Neural mechanisms underlying binocular fusion and stereopsis: Position vs. phase

AKIYUKI ANZAI, IZUMI OHZAWA, AND RALPH D. FREEMAN*

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

Communicated by Richard M. Held, Massachusetts Institute of Technology, Cambridge, MA, March 3, 1997 (received for review September 30, 1996)

Abstract

The visual system utilizes binocular disparity to discriminate the relative depth of objects in space. Since the striate cortex is the first site along the central visual pathways at which signals from the left and right eyes converge onto a single neuron, encoding of binocular disparity is thought to begin in this region. There are two possible mechanisms for encoding binocular disparity through simple cells in the striate cortex: a difference in receptive field (RF) position between the two eyes (RF position disparity) and a difference in RF profile between the two eyes (RF phase disparity). Although there have been studies supporting each of the two encoding mechanisms, both mechanisms have not been examined in a single study. Therefore, the relative roles of the two mechanisms have not been determined. To address this issue, we have mapped left and right eye RFs of simple cells in the cat’s striate cortex using binary m-sequence noise, and then we have estimated RF position and phase disparities. We find that RF position disparities are generally limited to small values that are not sufficient to encode large binocular disparities. In contrast, RF phase disparities cover a wide range of binocular disparities and exhibit dependencies on orientation and spatial frequency in a manner expected for a mechanism that encodes binocular disparity. These results indicate that binocular disparity is mainly encoded through RF phase disparity. However, RF position disparity may play a significant role for cells with high spatial frequency selectivity, which are constrained to small RF phase disparities.

An image of an object either in front of or behind the point of visual fixation projects onto slightly different locations of the retina in the two eyes. This difference, binocular disparity, is by itself a sufficient cue for our perception of depth (1, 2). Since the discovery that most neurons in the striate cortex of cats and monkeys are selective to binocular disparity (3–6), how these neurons encode binocular disparity has become an important issue for understanding neural mechanisms of binocular fusion and stereopsis (7–17).

There are two hypotheses for how cortical neurons encode binocular disparity. The traditional view, illustrated in Fig. 1a, is that left and right eye receptive fields (RFs) of a neuron have the same spatial profile, but their positions are not necessarily at retinal correspondence, creating RF position disparity through which binocular disparity can be encoded (7–9). In this scheme, the range of binocular disparity that can be encoded is limited by the range of RF position disparity.

Alternatively, binocular disparity can be encoded through a difference in RF profile or phase between the two eyes without RF position disparity (10–17) as shown in Fig. 1b. Since, by definition, RF phase disparity is limited to a range between ±180° phase angle, the range of binocular disparity that can be encoded with this scheme is proportional to the period of the RF or inversely proportional to the spatial frequency of the RF.

Since previous studies (7–13) have examined only one of the two hypotheses, the relative roles of the two encoding mechanisms have not been determined. It is still not clear whether simple cells encode binocular disparity through both RF position and phase disparities (16, 17) or through either one alone. Therefore, the question of how neurons encode binocular disparity still remains open and cannot be resolved unless one examines both RF position and phase disparities for individual neurons.

In this study, this long-standing issue is finally addressed. To assess the relative contribution of the two mechanisms of disparity encoding, we have estimated RF position and phase disparities for binocular simple cells in the cat’s striate cortex. We report here that the range of binocular disparity that can be encoded through RF position disparity is relatively small compared with that through RF phase disparity except for cells tuned to high spatial frequencies. In addition, RF phase disparity, not RF position disparity, exhibits characteristics that are expected for a mechanism of encoding binocular disparity. Therefore, we conclude that binocular disparity is mainly encoded through RF phase disparity. However, RF position disparity may play an important role in encoding binocular disparity for cells tuned to high spatial frequencies, which have necessarily small RF phase disparities.

Materials and Methods

Experimental Protocol. Extracellular recordings are made from simple cells in the striate cortex of anesthetized and paralyzed adult cats. Details of surgical procedures and the physiological recording setup are described elsewhere (13). Two tungsten-in-glass electrodes (18) are used to record from one or more well-isolated binocular simple cells simultaneously. Drifting sinusoidal gratings are used to obtain orientation and spatial frequency tuning for cells. Classification of simple cells is based on RF organization (19) and on the degree of temporal modulation in response to gratings (20). After orientation and spatial frequency tuning are determined, detailed RF maps are obtained and RF position and phase disparities are estimated.

RF Mapping. RFs of cells are mapped with white noise stimuli generated according to binary m-sequences (21, 22). Two kinds of stimulus configuration are used: Two-dimensional (2D) and one-dimensional (1D) patterns (Fig. 2).

PNAS is available online at http://www.pnas.org.

Abstract

The visual system utilizes binocular disparity to discriminate the relative depth of objects in space. Since the striate cortex is the first site along the central visual pathways at which signals from the left and right eyes converge onto a single neuron, encoding of binocular disparity is thought to begin in this region. There are two possible mechanisms for encoding binocular disparity through simple cells in the striate cortex: a difference in receptive field (RF) position between the two eyes (RF position disparity) and a difference in RF profile between the two eyes (RF phase disparity). Although there have been studies supporting each of the two encoding mechanisms, both mechanisms have not been examined in a single study. Therefore, the relative roles of the two mechanisms have not been determined. To address this issue, we have mapped left and right eye RFs of simple cells in the cat’s striate cortex using binary m-sequence noise, and then we have estimated RF position and phase disparities. We find that RF position disparities are generally limited to small values that are not sufficient to encode large binocular disparities. In contrast, RF phase disparities cover a wide range of binocular disparities and exhibit dependencies on orientation and spatial frequency in a manner expected for a mechanism that encodes binocular disparity. These results indicate that binocular disparity is mainly encoded through RF phase disparity. However, RF position disparity may play a significant role for cells with high spatial frequency selectivity, which are constrained to small RF phase disparities.

An image of an object either in front of or behind the point of visual fixation projects onto slightly different locations of the retina in the two eyes. This difference, binocular disparity, is by itself a sufficient cue for our perception of depth (1, 2). Since the discovery that most neurons in the striate cortex of cats and monkeys are selective to binocular disparity (3–6), how these neurons encode binocular disparity has become an important issue for understanding neural mechanisms of binocular fusion and stereopsis (7–17).

There are two hypotheses for how cortical neurons encode binocular disparity. The traditional view, illustrated in Fig. 1a, is that left and right eye receptive fields (RFs) of a neuron have the same spatial profile, but their positions are not necessarily at retinal correspondence, creating RF position disparity through which binocular disparity can be encoded (7–9). In this scheme, the range of binocular disparity that can be encoded is limited by the range of RF position disparity.

Alternatively, binocular disparity can be encoded through a difference in RF profile or phase between the two eyes without RF position disparity (10–17) as shown in Fig. 1b. Since, by definition, RF phase disparity is limited to a range between ±180° phase angle, the range of binocular disparity that can be encoded with this scheme is proportional to the period of the RF or inversely proportional to the spatial frequency of the RF.

Since previous studies (7–13) have examined only one of the two hypotheses, the relative roles of the two encoding mechanisms have not been determined. It is still not clear whether simple cells encode binocular disparity through both RF position and phase disparities (16, 17) or through either one alone. Therefore, the question of how neurons encode binocular disparity still remains open and cannot be resolved unless one examines both RF position and phase disparities for individual neurons.

In this study, this long-standing issue is finally addressed. To assess the relative contribution of the two mechanisms of disparity encoding, we have estimated RF position and phase disparities for binocular simple cells in the cat’s striate cortex. We report here that the range of binocular disparity that can be encoded through RF position disparity is relatively small compared with that through RF phase disparity except for cells tuned to high spatial frequencies. In addition, RF phase disparity, not RF position disparity, exhibits characteristics that are expected for a mechanism of encoding binocular disparity. Therefore, we conclude that binocular disparity is mainly encoded through RF phase disparity. However, RF position disparity may play an important role in encoding binocular disparity for cells tuned to high spatial frequencies, which have necessarily small RF phase disparities.

Materials and Methods

Experimental Protocol. Extracellular recordings are made from simple cells in the striate cortex of anesthetized and paralyzed adult cats. Details of surgical procedures and the physiological recording setup are described elsewhere (13). Two tungsten-in-glass electrodes (18) are used to record from one or more well-isolated binocular simple cells simultaneously. Drifting sinusoidal gratings are used to obtain orientation and spatial frequency tuning for cells. Classification of simple cells is based on RF organization (19) and on the degree of temporal modulation in response to gratings (20). After orientation and spatial frequency tuning are determined, detailed RF maps are obtained and RF position and phase disparities are estimated.

RF Mapping. RFs of cells are mapped with white noise stimuli generated according to binary m-sequences (21, 22). Two kinds of stimulus configuration are used: Two-dimensional (2D) and one-dimensional (1D) patterns (Fig. 2).

Abbreviations: RF, receptive field; 1D, one-dimensional; 2D, two-dimensional.

To whom reprint requests should be addressed. e-mail: freeman@pinoko.berkeley.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Copyright © 1997 by THE NATIONAL ACADEMY OF SCIENCES OF THE USA

0027-8424/97/0945438-652.00/0

PNAS is available online at http://www.pnas.org.

September 30, 1996

Communicated by Richard M. Held, Massachusetts Institute of Technology, Cambridge, MA, March 3, 1997 (received for review September 30, 1996)
an optimized m-sequence mapping (21). The size of the stimulus patch is adjusted so that it is large enough to cover the entire RF and the width of each element in the patch is no more than one fourth of the cells’ optimal spatial period (the inverse of the optimal spatial frequency) for gratings.

A stimulus patch is presented on each of the two cathode ray tube (CRT) displays for independent stimulation of the two eyes. The orientation of the patch is aligned at the cells’ optimal orientation for gratings. This is especially important for 1D mapping. When more than one cell is recorded simultaneously, optimal orientations of all cells have to be the same in order for 1D mapping to be valid. There is no such constraint for 2D mapping. The luminance of each element in the patches is modulated every 40 msec according to m-sequences and takes on binary values, either +18cd/m² or −18cd/m² around the mean luminance of the CRT displays (20cd/m²). The m-sequences used are identical for all elements except that they are temporally shifted by at least 2.5 sec from one another so that the luminance modulation of each element is uncorrelated in space and time within this period, i.e., the stimulus is “white” within the spatio-temporal passband of simple cells. A spike train for each cell is cross-correlated with the stimulus sequence to obtain RF maps (21).

**Data Analysis.** Each pair of left and right eye RFs, at the optimal cross-correlation delay, is fitted with a Gabor function (23), the product of a Gaussian envelope and a sinusoid, to obtain the RF center coordinate (the center coordinate of a Gaussian envelope) as well as RF phase (the phase of a sinusoid). RF phase disparity is obtained as a difference between left and right RF phases, measured in visual angle rather than phase angle so that it can be directly comparable to RF position disparity.

To estimate RF position disparity, we must design a method that is feasible for a paralyzed animal. In a paralyzed preparation, eye muscles are relaxed and the visual axes of the eyes deviate from a normal fixation position. As a consequence, it is very difficult to locate corresponding points on the retinae and use them to measure RF position disparity (see Fig. 1a). Instead, we have estimated RF position disparity using a reference-cell method (24–26) (Fig. 3a). Details of the method are described in the legend of Fig. 3a. Briefly, RFs of two or more cells are mapped simultaneously. For each recording, cells are grouped in distinct pairs. One member of each pair, which is chosen arbitrarily, is regarded as a reference cell and RF position disparity of the other member is measured with respect to the RF position of the reference cell. That is, RF position disparity of a cell is obtained as the distance in visual angle between the centers of left and right eye RFs while the RF position disparity of a reference cell is assumed to be 0. In other words, the RF position disparity measured here is the relative position disparity of one cell to that of a reference cell. However, as illustrated in Fig. 3b, the population distribution for true position disparity can be determined from the distribution of relative position disparities for a population of cells. Furthermore, the true position disparities of individual neurons can be estimated with a specified amount of uncertainty (see the legend of Fig. 3b). Using this method, we have obtained RF position disparity along the direction perpendicular to RF orientation (position disparity-X), which is also the direction in which RF phase disparity is measured. In addition, for 2D RF data, we have determined RF position disparity along the direction parallel to RF orientation (position disparity-Y).

**RESULTS**

We have recorded from 97 binocular simple cells in 14 adult cats and have obtained 2D or 1D (or both) profiles of RFs. Of
these, 48 cells were recorded individually for which RF position disparity could not be determined. The remaining 49 cells were from either pair recordings (20 cases) or trio recordings (three cases). For each cell, a RF phase disparity was obtained. A total of 23 multiple-cell recordings yielded 29 distinct cell pairs and a RF position disparity-X was estimated for each of these pairs using a reference-cell method. Among these pairs, there were 15 cases in which 2D RF maps were obtained so that a RF position disparity-Y was also estimated.

Fig. 4 shows examples of left and right eye RFs for a pair of simple cells recorded simultaneously. The RF map of each eye is shown separately for each cell for clarity. Reference cells are presented as cell-A. (a) An example of 2D RF maps. The solid and dashed contours represent bright- and dark-excitatory regions, respectively. The contours are drawn such that they divide the response amplitude between 0 and either a positive or negative peak, whichever is greater, into seven equally spaced levels. Both cell-A and -B show different RF profiles in the two eyes, indicating RF phase disparities. Phase disparities for cell-A and -B are 0.47° and 0.36° in visual angle, respectively. Position disparity-X and -Y are $-0.10°$ and $-0.05°$ in visual angle, respectively.

(b) Another example of 2D RF maps. Phase disparities for cell-A and -B are $-0.28°$ and $-0.82°$, respectively. Position disparity-X and -Y are 0.32° and 0.05°, respectively. (c) An example of 1D RF profiles. The amplitude of each profile is normalized to its peak. Both cell-A and -B show relatively similar RF profiles in the two eyes. Phase disparities for cell-A and -B are $-0.12$ and $-0.38°$, respectively. Position disparity-X is $-0.32°$.

and right eye RFs for cell-B minus that for cell-A, a reference cell) appear to be relatively small in the examples shown.
Fig. 5 shows histograms of RF position and phase disparities for a population of simple cells. Phase disparity is expressed in terms of visual angle rather than phase angle so that it has the same unit as that of position disparity. Both position and phase disparities are distributed around 0, and most are within ± 1°. This range corresponds roughly to the limits of binocular fusion in cats (27). The standard deviations of the distributions for position disparity-X and -Y are 0.52° and 0.62°, respectively. These values divided by \( \sqrt{2} \), i.e., 0.37 and 0.44, are the estimated standard deviations of the distributions for true position disparities (see Fig. 3b). The SD for the phase disparity distribution is 0.59°, which is 1.6 times greater than that of the distribution for true position disparity-X. Therefore, position disparity is limited to a relatively small extent compared with that of phase disparity.

It is possible that the difference in distributions may be due to differences in the amount of error associated with the estimates of position and phase disparities. To examine this possibility, we have conducted a Monte Carlo simulation to obtain SEs for each disparity estimate. More than 90% of the SEs are <0.25°. Mean values of the SEs for position disparity-X and -Y, and phase disparity are 0.12°, 0.1°, and 0.12°, respectively. Therefore, errors in the disparity estimates are comparable for position and phase disparities, and cannot account for the difference between the distributions of position and phase disparities.

The overall preference of each cell to binocular disparity is determined by the sum of position disparity-X and phase disparity. Therefore, position disparity, though it is small, may still be actively used to make up the cell’s disparity preference by having the same sign as that of phase disparity. However, this is not the case. In Fig. 6, position disparities of individual cells are plotted against their phase disparities. We find no correlation between position and phase disparities (correlation coefficient \( r = 0.12 \), \( R^2 = 1.45\% \)), suggesting that they are largely independent of each other.

Previously, we have reported that RF profiles for the left and right eyes are relatively matched for cells tuned to horizontal orientations whereas those for cells tuned to vertical orienta-
tions are predominantly dissimilar (11–13). We have confirmed this finding for the data reported here. This implies that cells tuned to horizontal orientations encode a small range of binocular disparity compared with cells tuned to vertical orientations. This orientation anisotropy is expected because binocular parallax yields a larger range of binocular disparities along horizontal compared with vertical directions due to the fact that the eyes are displaced laterally.

In Fig. 7a, we plot magnitudes of position disparity-X (●) and phase disparity (○) as a function of RF orientation. RF orientations of 0° and 90° correspond to horizontal and vertical, respectively. There is a tendency for cells tuned to horizontal orientations to have small phase disparities compared with those tuned to vertical orientations. The distribution of phase disparity for cells tuned to orientations within ±20° from horizontal has a smaller variance compared with the distribution for cells tuned to orientations within ±20° from vertical (F-test, $P < 0.01$). Barlow et al. (4) reported a similar orientation anisotropy (but see refs. 25 and 26). No orientation anisotropy is apparent for position disparity, which is consistent with most previous reports (7, 28, 29).

Fig. 7b shows how position disparity-X and phase disparity depend on RF spatial frequency. By definition, phase disparity is limited to a 180° phase angle as indicated (—). As a reference, disparities equivalent to a 90° phase angle are also shown (—). Phase disparities are scattered below the solid line, suggesting that they can be used to encode a wide range of binocular disparities. A regression analysis indicates that there is a tendency for phase disparity to decrease with spatial frequency (slope = −0.86, $P < 0.01$). In contrast, position disparities are in general very small (note that the spread of the position disparities along the vertical axis in the figure would be even smaller for true position disparities by a factor of $\sqrt{2}$) and are relatively constant across spatial frequency (regression slope = −0.75, $P = 0.089$). Therefore, if the visual system were to encode binocular disparity through position disparity, its performance in binocular fusion and stereo tasks would not depend on stimulus spatial frequency. On the other hand, if phase disparity were to be used, dependence on spatial frequency would be expected.

It has been reported that performance of human observers in binocular fusion and stereo tasks depends on stimulus spatial frequency (30–38). For example, Schor et al. (36) found that the fusion limit of human observers decreases with stimulus spatial frequency (size-disparity correlation) in a manner similar to the prediction of a phase encoding model, up to a spatial frequency of about 2.5 c/deg. Beyond this spatial frequency, however, the performance of human observers deviates from the prediction and becomes constant. Our results are concordant with theirs in the sense that phase disparity seems to provide the upper limit of binocular disparity at low spatial frequencies and position disparity provides a constant limit at high spatial frequencies for which phase disparities become necessarily small.

**DISCUSSION**

Using a quantitative RF mapping technique, combined with a reference-cell method, we have estimated RF position and phase disparities for simple cells in the cat’s striate cortex. Position disparities are generally small and are only suitable for encoding small binocular disparities. They do not show any correlation with RF orientation or spatial frequency. It seems, therefore, that RF position disparity may be a by-product of random jitter in RF position. On the other hand, phase disparities cover a wide range of binocular disparities and exhibit orientation anisotropy. They also provide a basis for the size-disparity correlation observed in psychophysics. Considered together, these results strongly favor the notion that binocular disparity is mainly encoded through RF phase disparity. However, RF position disparity may still play an important role in encoding binocular disparity at high spatial frequencies for which RF phase disparity becomes necessarily small.

As described earlier, psychophysics data such as those of Schor et al. (36) indicate that the performance of human observers in binocular fusion and stereo tasks consists of two parts: a spatial frequency dependent portion (at low spatial frequencies) and an independent portion (at high spatial frequencies). Interestingly, this dual behavior is apparently not unique to binocular fusion and stereopsis, but is also found in various spatial tasks (39–48). For example, DeValois and DeValois (39) measured the threshold of human observers for...
displacement of sinusoidal gratings at various spatial frequencies. They found that, at spatial frequencies below 2 c/deg, the threshold decreased with spatial frequency (see also refs. 41 and 42), but for higher spatial frequencies, the threshold was approximately constant (see also ref. 40). Similar results have also been reported for measurements of maximum displacement ($D_{max}$) for correct identification of direction in short-range apparent motion (43–48).

It is tempting to speculate that the dual behavior observed for binocular fusion and stereopsis, monocular displacement detection, and short-range apparent motion all share the same neural basis: a phase encoding mechanism for low spatial frequencies and a position encoding mechanism for high spatial frequencies. In order for the phase encoding mechanism to work properly, RF centers have to be at the same position. However, RF position is subject to slight random jitter. For RFs with low spatial frequency selectivity, this is not a problem since the amount of jitter is very small compared with the size of the RFs. For RFs with high spatial frequency selectivity, however, the amount of jitter may be significant compared with the size of the RFs, and position encoding becomes more reliable than phase encoding. To explain the dual behavior of their displacement threshold data, DeVilbiss and DeVilbiss (33, 39) proposed a two-stage model in which a phase processing stage (presumably at the striate cortex level) is followed by a position processing stage (extrastriate cortex). Our results suggest that both mechanisms may reside at the level of the striate cortex.

We are grateful to Dr. Erich Sutter for his advice on binary m-sequences and their application to receptive field mapping. We also thank Drs. Karen DeVilbiss and Clifton Schor for helpful discussions, and Greg DeAngelis for valuable comments and suggestions. This work was supported by research and CORE grants from the National Eye Institute (EY01175 and EY03176) and by a grant from the Human Frontier Science Program.