$, and $85^\circ$ respectively (top to bottom). (b) Relationship between the absolute value of phase difference, $[\Phi]$, and the preferred orientation is shown for sixty-five simple cells. In this polar-coordinate scatter diagram, the radial coordinate is phase difference, and the angular coordinate is preferred orientation. Orientation is given as the number of degrees away from horizontal, so that $90^\circ$ represents vertical. Note that cells preferring orientations near horizontal generally have values of $[\Phi]$ close to $0^\circ$, whereas cells preferring orientations near vertical exhibit a wide range of values of $[\Phi]$.

If disparity is encoded in terms of RF phase (figure 2c) and RF size varies inversely with preferred spatial frequency (Maffei and Fiorentini 1977), then the range of disparities that any simple cell can signal is determined by the preferred spatial frequency of the cell (size-disparity correlation; see section 4.3). Thus, it is of interest to examine the distribution of phase differences as a function of spatial frequency. Figure 8a shows phase difference, $[\Phi]$, plotted against preferred spatial frequency for the same population of sixty-five simple cells as shown in figure 7b. A wide range of phase differences is seen throughout the spatial frequency range. Linear regression analysis yields a correlation coefficient of $r = 0.17$, which is not significant ($p = 0.18$). The slope of the regression line (solid line in figure 8a) is not significantly different from zero, indicating that the distribution of $[\Phi]$ does not depend on spatial frequency. This result suggests that simple cells are capable of encoding disparity in terms of phase across a broad range of spatial frequencies.

In previous studies (Poggio and Fischer 1977; Ferster 1981; LeVay and Voigt 1986) it has been reported that neurons that prefer nonzero disparities (ie 'near', 'far', and 'tuned-inhibitory' cells) tend to be ocularly imbalanced, responding much more strongly to stimulation through one eye than through the other, whereas cells tuned to near-zero disparities (tuned-excitatory' cells) are ocularly balanced. Thus, one might predict, on the basis of the model in figure 11 (see also Nomura et al 1990), that ocularly imbalanced cells exhibit large phase differences between RFs for the two eyes, whereas ocularly balanced cells do not. It can be seen from figure 8b that this is not the case. An ocular balance index (OBI) is defined as

$$\text{OBI} = 1 - \frac{R_l}{R_l + R_c} - 0.5,$$

where $R_l$ and $R_c$ denote the response amplitudes of a given neuron to drifting sine-wave-grating stimuli (having optimal parameters) that are presented to the ipsilateral and contralateral eyes, respectively. Figure 8b shows the interocular phase difference,
[δΦ], plotted as a function of ocular balance for our population of simple cells. Values of the OBI near 0 denote cells that are strongly dominated by the contralateral eye (open symbols) or the ipsilateral eye (filled symbols). Values of the OBI near 1 represent cells that are equally responsive to stimulation of either eye (ocularly balanced). Linear regression analysis yields a correlation coefficient of \( r = 0.11 \), which is not significant (\( p = 0.38 \)). Note that the regression line in figure 8b is almost flat. It is clear from these data that [δΦ] does not depend on ocular balance. In particular, cells with large interocular phase differences do not tend to be monocularly driven. In fact, many of the cells with large phase differences have an OBI near 1. However, it should also be noted that most of the simple cells studied here (47/65) have values of the OBI between 0.5 and 1. There are very few cells in this sample that respond only to stimulation of one eye. In contrast, other studies (e.g., Hubel and Wiesel 1962; LeVay and Voigt 1988) show that the ocular-dominance distribution for simple cells is approximately uniform. This discrepancy arises because our sample of neurons is biased in favour of binocularly driven cells. The reason for this bias is that the reverse-correlation mapping procedure only works well on cells that give a reasonably vigorous response (at least 10–15 spikes per second to a drifting grating). Thus, to obtain RF profiles for both eyes, each cell must give a robust response to stimulation of both eyes.

To determine whether [δΦ] depends on any response parameter other than the preferred orientation, we performed a multiple regression analysis in which we tested the relationships between [δΦ] and eight different response parameters. The parameters tested were the preferred orientation, spatial frequency, and temporal frequency of each cell (averages of values for the two eyes), ocular balance, direction selectivity, and the interocular differences in preferred orientation, spatial frequency, and direction selectivity. Only the preferred orientation made a significant (\( p = 0.011 \)) contribution to the predictive power of the statistical model; none of the other parameters made a significant contribution (\( p > 0.1 \)). This analysis shows that the phase difference, [δΦ], between left and right RF profiles depends only on the preferred orientation of simple cells. Thus, the specialization of RF structure shown in figure 7b is exclusive to the domain of orientation.

3.3 Effect of space–time inseparability on phase difference

Until now, all comparisons between RF profiles for the two eyes have been made at a single (optimal) value of \( T \) (see figures 5 and 6). Let us now consider how the phase difference, \( \delta \Phi \), between RF profiles for the two eyes varies as a function of time. As mentioned above, for many simple cells spatial and temporal aspects of RF structure cannot be considered independent. The \( X - T \) profiles of simple cells can be assigned to one of two categories—those that are space–time separable and those that are not (McLean and Palmer 1989; Reid et al 1991; DeAngelis et al 1993a). It should be noted, however, that we observe a continuum with regard to space–time separability, rather than a clear dichotomy. For cells with space–time-separable RFs, the \( X - T \) profile can be approximated as the product of two 1-D profiles, one a function of space and the other a function of time \( [\text{ie } R(X, T) = G(X)H(T)] \). In these cases, the spatial phase \( (\Phi) \) of the RF is approximately constant over time, although there may be ±180° jumps in \( \Phi \) if the \( X - T \) profile is polyphasic in time (see figure 10 of DeAngelis et al 1993a). In contrast, for cells with inseparable \( X - T \) profiles, the spatial phase of the RF changes gradually with time (whereas other RF parameters remain relatively constant). In other words, there is no unique spatial RF profile for these cells. This behavior is illustrated in figure 9. Figure 9a shows an \( X - T \) profile for a simple cell with a space–time inseparable RF. This profile was obtained by stimulation of the left eye. Note that bright-excitatory and dark-inhibitory subregions are tilted to the right in the space–time domain. As a result, the region responsive to dark stimuli (dashed contours in figure 9a) shifts gradually to the right over time. This type of behavior is typical of simple cells with space–time inseparable RFs, and is thought to serve as a linear mechanism for motion selectivity (McLean and Palmer 1989; Reid et al 1991; DeAngelis et al 1993b). It can be seen from figure 9b that the spatiotemporal organization of the right-eye RF is very similar to that of the left eye.

To examine how the spatial phase difference, \( \delta \Phi \), between RFs for the two eyes varies over time, we take cross sections, at different values of \( T \), through the \( X - T \) data for each eye, and fit a Gabor function to each cross section. Figure 9c shows four 1-D RF profiles (filled circles) for each eye; these 1-D profiles are obtained by taking cross sections through the \( X - T \) data at four different values of \( T \) (\( T = 20, 60, 100, \) and 140 ms). Solid curves in figure 9c show the best-fitting Gabor function for each 1-D profile. Clearly, the shape of each RF changes as a function of time, and this change is reflected in the phase parameter, \( \Phi \), of the Gabor function. Figure 9d shows RF phase as a continuous function of time for both the left eye and the right-eye RFs (solid and dashed curves, respectively). Although \( \Phi \) changes gradually over time, these changes are closely matched for the two eyes. As a result, the phase difference, \( \delta \Phi = \Phi_L - \Phi_R \), remains approximately constant over time (see figure 9e). For this cell, the average value of \( \delta \Phi \), over the period of time shown in figure 9e, is \( -1.8° \) (± 10.4° SD).

(1) The \( X - T \) profiles of many simple cells are intermediate between the canonical separable and inseparable types seen in theoretical work (e.g., Adelson and Bergen 1985). An indirect metric of space–time inseparability is the linear estimate of direction selectivity obtained from the Fourier transform of the \( X - T \) data (see figures 11 and 17 of DeAngelis et al 1993a). This direction-selectivity index, which is correlated with space–time inseparability, shows a unimodal distribution.
Figure 10a is a summary of the time dependence of $\delta \Phi$ for the population of sixty-five binocular simple cells that we have studied. For each cell, we computed the phase difference as a function of time, as shown in figure 9e. We then subtracted the time-average value of $\delta \Phi$ from each curve, and plotted the resulting curves for all cells on the same set of axes. Thus, figure 10a shows the deviation of $\delta \Phi$ around the time-average value for each cell. For the vast majority of cells, $\delta \Phi$ deviates by no more than ±30° from the time-average value. However, there are several cells for which $\delta \Phi$ clearly changes as a function of time. For these cells, we may obtain very different estimates of $\delta \Phi$, depending on the value of $T$ that we choose for analysis. It may be argued that this time dependence accounts for some of the large phase differences that we have observed among cells tuned to orientations near vertical (figure 7b). However, it can be seen from figure 10b that this is not the case.

Figure 10b shows the standard deviation of $\delta \Phi$ (over time) plotted against the time-average value of $\delta \Phi$ for each cell. If $\delta \Phi$ is approximately constant over time, then the standard deviation will be relatively small. If $\delta \Phi$ changes progressively with time, or if there is a great deal of variability in $\delta \Phi$ over time, then the standard deviation will be relatively large. It can be seen from figure 10b that most cells with large positive or negative (time-average) phase differences have small standard deviations. There is no tendency for the standard deviation of $\delta \Phi$ to increase with the absolute (time-average) value of $\delta \Phi$. Thus, our findings concerning the orientation dependence of $[\delta \Phi]$ (figure 7b) remain valid when one incorporates the temporal dimension of the RFs. Note that, for most cells, the time-average value of $\delta \Phi$ is very similar to the value obtained at the optimal reverse-correlation delay (median difference = 9°), the latter of which was used in figures 5–8.

Overall, it can be seen from figure 10 that the phase difference between RF profiles for the two eyes does not change substantially over time for most binocular simple cells. If one assumes that simple cells encode binocular disparity through RF phase (figure 2c), this result means that the preferred disparity of most cells does not change as a function of time, even though there may be large changes in the absolute phase of the RFs. For a few cells, $\delta \Phi$ does change gradually over time; these cells may respond best in situations where binocular disparity changes as a function of time (eg when an object moves in depth; see section 4.4).

Figure 10. Summary of changes in $\delta \Phi$ over time for sixty-five simple cells. (a) Deviation, $D$, of $\delta \Phi$ around the time-average value for each cell as a function of time. (b) The standard deviation of $\delta \Phi$ (over time) is plotted against the time-average value for each simple cell. Filled triangles represent cells that have preferred orientations within 30° of vertical; the remaining cells are denoted by filled circles. The scatter of points shows that the deviation in $\delta \Phi$ over time is not correlated with the time-average value of $\delta \Phi$: thus, large phase differences between RFs for the two eyes (figure 7b) are not an artifact of the time dependence of $\delta \Phi$.

3.4 Phase encoding and classes of disparity-tuned cells

In the previous sections, we have shown that neurons tuned to orientations near vertical exhibit a wide range of phase differences between RF profiles for the two eyes. We suggest that this specialization of RF structure is used by the brain to encode binocular disparities in the visual image. Since this idea represents a major modification to the standard idea of the neural basis of stereopsis (eg Barlow et al 1967; von der Heydt et al 1976; Maske et al 1984; Poggio et al 1988), it is important to consider how the phase-encoding scheme proposed here relates to the traditional types of disparity tuning exhibited by cells in the visual cortex.

From the results of experiments performed on alert monkeys (Poggio and Fischer 1977; Poggio 1984; Poggio et al 1988), four basic patterns of disparity tuning have
been identified: tuned excitatory, tuned inhibitory, near, and far. Tuned-excitatory cells respond best to objects that are located within the plane of fixation (ie when the image has zero disparity). Conversely, the response of tuned-inhibitory cells is suppressed when the visual image has zero disparity. Near or far cells respond optimally to objects that are in front of or behind the plane of fixation, respectively. Thus, near cells prefer negative (crossed) disparities, whereas far cells prefer positive (or uncrossed) disparities. Studies of disparity tuning in cat visual cortex (von der Heydt et al 1978; Fischer and Kruger 1979; Ferster 1981; LeVay and Voigt 1988) have also uncovered these four basic disparity response patterns. However, recent studies in both the monkey (Poggio et al 1988) and the cat (LeVay and Voigt 1988) suggest that there are patterns of disparity selectivity intermediate between the tuned-excitatory and near/far types. These intermediate types have been called tuned near and tuned far by Poggio et al (1988). It appears, therefore, that there is a continuum, rather than discrete categories, of disparity response patterns.

Let us now consider whether a phase-encoding scheme can predict the types of disparity tuning described above. Nomura et al (1990) have proposed a model of the simple cell which incorporates the idea of mismatched left and right RFs to account for the basic types of disparity tuning described in the literature. The model of Nomura et al is essentially identical to the phase-encoding scheme that we have proposed (DeAngelis et al 1991; see figure 2c), in that the left and right RFs of a simple cell are allowed to have different shapes, but their centers (ie Gaussian envelopes) are aligned in retinal correspondence (ie at zero disparity). Figure 11a illustrates the basic model for a binocular simple cell. It is assumed that the simple cell sums its inputs from the left and right eyes in a linear fashion, and produces output through a threshold nonlinearity. These assumptions are supported by the results of experiments in which simple cells are stimulated with dichoptically presented sinusoidal gratings (Ohzawa and Freeman 1986).

Figure 11b shows predictions of the binocular-simple-cell model for various phase differences between the left and right RF profiles (see also Nomura et al 1990). When the phase difference, $\delta \Phi$, is $0^\circ$ (top row of figure 11b), the disparity-tuning curve of the model cell has a peak at zero disparity. This is similar to the disparity tuning of a tuned-excitatory cell. In fact, the disparity-tuning curve predicted by the model for $\delta \Phi = 0^\circ$ is quite similar in shape to disparity-sensitivity profiles reported in the literature for some tuned-excitatory cells [see figure 3b of LeVay and Voigt (1988) and figure 1 of Ferster (1981)]. When $\delta \Phi = 180^\circ$ (second row of figure 11b), the model predicts a disparity-tuning function that is analogous to the response of tuned-inhibitory cells. When $\delta \Phi = 90^\circ$ or $\delta \Phi = -90^\circ$ (bottom two rows of figure 11b), the predicted disparity tuning has a peak at negative (crossed) or positive (uncrossed) disparities, respectively. Thus, the prediction of the model can be similar to the disparity tuning of a near cell ($\delta \Phi = 90^\circ$) or a far cell ($\delta \Phi = -90^\circ$). Note that the disparity-tuning curves predicted by the model for phase differences of $\pm 90^\circ$ are asymmetric about zero disparity. On one side of zero disparity, there is an excitatory peak; on the other side, there is a trough of inhibition. These predictions are actually quite similar in shape to disparity-sensitivity profiles of near (or tuned-near) and far (or tuned-far) cells that have been reported in the literature [see figure 10 of Poggio et al (1984), figure 8 of Poggio et al (1988), and figures 4a and 4c of LeVay and Voigt (1988)].

It can be seen from figure 11b that the predictions of a simple model, in which binocular disparity is encoded through RF phase, are consistent with the traditional types of disparity tuning reported in the literature, provided that $\delta \Phi$ takes on a broad range of values in a population of cells. It is clear from figure 7b that this requirement holds for simple cells with preferred orientations close to vertical, implying that horizontal disparities are encoded by differences in RF phase between the two eyes.

A large range of $\delta \Phi$ is not necessary for cells tuned to orientations near horizontal, because the range of naturally occurring vertical disparities is relatively small. It should be noted that figure 11b only shows predictions of the model for four values of $\delta \Phi$ ($0^\circ$, $180^\circ$, $90^\circ$, and $-90^\circ$). For values of $\delta \Phi$ that fall in between these examples, the model will predict disparity-tuning curves that are intermediate between the standard types defined by Poggio et al (1988). Thus, the model predicts that some cells in the cat's striate cortex should exhibit disparity selectivities that are not easily categorized into one of the standard types. Indeed, LeVay and Voigt (1988) report that some cells have disparity-sensitivity profiles that are difficult to categorize, and they suggest that the traditional discrete categories of disparity tuning are not adequate. Rather, they conclude that cortical cells display a continuum of disparity-tuning properties, a finding that is consistent with the phase-encoding scheme described here.

![Figure 11](image-url)

**Figure 11.** Relationship between the phase-encoding scheme of figure 2c and the traditional types of disparity tuning exhibited by neurons in the visual cortex. (a) A basic model for the binocular simple cell. In this model, the cell linearly sums its excitatory inputs from the two eyes and produces output through a threshold nonlinearity. (b) Predictions of disparity tuning based upon phase differences. Curves in the left-hand panel are Gabor functions that represent RF profiles for the left and right eyes of a model simple cell. As in figure 2c, the left and right RF profiles may be different in shape, but their centers are located at corresponding points on the two retinas. Curves in the right-hand panel are predicted disparity-tuning functions. These curves give the average response of the model cell to all left-right pairs of stimulus positions that correspond to a particular disparity. Thus, these predicted disparity-tuning curves are analogous to those obtained experimentally by sweeping a pair of bright bars across the RFs at various disparities and computing the average response rate (eg Ferster 1981).
4 Discussion

Previously, it has been reported (Hubel and Wiesel 1962; Maske et al 1984) that the left and right RFs of binocular simple cells are nearly identical in structure. On the basis of these observations, it has generally been assumed (eg Barlow et al 1967; Joshua and Bishop 1970; von der Heydt et al 1978; Maske et al 1984) that horizontal binocular disparities are encoded by a group of cells with RFs (see figure 2a) that are spatially displaced from the point of retinal correspondence (ie zero disparity). In this study, we have proposed an alternative scheme for disparity encoding, in which the left and right RF profiles of simple cells may differ in shape (or phase), but are centered at corresponding retinal locations (see also Nomura et al 1990).

The data we report here show that simple cells tuned to orientations near vertical exhibit a broad range of phase differences between RFs for the left and right eyes. In contrast, cells tuned to orientations near horizontal tend to have matched RFs for the two eyes. For most cells with space–time inseparable RFs, the temporal dynamics of RF structure are matched for the two eyes, so that the interocular phase difference remains constant over time. We also demonstrate here that phase differences between RF profiles for the two eyes can account for the traditional types of disparity tuning observed in studies of the visual cortex (eg Poggio and Fischer 1977; Fischer and Kruger 1979; Ferster 1981; LeVay and Voigt 1988; Poggio et al 1988). Overall, these findings are consistent with the idea (see figure 2c; see also Nomura et al 1990) that simple cells encode horizontal disparity through differences in the phases, rather than through spatial displacement, of RFs for the two eyes.

It should be noted, however, that our findings do not rule out the possibility that positional disparities (or incongruities) are also used in the neural processing of depth information (as discussed below), because positional incongruities between the left and right RF envelopes have not been measured in this study. In an anesthetized, paralyzed animal preparation, it is quite difficult to determine whether or not the RFs of a binocular neuron are aligned in retinal correspondence (ie at zero disparity). Several researchers (eg Barlow et al 1967; Joshua and Bishop 1970; Hubel and Wiesel 1973; von der Heydt et al 1978; Ferster 1981) have attempted to measure positional disparities (or incongruities) between RFs for the two eyes, but the findings are not consistent. These attempts are generally limited by inadequate methods for determining the location of the centre of the RFs and by problems associated with the monitoring of eye position (see Ferster 1981; LeVay and Voigt 1988). Future work will be focused on assessing, with adequate precision, the relative contributions of phase and position disparities.

4.1 Relationship to previous findings

As noted above, our findings concerning the similarity of RF profiles for the left and right eyes are not in agreement with previous work. Hubel and Wiesel (1962) originally reported that the left and right RFs of binocular simple cells are closely matched in structure. However, because Hubel and Wiesel mapped RFs qualitatively with spots of light, they may have been unable to discern differences between the shapes of the two RFs. In addition, many simple cells have RFs that are not space–time separable (see figure 9), and it is not clear how the static RF maps of Hubel and Wiesel (1962) relate to the dynamic spatiotemporal RF structure of these cells. As discussed elsewhere (DeAngelis et al 1993a), estimates of RF phase obtained from a static spatial RF profile may be inaccurate.

Maske et al (1984) also report that the left and right RFs of simple cells have a similar spatial organization. However, there are methodological problems with this study, as well. Maske et al obtained RF profiles by recording peristimulus time histograms in response to moving bright and dark bars. With this method, the shape of RF profiles is likely to be distorted by the effects of sequential stimulation of neighboring subregions. In other words, spatial and temporal response properties are confounded. This is especially problematic for simple cells with direction-selective (ie space–time inseparable) RFs, because a unique spatial RF profile does not exist for these cells (see figure 9). Another problem with the study of Maske et al (1984) is that a quantitative method (such as fitting Gabor functions) was not used to compare the left and right RF profiles. Thus, it may have been difficult to discern phase differences of 90° or less. In fact, some of the data reported by Maske et al (see figure 4) exhibit a clear phase difference between RFs for the two eyes. Moreover, data are presented for only fifteen simple cells in their study, and orientation preferences are not given.

4.2 Population coding of binocular disparity

In figure 11, we have shown how differences in RF structure between the two eyes may yield various types of disparity tuning. We now consider how a population of neurons can faithfully encode, on the basis of phase, the entire range of horizontal disparities to which the visual system is sensitive. For the cat, behavioral studies (Packwood and Gordon 1975) indicate that horizontal disparities as large as 5 deg can be utilized to perceive depth. If simple cells encode disparity through RF phase, then the precision with which each neuron can signal disparity is commensurate with its optimal spatial frequency. Moreover, if the two RFs of a cortical neuron are aligned in retinal correspondence (at zero disparity), then the range of disparities that a given cell can encode is determined by the size of its RFs. Previously, it has been reported (Maffei and Fiorentini 1977) that the size of cortical RFs is inversely related to their preferred spatial frequency. This suggests that the average number of bright-excitatory and dark-inhibitory subregions within the RFs of simple cells remains constant as a function of spatial frequency (ie RFs simply scale), so that cells tuned to low spatial frequencies generally have large RFs and cells tuned to high spatial frequencies generally have small RFs. Thus, in a phase-encoding scheme (figure 2c), the range of disparities that can be encoded by binocular simple cells should be inversely related to spatial frequency (see also Marr and Poggio 1979).

This relationship between RF size (or spatial frequency) and disparity selectivity is illustrated schematically in figure 12 (solid curves). This figure shows disparity-tuning curves of hypothetical simple cells at each of five different spatial-frequency scales, ranging from low spatial frequency (top) to high spatial frequency (bottom). For simplicity, there are just three different simple cells at each spatial frequency and each simple cell has a disparity-tuning curve, shown as a Gaussian (solid curves), which is determined by the relative phases of the left and right RFs. One (tuned-excitatory) cell has a preference for zero disparity by virtue of having matched RFs for the two eyes (θ = 0). The two remaining cells, at each spatial frequency scale, are tuned to crossed or uncrossed disparities by having θ = 90° or θ = -90°, respectively. It is assumed here (solid curves) that RFs at different spatial frequencies are scaled replicas of one another (ie RF size changes, but RF periodicity does not).

At the lowest spatial-frequency scale (top of figure 12), each simple cell is broadly tuned for disparity. Thus, each cell can reliably discriminate only coarse disparities, because small changes in disparity produce only a small change in the response level of the neuron. On the other hand, this pool of cells tuned to low spatial frequencies can encode a fairly broad range of disparities. At the highest spatial frequency scale (bottom of figure 12), each neuron can reliably discriminate fine disparities, but the range of disparities that the cells can collectively encode is rather small. Thus, there is a trade-off between the range of disparities that can be encoded in terms of phase and the precision with which these disparities can be signalled. However, by pooling
the responses of cells over all spatial-frequency scales, it should be possible for the visual system to encode faithfully a large range of horizontal disparities.

Figure 12. A scheme for encoding binocular disparities in terms of RF phase at multiple spatial scales. Idealized disparity-tuning curves are shown at each of five spatial-frequency scales, ranging from low (top) to high (bottom). For each set of curves, the abscissa is binocular disparity, and the ordinate represents response strength. For the sake of clarity, these tuning curves are shown as Gaussians rather than more realistic shapes (as in figure 11b). Also, only three cells, having phase differences of $\delta \Phi = -90^\circ$, $0^\circ$, and $90^\circ$, are represented at each spatial frequency scale (solid curves). These three classes of cells represent tuned-excitatory ($\delta \Phi = 0^\circ$), near ($\delta \Phi = 90^\circ$), and far ($\delta \Phi = -90^\circ$) types. Cells tuned to low spatial frequencies have large RFs and can encode a large range of disparities. Cells that are tuned to high spatial frequencies encode a much more limited range of disparities, but with higher precision. Dashed curves at high spatial frequencies show that a larger range of binocular disparities can be encoded if positional disparities between RFs for the two eyes are allowed (as shown in figure 2a). The combination of phase differences and positional disparities (incongruities) may be necessary to account for behavioral performance at high spatial frequencies (see section 4.3).

4.3 The size—disparity correlation

The multiscale phase-encoding scheme of figure 12 makes a specific prediction for the way that depth is encoded by the visual system. This model suggests that the range of disparities to which the visual system is sensitive depends upon the spatial-frequency content of the image. This relationship is referred to as the size—disparity correlation (eg Schor and Wood 1983). It arises in our scheme because coarse disparities must be signalled by cells tuned to low spatial frequencies, whereas fine disparities are signalled by cells tuned to high spatial frequencies (see also Marr and Poggio 1979).

Human psychophysical studies provide limited support for the existence of a size—disparity correlation in stereopsis [see Smallman and MacLeod (1994) for a recent review]. For example, it has been reported that lower disparity limits for stereopsis (ie stereothresholds—Pulliam 1982; Schor and Wood 1983; Schor et al 1984a; Legge and Gu 1989), binocular fusion limits (Schor et al 1984b, 1989), and upper disparity limits for perceived depth (Schor and Wood 1983; Schor et al 1984b; Smallman and MacLeod 1994) all decrease with stimulus spatial frequency. It should be noted, however, that some evidence does not support a size—disparity correlation in stereopsis (eg Frishby and Mayhew 1978; Mayhew and Frishby 1979; Badcock and Schor 1985).

Among studies that support the existence of a size—disparity correlation in stereopsis, there is considerable evidence that this correlation breaks down at high spatial frequencies. For example, Schor et al (1984b, 1989) have examined binocular fusion limits as a function of the spatial frequency of three types of visual stimuli. Their results show that, for spatial frequencies below ~2 cycles deg$^{-1}$, fusion limits decrease in inverse proportion to the peak spatial frequency of the stimulus (on log-log coordinates), as predicted by a phase-encoding model. However, the size—disparity correlation breaks down at high spatial frequencies (>2 cycles deg$^{-1}$), and, in some cases, the fusion limit becomes roughly constant at ~10 min arc (see figure 1 of Schor et al 1984b). Similar behaviour at high spatial frequencies has been reported for stereothresholds (eg Schor and Wood 1983; Legge and Gu 1989) and for upper depth limits (eg Schor et al 1984b). The recent data of Smallman and MacLeod (1994) show a reduction in slope of the size—disparity relationship at high spatial frequencies, but do not suggest that the relationship flattens out.

Although it is not known whether the results of human psychophysical studies are applicable to binocular vision in the cat, these psychophysical results suggest that a disparity-encoding mechanism based solely on phase may be insufficient to explain behavioral performance. To address this issue, consider that a perfect size—disparity correlation is expected for stereopsis only if the following two conditions are met: (i) disparity is encoded in terms of phase, with no positional disparities between RFs for the two eyes, and (ii) RFs simply scale as a function of spatial frequency (ie RF periodicity remains constant). If either of these conditions is not met, then one may not expect an inverse relationship (slope = -1) between disparity and spatial frequency (on log-log coordinates).

If we consider these requirements, one possible explanation for breakdown of the size—disparity correlation is that neurons tuned to high spatial frequencies exhibit positional disparities (or incongruities) between RFs for the two eyes, in addition to having phase differences. Cells tuned to low spatial frequencies may encode disparity strictly on the basis of phase, whereas neurons tuned to high spatial frequencies may utilize both phase and position encoding. The implications of such a hybrid disparity-encoding scheme are illustrated schematically by the dashed curves in figure 12. If simple cells tuned to high spatial frequencies exhibit both phase and position disparities between RFs for the two eyes, these neurons can encode large disparities that are carried by high-spatial-frequency components in the image. Recent theoretical work (Jacobsen et al 1993) suggests that a hybrid disparity-encoding scheme, incorporating both phase differences and position disparities, may be necessary to encode reliably nonzero disparities. Alternatively, it is possible that there is a fixed degree of positional jitter (ie small random incongruities) between RFs for cells at all spatial-frequency scales. At high spatial frequencies, this jitter may become substantial relative to the spatial period of the RFs, thus interfering with phase encoding. This could conceivably explain, for example, why stereocuity plateaus (ie the size—disparity relationship flattens out) at high spatial frequencies (eg Schor and Wood 1983; Legge and Gu 1989).

A second possible explanation for psychophysically observed deviations from the size—disparity correlation is that RFs do not simply scale with spatial frequency. To address this possibility we determined the number of subregions within the RFs of simple cells. Figure 13a shows the number of RF subregions plotted against preferred spatial frequency for each of the sixty-five binocular simple cells studied here.
Linear regression (solid line) shows that there is a significant trend ($r = 0.48$, $p < 0.001$) for the number of RF subregions to increase with spatial frequency. Thus, cells tuned to high spatial frequencies tend to have more-periodic RFs than cells tuned to low spatial frequencies. This result is consistent with the finding that spatial-frequency-tuning bandwidths (in octaves) decrease somewhat with preferred spatial frequency (DeValois et al 1982), but is not consistent with the findings of Maffei and Fiorentini (1977), who report that RFs simply scale with spatial frequency. Thus, the data of figure 13a do not agree with the requirement (dashed line) for existence of a perfect size-disparity correlation. Despite the trend shown in figure 13a, RF size does tend to decrease with preferred spatial frequency (figure 13b, solid line; $r = 0.52$, $p < 0.001$ by linear regression in the log-log domain), but not as steeply as would be expected (figure 13b, dashed line) if RFs simply scaled with spatial frequency.

The finding that RF periodicity increases with spatial frequency may account for a reduction in slope of the size-disparity relationship at high spatial frequencies (Smallman and MacLeod 1994), because cells with more-periodic RFs can potentially encode a larger range of disparities (at a given spatial-frequency scale). However, the data of figure 13 are unlikely to account for the observation that, for some subjects, the size-disparity relationship becomes flat at high spatial frequencies (Schor and Wood 1983, Schor et al 1984b, 1989). In any case, we can conclude, on the basis of physiology, that it may be inappropriate to expect a perfect size-disparity correlation in behavioral performance.

It should be pointed out that simple cells with more-periodic RFs (ie those tuned to high spatial frequencies) will have multiple peaks in their disparity-tuning curves (see figure 11b). Thus, the responses of these binocular neurons will be ambiguous with regard to disparity, in the same way that their responses to monoptic stimulation will be ambiguous with regard to stimulus position. However, this ambiguity in the disparity domain can be eliminated by combining the outputs of two or more neurons with different preferred disparities, just as the ambiguity with regard to position in monoptic responses can be eliminated by combining the outputs of two or more neurons, having different RF phases (ie sine and cosine phases), that comprise a basis set (eg Sakitt and Barlow 1982).

Last, we note that it is possible that simple cells encode disparity on the basis of phase, whereas cells at some higher level in the visual pathway encode disparity on the basis of positional offsets (for an analogous idea, see DeValois 1982, DeValois and DeValois 1988). At low spatial frequencies, phase-based encoding would constrain fusion limits, whereas position-based encoding would determine fusion limits at high spatial frequencies. To determine the relative roles of phase and position with regard to disparity encoding by simple cells, it will be necessary to make accurate measurements both of the phase and of the position disparities between RFs for the two eyes.

4.4 Phase encoding and space-time inseparability

We have recently reported that there is a correlation between space-time inseparability and the degree of direction selectivity exhibited by simple cells. Cells with space-time separable RFs tend to exhibit little direction selectivity in response to drifting sinusoidal gratings, whereas cells with clearly inseparable $X-T$ profiles tend to be strongly direction selective (see figure 9 of DeAngelis et al 1993b). Thus, it appears that space-time inseparability forms the basis for direction selectivity in simple cells (see also McLean and Palmer 1989; Reid et al 1991; Heeger 1993).

For simple cells with inseparable $X-T$ profiles, both excitatory and inhibitory subregions of the RF appear as tilted slabs in the space-time domain. As shown in figure 9, this spatiotemporal organization can be described as a change in the spatial phase of the RF over time. These temporal dynamics of RF structure would appear to make it difficult for simple cells to encode binocular disparity through differences in phase between RFs for the two eyes. In order to perceive the depth of stationary objects, it seems necessary to have neurons that respond to a fixed disparity over time. For binocular cells with space-time inseparable RFs, this requirement can only be met if changes in phase over time are matched for the left and right RFs, so that the interocular phase difference remains constant. The results reported here (see figures 9 and 10) show that changes in phase over time are, indeed, matched for the two eyes. This matching of temporal dynamics for the two RFs allows the phase-based mechanism for disparity encoding to coexist with the linear mechanism for direction selectivity.

For a few of the sixty-five simple cells studied here, changes in RF phase over time are not matched for the two eyes. This situation generally arises when RFs for the two eyes have somewhat different velocity preferences. As a result, the phase difference, $\delta \Phi$, changes over time for these cells (see figure 10a). If simple cells encode depth through interocular differences in RF phase, then these cells may have a special functional role. When an object moves in depth (ie toward or away from an observer), the horizontal disparity between images on the two retinas changes over time. Thus, it is plausible that motion in depth is encoded by simple cells for which $\delta \Phi$ changes with time (although this possibility has not been tested directly in this study). Neurons that have different velocity selectivities for the two eyes have previously been recorded from the extrastriate visual cortex of cats (Cynader and Regan 1978, 1982; Spieler et al 1990) and monkeys (Poggio and Talbot 1981). It has been suggested (eg Cynader and Regan 1978) that these cells are specialized to signal object
motion in depth. The findings reported here suggest that only a small proportion of simple cells in the cat’s striate cortex may be suited for encoding motion in depth.

4.5 The role of complex cells

Thus far, we have focused exclusively on the role of simple cells in processing binocular disparity. Because simple cells receive most of the geniculate input to the striate cortex (for reviews, see Gilbert and Wiesel 1981; Gilbert 1983; Martin 1984), it is reasonable to suppose that simple cells are responsible for the initial encoding of binocular disparity. Indeed, we have shown that the RFs of binocular simple cells are well suited to encode horizontal disparities in terms of phase at multiple spatial scales.

Although simple cells may serve to encode disparity, these neurons are not well suited to access disparity by detectors. There are several limitations of a phase-based mechanism for disparity encoding, and these limitations suggest that other neurons in the visual cortex must perform a more advanced processing of disparity signals. To illustrate these shortcomings, let us consider some properties of an ideal disparity detector (see also Ohzawa et al. 1990). First, the preferred disparity of an ideal detector should be constant for all stimulus positions within its RF. In other words, the ideal detector should not confound disparity information with position information. Simple cells clearly do not meet this requirement. Suppose that a bright bar stimulus is located within a bright-excitatory subregion in each of the two RFs of a binocular simple cell. This stimulus will elicit a vigorous response. If the binocular disparity is kept constant but the positions of the bright bars are shifted by 180° (so that they now fall within dark-excitatory subregions in each eye), then the response of the cell will be suppressed. Thus, simple cells do not encode disparity independently of position. Another requirement of an ideal detector is that disparity should be signalled independently of contrast. A pair of bright bars (one being presented to each eye) and a pair of dark bars should elicit the same response when presented at a particular disparity, independently of position. Simple cells do not behave this way. If a simple cell responds well to a pair of bright bars, it will not respond well to a pair of dark bars at the same locations in the two RFs. Last, for an ideal disparity detector, incorrect contrast polarity combinations should be ineffective if presented at the optimal disparity for a matched polarity pair. In other words, the combination of a bright bar to one eye and a dark bar to the other should not, for any appropriate pair of positions in the two eyes, elicit a response at the preferred disparity of the detector. This would allow the detector to reject incorrect (ie opposite-polarity) feature matches.

Using a binocular version of the reverse-correlation technique, we have recently mapped the disparity selectivities of simple and complex cells in the cat’s striate cortex (Ohzawa et al. 1990). The results of that study show that many complex cells exhibit the desired properties of a disparity detector, whereas simple cells do not. In particular, we have shown that binocular complex cells signal disparity independently of position and contrast polarity. These complex cells also act to reject incorrect feature matches because they do not respond to mismatched contrast polarities (ie a bright bar to one eye and a dark bar to the other) at the disparity which is optimal for a matched polarity pair. Thus, it is clear that complex cells perform a higher-level processing of binocular disparity signals. Although complex cells exhibit some ideal characteristics for disparity processing, it is not strictly appropriate to label these neurons as disparity detectors, since their responses depend on other stimulus parameters, such as orientation and spatial frequency. However, owing to the nonlinear organization of their RFs (see Ohzawa et al. 1990), complex cells appear to form the first stage in the neural process of disparity detection.

References


Cynader M, Regan D, 1982 “Neurons in cat visual cortex tuned to the direction of motion in depth: effect of position disparity” Vision Research 22 967–982

DeAngelis G C, Freeman R D, Ohzawa I, 1994 “Length and width tuning of neurons in the cat’s primary visual cortex” Journal of Neurophysiology 71 347–374

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

DeAngelis G C, Ohzawa I, Freeman R D, 1993a “Spatiotemporal organization of simple-cell receptive fields in the cat’s striate cortex. I. General characteristics and postnatal development” Journal of Neurophysiology 69 1091–1117

DeAngelis G C, Ohzawa I, Freeman R D, 1993b “Spatiotemporal organization of simple-cell receptive fields in the cat’s striate cortex. II. Linearity of temporal and spatial summation” Journal of Neurophysiology 69 1118–1135


Neuronal mechanisms underlying stereopsis


Ohtsza I, Freeman R D, 1986 “The binocular organization of single cells in the cat’s visual cortex” *Journal of Neurophysiology* 56 221-242

Ohtsza I, DeAngelis G C, Freeman R D, 1990 “Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors” *Science* 249 1037-1041

Packwood J, Gordon B, 1975 “Stereopsis in normal domestic cat, siamese cat, and cats raised with alternating monocular occlusion” *Journal of Neurophysiology* 38 1485-1499


Pullman K, 1982 “Spatial frequency analysis of three-dimensional vision” *Proceedings of the Society of Photo-Optical Instrumentation Engineers* 303 17-23


Sakai T, Barlow H B, 1982 “A model for the economical encoding of the visual image in cerebral cortex” *Biological Cybernetics* 43 97-108

Schor C M, Wood I, 1983 “Disparity range for local stereopsis as a function of luminance spatial frequency” *Vision Research* 23 1649-1654

Schor C M, Heckmann T, Tyler C W, 1989 “Binocular fusion limits are independent of contrast, luminance gradient and component phases” *Vision Research* 29 821-835


Schor C M, Wood I, Ogawa J, 1984b “Binocular sensory fusion is limited by spatial resolution” *Vision Research* 24 661-665


Smallman H S, MacLeod D I A, 1994 “Size-disparity correlation in stereopsis at contrast threshold” *Journal of the Optical Society of America* A 11 2169-2183

Spileers W, Orban G A, Gulyás B, Maes H, 1990 “Selectivity of cat Area 18 neurons for direction and speed of drift” *Journal of Neurophysiology* 63 936-954

Wheatstone C, 1838 “Contributions to the physiology of vision. Part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision” *Philosophical Transactions of the Royal Society of London* 2 371-393