

# Neuronal Computation of Disparity in V1 Limits Temporal Resolution for Detecting Disparity Modulation

Hendrikje Nienborg,<sup>1</sup> Holly Bridge,<sup>2</sup> Andrew J. Parker,<sup>2</sup> and Bruce G. Cumming<sup>1</sup>

<sup>1</sup>Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Bethesda, Maryland 20892, and <sup>2</sup>University Laboratory of Physiology, University of Oxford, Oxford, OX1 3PT, United Kingdom

The human ability to detect modulation of binocular disparity over time is poor compared with detection of luminance modulation. We examined the physiological origin of this limitation by analyzing neuronal responses to temporal modulation of binocular disparity in striate cortex of awake monkeys.

When neurons were presented with random-dot stereograms in which disparity varied sinusoidally over time, their responses modulated at the stimulus temporal frequency, with little change in mean firing rate. We calculated modulation amplitude as a function of temporal frequency and compared this with the psychophysical performance of four human observers. Neuronal and psychophysical functions showed similar peak frequencies (2 Hz) and comparable high-cut frequencies (10 and 5.5 Hz, respectively). Thus, V1 (primary visual cortex) neurons appear to limit psychophysical performance.

The temporal resolution of the same neurons for contrast modulation was  $\sim 2.5$  times greater, which parallels the superior psychophysical performance for contrast. There is a simple mathematical explanation for this difference: it results from calculating cross-correlation between temporally broadband monocular images that are bandpass filtered before measuring correlation.

The limit on temporal resolution is a direct consequence of the binocular energy model that adds to the list of properties of human stereoscopic performance that are explained by this simple model of disparity encoding in V1: the same neurons can account for the performance of psychophysical tasks that result in either high (contrast) or low (disparity) temporal resolution. Because this principle holds whenever a broadband input is bandpass filtered before computing correlation, it may limit the resolution of other neuronal systems.

**Key words:** striate cortex; temporal processing; stereopsis; macaque monkey; correlation; energy model; human psychophysics

## Introduction

Binocular disparity can be used by the visual system to construct depth percepts (stereopsis). The striate cortex (area V1, primary visual cortex) is the earliest brain area in which selectivity for binocular disparity occurs in visual neurons (Barlow et al., 1967; Pettigrew et al., 1968). Neurons in V1 do not directly account for many perceptual features of stereopsis (Cumming and Parker, 1997, 1999, 2000). Nonetheless, they appear to play a critical limiting role for some aspects of behavioral performance, such as the threshold for detection of binocular disparity (Prince et al., 2000) and the acuity for perceiving spatial modulations of disparity (Nienborg et al., 2004). Here we examine the relationship between the temporal properties of disparity processing in single V1 neurons of awake macaque monkeys and in human observers.

A central objective was to understand whether V1 neurons

provide an account for the discrepancy between the psychophysical temporal resolution for disparity and that for contrast. Psychophysical studies have shown that the temporal resolution for disparity modulation (Norcia and Tyler, 1984) is poorer than for contrast modulation (Kelly, 1971b; Kelly et al., 1976). The temporal resolution of single V1 neurons for contrast modulation is high enough to support psychophysical performance (Hawken et al., 1996; Williams et al., 2004). If the temporal resolution of single V1 neurons reflects the biophysics of the cell under study, one might assume that the resolution should be similar regardless of the stimulus that is used to modulate the synaptic input. The temporal resolution of single neurons should then be similar for modulation of both disparity and contrast. It follows that such neurons would signal disparity modulations over a substantial range of frequencies that are psychophysically undetectable. The alternative is that cortical neurons may have different temporal frequency tuning functions for different types of stimuli, something for which there is little evidence at present (Ringach et al., 1997; Bredfeldt and Ringach, 2002). Our data argue that neither position is correct. Rather, we find that, although the temporal resolution for two stimuli can be quite different, a single mechanism explains both observations.

We examined temporal responses to disparity and contrast. To compare these data with behavioral performance, we mea-

Received June 8, 2005; revised Sept. 20, 2005; accepted Sept. 20, 2005.

This research was supported by the Intramural Research Program of the National Institutes of Health (NIH), National Eye Institute. Work at Oxford University was supported by a Wellcome Trust Programme Grant and a Royal Society Leverhulme Senior Research Fellowship (A.J.P.). We also thank Chris Hillman and Mark Szarowicz (NIH) and Stephen Laird (Oxford University) for excellent animal care and Jenny Read for help with the math.

Correspondence should be addressed to Hendrikje Nienborg, 49/2A50 Convent Drive, Bethesda, MD 20892-4435. E-mail: hn@sr.nei.nih.gov.

DOI:10.1523/JNEUROSCI.2342-05.2005

Copyright © 2005 Society for Neuroscience 0270-6474/05/2510207-13\$15.00/0

sured human psychophysical thresholds to disparity modulation as a function of temporal frequency using a procedure modified from Norcia and Tyler (1984). This then allowed us to construct a sensitivity curve for comparison with the behavior of V1 neurons.

Our data show that the neuronal temporal frequency tuning for disparity modulation differs radically from that to luminance modulation. This difference may offer a sufficient explanation of the human psychophysical performance. We find a simple mathematical explanation for this difference in temporal resolution to disparity and contrast modulation, which is consistent with current models of disparity selectivity in V1.

## Materials and Methods

**Animals and training.** We studied striate cortical neurons of one female (Rb) and three male (Df, Hg, and Rf) awake monkeys (*Macaca mulatta*) performing a standard fixation task for fluid reward. The animals were implanted with a head fixation post and scleral search coils in both eyes (Judge et al., 1980) under general anesthesia. The surgical and training procedures are described in detail by Cumming and Parker (1999).

For monkeys Hg and Rb, all procedures were performed in accordance with the United Kingdom Home Office regulations on animal experimentation. For monkeys Df and Rf, all procedures complied with the Public Health Service policy on the humane care and use of laboratory animals, and all protocols were approved by the Institute Animal Care and Use Committee.

**Single-unit recording.** Extracellular recording from V1 neurons was done by introducing tungsten in glass microelectrodes [monkeys Hg and Rb (Merrill and Ainsworth, 1972)] or glass-coated platinum–iridium electrodes (monkeys Df and Rf; Frederick Haer Company, Bowdoinham, ME) transdurally into the striate cortex on each day of recording. The signal was amplified (Bak Electronics, Mount Airy, MD), filtered (200 Hz to 2 kHz), digitized (32 kHz), and stored to disk. Unit isolation was always rechecked off-line. The horizontal and vertical positions of both eyes were measured with a magnetic scleral search system (CNC Engineering, Seattle, WA) and digitized at 800 or 587 Hz. Only data from successfully completed trials (fixation to within  $0.4\text{--}1^\circ$ , depending on monkey) were analyzed.

We mapped the minimum response field of each neuron by hand with a high contrast bar and subsequently centered stimuli over it.

**Stimulus presentation for electrophysiology.** The stimuli were displayed on two EIZO (Ishikawa, Japan) Flexscan 78 monitors (for monkeys Hg and Rb) and EIZO Flexscan F980 monitors (for monkeys Df and Rf) and were viewed stereoscopically at a distance of 89 cm through two small mirrors, positioned  $\sim 1.5$  cm in front of the animals' eyes. A Silicon Graphics (Mountain View, CA) workstation generated the stimuli at a mean luminance of 37 or 42  $\text{cd/m}^2$ , maximum contrast of 99%, and at a frame rate of 72 Hz.

The stimuli were generally presented for a duration of 2 s. For monkeys Df and Rf, disparity tuning was initially tested with stimuli lasting only 420 ms, separated by 100 ms, allowing four stimulus presentations within one 2 s trial. For all measures of temporal frequency response, 2 s stimulus durations were used.

**Measurements of disparity tuning functions and responses to disparity modulation.** We measured selectivity for disparity using circular patches of dynamic random-dot stereograms (RDSs). To keep mean luminance constant, the RDS consisted of equal numbers of randomly distributed black and white dots of 99% contrast ( $0.1^\circ \times 0.1^\circ$  size and with an overall density of 50%) on a midgray background. A new RDS was presented in each video frame. The RDS was centered on the minimum response field and extended beyond its limits. The disparity of the central region of the RDS varied from trial to trial, while a surrounding annulus ( $0.5\text{--}2^\circ$  wide) was kept at zero disparity. The width of the annulus was always greater than the largest disparity used, eliminating monocularly detectable changes in the stimuli and keeping variation of vergence to a minimum. If the neuron exhibited sensitivity to disparity, responses to sinusoidal modulations of disparity were studied. The neurons were presented with an RDS whose disparity varied as a sinusoidal function of space and time.

The RDS contained a sinusoidal variation in disparity as a function of vertical position (described in detail by Nienborg et al., 2004). For this study of temporal responses, the spatial frequency was low enough (usually 0.125 cycles per degree) that the disparity at any moment was essentially uniform within a receptive field. We chose to use a low-frequency sinusoidal spatial variation in disparity (rather than a planar patch) to maintain compatibility with our previous study on spatial summation of disparity (Nienborg et al., 2004). As a result, the two studies combined provide spatial and temporal slices through the same spatiotemporal stimulus space. The disparity modulated as a sinusoidal function of time between a value close to the preferred disparity and a value close to the null disparity (so that firing rate was a monotonic function of disparity within the range used). The total stimulus size was  $5^\circ \times 5^\circ$  ( $4^\circ$  modulating region plus  $1^\circ$  surrounding frame of zero disparity; 50% black and 50% white dots;  $0.1^\circ \times 0.1^\circ$  size; usually 99% contrast). Although a hard edge was visible between the  $1^\circ$  surrounding frame and the central stimulus, this lay far outside the receptive field of the neurons. We varied the temporal frequency of the corrugation in one octave increments (1–32 Hz for monkeys Hg and Rb; 1.12–36 Hz for monkeys Rf and Df), presented in pseudorandom order. Each stimulus condition was presented a minimum of four times (maximum of 27; mean of 10.3).

We interleaved an otherwise identical planar stimulus with a constant disparity equal to the mean disparity of the disparity modulation. Because there was no temporal or spatial modulation of disparity in this stimulus, it allowed us to estimate the extent of temporal modulation in the neuronal firing that was unrelated to the disparity modulation of the stimulus.

**Temporal frequency tuning in response to disparity modulation.** For entry into this study, we required that all neurons show significant modulation of firing rate with changes in disparity (in the RDS with uniform disparity) on a one-way ANOVA ( $p < 0.05$ ), and that their mean response at the preferred disparity exceeded 10 spikes/s. These criteria were met by 155 of 242 neurons (63%), not greatly different from the proportions reported previously for cortical area V1 (Prince et al., 2002).

Because the disparity modulation was chosen to cover a monotonic response range of the neuron, the main manifestation of disparity selectivity in response to the disparity modulations is a periodic modulation in firing rate at the drift frequency (Nienborg et al., 2004). We never observed substantial changes in mean firing rate (as a function of temporal frequency) without changes in this periodic modulation. We quantified this modulation using the relative modulation (RM) (Movshon et al., 1978a; Nienborg et al., 2004). It is the ratio of the amplitude of the modulation at the fundamental frequency ( $f_1$ , the temporal frequency of the disparity modulation) and mean response rate ( $f_0$ ). The value for  $f_1$  was obtained by averaging the responses for all trials and then calculating the modulation of this mean response at the frequency  $f_1$ . The same calculation (repeated for all temporal frequencies of the disparity modulation) applied to responses to the planar stimulus served as a control for the extent of modulation that was not attributable to the disparity modulation. To eliminate artifactual modulation related to response latency and the onset transient, the first 500 ms of each trial was discarded.

We used a resampling (1000 cycles) and bootstrapping method (Davison and Hinkley, 1997) to determine confidence intervals. Temporal frequency tuning curves for disparity modulation were only analyzed in cells whose relative modulation in response to at least one temporal frequency was significantly higher (on the 5% level, corrected for multiple comparisons) than in response to the planar stimulus.

**Temporal frequency tuning for contrast.** For comparison with the responses to disparity modulation, temporal frequency tuning was also measured in response to high contrast (99%), sinusoidal luminance gratings, at the preferred orientation and spatial frequency; both drifting and counterphase-modulating gratings were used (although both datasets were not collected for every cell). The stimuli were presented over the same range of temporal frequencies that was used for disparity modulation, a minimum of four times each.

Comparing temporal tuning to disparity in RDSs with that for contrast in gratings assumes that temporal summation is a fixed property of the neurons involved, not something that varies with the spatial structure of a stimulus. In fact, simple changes in a spatial stimulus may affect tem-

poral tuning, e.g., changes in contrast (Holub and Morton-Gibson, 1981; Albrecht, 1995; Carandini et al., 1997). These may in part reflect increases in membrane conductance caused by the inhibitory inputs that mediate contrast normalization. If the extent of contrast normalization elicited by gratings and RDSs were different, this alone could lead to differences in temporal response. We therefore also measured the temporal frequency tuning for contrast in an RDS stimulus, using dot patterns in which the dot locations remained fixed throughout an entire 2 s trial. However, the luminance of the dots was modulated sinusoidally (as dark dots got lighter, bright dots got dimmer, until the contrast was reversed). This is equivalent to applying counterphase modulation to all of the spatial Fourier components of the dot pattern, so we describe this stimulus as a counterphase-modulating RDS. The RDSs were otherwise similar to the stimuli used for measuring disparity tuning. The RDSs and gratings were presented monocularly to the dominant eye, or binocularly at the preferred disparity. For complex cells, the dominant response to contrast reversal is modulation at the second harmonic of the stimulus (Movshon et al., 1978b; Hawken et al., 1996). Tuning curves plotted therefore  $f_2$  (modulation amplitude at the second harmonic of stimulus temporal frequency).

**Gaussian fits.** We used Gaussian functions of linear or logarithmic temporal frequency [whichever minimized the least square residuals in a nonlinear optimization algorithm (Lagarias et al., 1998)] to quantify temporal frequency tuning in response to disparity modulation, luminance gratings, and counterphase-modulating RDS and to characterize the psychophysical performance. The Gaussian functions had four free parameters (mean, amplitude, SD, and baseline), which, except for the mean, were constrained to values  $\geq 0$ . Only fits that explained  $>65\%$  of the variance in the data were used for additional analysis. Across all fits (to both disparity and contrast data), the mean variance explained by all fits was 90%.

**Onset transient.** Characterization of the response at onset (time to 60% peak response) was obtained from spike density functions (SDFs). We first identified the peak value in the first 250 ms of the SDF and then determined the first bin in which the value exceeded 60% of this maximum.

**Temporal integration time.** Any delay in the response will introduce a phase difference between the sinusoidal input and the sinusoidal response. As frequency increases, a fixed delay corresponds to a proportionally larger fraction of the period, so the phase lag increases with frequency (linearly for a fixed delay). In addition to any absolute latency, the temporal kernel itself introduces an effective delay. To determine the sum of these components, the phase of the response modulation was plotted as a function of stimulus frequency. The slope of the resulting function is used to estimate the sum of these delays, and, following previous usage (Reid et al., 1992; Hawken et al., 1996), we refer to this as the temporal integration time.

To reduce error caused by noise, phases were only calculated at the temporal frequencies for which RM was  $>0.5$  peak RM. The slope of the function relating phase and input temporal frequency was then determined by linear regression.

**Psychophysical experiments.** We measured sensitivity to disparity modulation in four human subjects (one female, two naive observers), with normal or corrected-to-normal vision. The subjects viewed the stimuli in a Wheatstone stereoscope similar to that used for recordings. They were presented, in a two-interval forced-choice procedure, with a signal stimulus containing temporal modulation of disparity and a no-signal stimulus without disparity modulation. Stimuli were planar dynamic random-dot stereograms (size,  $8^\circ \times 8^\circ$ ; dot size,  $0.1^\circ \times 0.1^\circ$ ; 50% black and 50% white dots; dot density, 50%). Signal stimuli were disparity-modulating RDSs (temporal frequency of 1.5–24 Hz, varying between blocks), whose amplitude of disparity modulation was varied pseudorandomly between trials to determine detection thresholds (using the method of constant stimuli). Above the temporal frequency limit of disparity modulation, a disparity-modulating RDS is perceived as a transparent cuboid (Norcia and Tyler, 1984). The no-signal interval therefore consisted of an RDS giving the same percept. On each frame of the no-signal RDS, multiple disparities were present simultaneously and assigned randomly to the dots in the display. The distribution of these

disparity values was chosen to match exactly the distribution in the signal interval for that individual trial. (On any given frame in the signal interval, all of the dots had the same disparity, and this disparity changed from frame to frame as a sinusoidal function of time). Thus, for each stimulus pair, the total number of dots at each disparity was the same in the signal and in the no-signal interval. The no-signal interval contained the same random distribution of disparities on every frame, whereas the signal interval contained only one disparity on each frame.

We chose to use a planar stimulus for the human psychophysics (rather than a low spatial frequency sinusoidal modulation as in the physiological experiments) for two reasons. First, although for the physiological experiments the disparity was nearly constant over any one receptive field, human observers could see the whole image, so the approximation to constant disparity was much poorer. Second, this spatial variation would have produced a spatial cue that could have been used by the subjects, especially at high temporal frequencies: when the number of video frames in a stimulus cycle is small, then the time-averaged disparity distribution would be different at different stimulus locations. This problem was avoided by using stimuli that were planar at any instant.

To avoid monocularly detectable changes in the stimuli, the disparity-modulating RDSs were surrounded by a  $1^\circ$  frame at zero disparity in both the signal and the no-signal interval. Total stimulus size was  $9^\circ$  ( $8^\circ$  plus  $1^\circ$  frame). The frame therefore generated a hard disparity edge at the transition to the part that contained the disparity modulation (signal interval) or multiple disparities (no-signal interval). At the higher frequencies, this edge looked identical in the no-signal and the signal intervals (because both stimuli gave the thickened percept). However, at the lowest frequencies used ( $<3$  Hz), the presence of this edge (a nonzero relative disparity) at any point in the trial might have been used to detect the interval with modulation. At these low frequencies, we therefore added a small pedestal disparity (exceeding the maximal disparity modulation amplitude, constant across trials) to both signal and no-signal intervals. Consequently, the stimulus remained on the same side of the fixation point throughout all low-frequency trials, and the edges were always visible. Thus, subjects were forced to detect modulation in disparity over time to do the task at all frequencies. Finally, to ensure that temporal components of the onset did not distort the results, all disparities in the first 100 ms were tapered with a temporal Gaussian window, for both the no-signal and signal intervals. (The Gaussian taper at the onset seemed to be sufficient, although in theory the same concern applied at stimulus offset. In a control experiment for one temporal frequency in two subjects, performance was statistically indistinguishable whether the Gaussian taper was used only for onset or for both on and offset.) The monitor frame rate was set to 96 Hz for the psychophysical experiments to improve the frequency resolution. The amplitude of the disparity modulation was varied pseudorandomly between trials to determine detection thresholds, using the method of constant stimuli. The range of these stimuli was chosen after collecting pilot data to ensure that subjects were performing  $>90\%$  correct for the largest amplitude.

Subjects were asked to fixate on a cross in the center of the monitors. Each trial was initiated by a button press. The stimuli were presented foveally and usually lasted for 1 s each (separated by a 100 ms interval) unless subjects terminated trials earlier by making their choice before the end of the stimulus presentation. Subjects received visual feedback after each trial. Cumulative Gaussian curves were fit (using a maximum likelihood estimator) to the psychometric functions to determine the point at which 75% correct performance was achieved. (We will refer to this as the disparity threshold at the respective temporal frequency of disparity modulation.)

## Results

We recorded from 242 isolated units in striate cortex of one female and three male monkeys: 110 from monkey Hg, 62 from monkey Df, 59 from monkey Rf, and 11 from monkey Rb. Of these, 155 units were significantly selective for disparity (5% level in ANOVA) and fired  $>10$  spikes/s in response to their preferred disparity. Responses to disparity-modulating RDSs were analyzed for 73 of the 155 cells. (For the remainder, the unit was

either lost before a complete dataset was recorded or the range of disparity modulation did not match the range of disparity selectivity of the cell.)

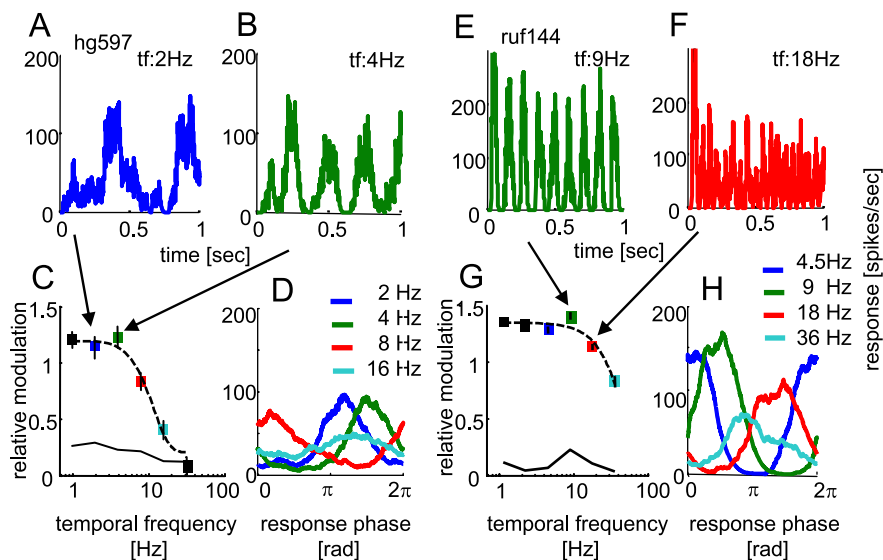
### Temporal frequency tuning in response to disparity modulation

During presentation of these stimuli, the disparity at any point in the receptive field modulates as a sinusoidal function of time. Figure 1, *A, B* and *E, F*, depicts the SDFs of two disparity-tuned neurons in response to disparity-modulating RDSs at different temporal frequencies. *D* and *H* show cycle averages of these SDFs, allowing clearer comparison of the response amplitude and phase. These neurons, like the vast majority of disparity-selective neurons, modulated their firing at the stimulus frequency. (This is a direct consequence of the fact that we chose the range of disparity modulation to match the monotonic region of the disparity tuning curve.) We quantified the extent of modulation by calculating the proportion of the firing rate of the neuron that varied synchronously with the sinusoidal changes of disparity (RM; see Materials and Methods).

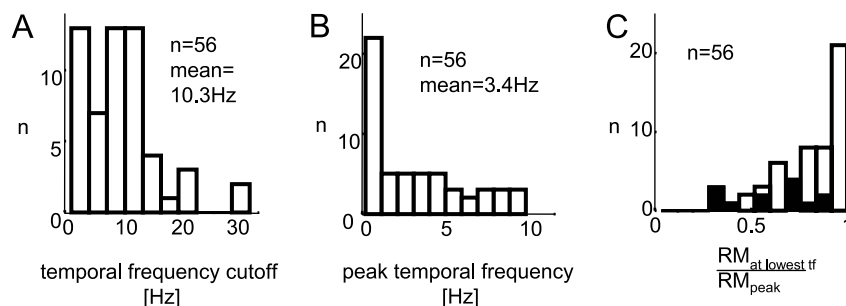
To control for fluctuations in the absence of any changes in disparity, we also measured the response to an RDS whose depth modulation was 0: this control was a planar stimulus with a constant disparity over space and time (although the dots were still dynamic), interleaved with the modulating stimuli. In 69 of 73 neurons, the modulation was significantly stronger for at least one of the temporal frequencies tested than in response to the planar control stimulus (by resampling,  $p < 0.05$ , corrected for multiple comparisons).

The mean firing rate was only moderately affected by temporal frequency. Only 32 of 69 cells showed any significant change (one-way ANOVA,  $p < 0.05$ ), and, even for these cells, the modulation in mean firing was modest and rarely showed systematic changes as a function of temporal frequency. We therefore used the RM values to obtain temporal frequency tuning curves in response to disparity modulation (Fig. 1*C, G*) and described these by fitting Gaussian curves. For 56 of 69 cells, the Gaussian fits explained  $>65\%$  of the variance and were used to describe the temporal properties [high-cut temporal frequency (frequency at two-thirds peak RM), peak frequency, and frequency bandwidth obtained from the mean and the SD of the Gaussian fit, respectively]. The majority of fits were better than this criterion value: in 51 of 56 neurons, the fits explained  $>75\%$  of the variance.

Unless stated otherwise, the remainder of our quantitative analysis is restricted to these 56 cells. The mean high temporal frequency cutoff was  $10.3 \pm 6.6$  Hz SD (Fig. 2*A*); the median was 10.2 Hz. The mean peak frequency was  $3.4 \pm 3.0$  Hz SD (Fig. 2*B*), with a mean frequency bandwidth of  $6.4 \pm 6.0$  Hz SD (data not



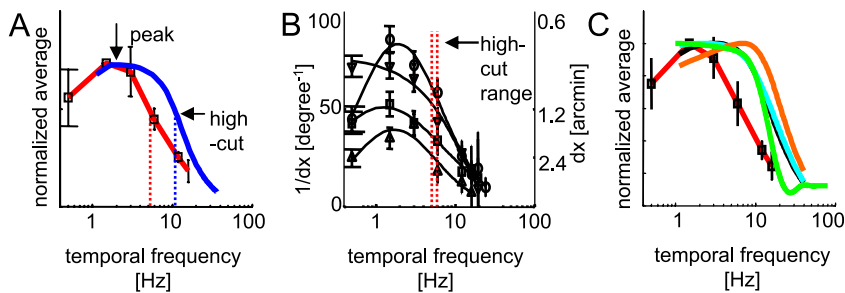
**Figure 1.** Responses to disparity-modulating RDSs for two neurons (hg597, ruf144). *A, B* and *E, F* depict SDFs in response to disparity-modulating RDSs at different temporal frequencies (2 and 4 Hz for *A* and *B*; 9 and 18 Hz for *E* and *F*). *D* and *H* show cycle averages of the SDFs (2–16 and 4.5–36 Hz, respectively). The RM is plotted as a function of temporal frequency in *C* (cell hg597) and *G* (cell ruf144). Baseline RM (solid black line) is the modulation to a control stimulus (no disparity modulation) measured for the same temporal frequencies. The arrows point to the RM calculated for the data depicted in the respective panels *A, B* and *E, F*. RM plotted as a function of temporal frequency is fitted by Gaussian curves (dashed lines in *C* and *G*). High cutoff values were usually similar to that shown in *C* (8.1 Hz) but were significantly higher in a few neurons, as seen in *G* (35.3 Hz).



**Figure 2.** Temporal frequency tuning of the cell population. *A* summarizes the high-frequency cutoff for 56 neurons. Mean value is 10.3 Hz. A frequency histogram of the peak temporal frequencies of the 56 neurons is shown in *B* (mean of 3.4 Hz). Most neurons had low peak temporal frequencies. The ratio of RM at the lowest temporal frequency over peak RM was used to estimate the extent of low-frequency attenuation (frequency histogram in *C*;  $n = 56$ ). Filled bars correspond to cells with statistically significant low-frequency attenuation ( $n = 13$ ;  $p < 0.05$  by resampling).

shown). The distribution of the peak temporal frequencies shows that many cells have peaks at 1 Hz (the lowest frequency tested), suggesting that they are low pass. To quantify the extent to which cells were low pass/bandpass, we compared the RM at the lowest temporal frequency against peak RM. This comparison is shown in Figure 2*C* as a ratio, which was calculated directly from the raw RM data, not from the fitted curve. The distribution of this ratio is strongly biased toward values of 1. Nonetheless, 13 of 56 cells (filled bars) showed statistically significant low-frequency attenuation ( $p < 0.05$ , two-tailed test by resampling, corrected for multiple comparisons). The magnitude of the attenuation was generally modest, with only 4 of 56 neurons showing responses that fell below half their peak value.

It is possible that these figures might overestimate the true magnitude of low-frequency attenuation, if the disparity modulation were to have induced vergence eye movements. Monkeys trained to make tracking vergence eye movements are able to do so only up to frequencies of 1–4 Hz (Cumming and Judge, 1986),



**Figure 3.** Comparing the neuronal population response with the psychophysical performance. **A** depicts the averaged Gaussian fits to the RM as a function of temporal frequency for the 56 neurons (blue solid line). The arrows indicate the high cutoff (10.5 Hz) and peak (2 Hz) for the averaged fits. Superimposed is the mean of the psychophysical performance (normalized for each subject by the value at 1.5 Hz) in response to disparity modulation for four human subjects (open squares, red solid line). Dotted lines indicate temporal frequency high-cut values. Psychophysical performance for each subject ( $n = 4$ ) is shown in **B**. Sensitivity ( $1/\text{disparity threshold}$ ; see Materials and Methods) is plotted as a function of temporal frequency. Mean high cutoff (two-thirds peak sensitivity) is  $5.5 \pm 0.4$  Hz SD, and the range of the temporal frequency cutoffs between subjects is indicated by the dotted lines. In **C**, the influence of different pooling schemes is examined. The red curve shows the mean psychophysical performance and is identical to that in **A**. The orange curve depicts the average of the Gaussian fits for the 25% of the neurons ( $n = 14$ ) with the highest temporal frequency cutoffs, the cyan of those 25% of the neurons ( $n = 14$ ) with the highest RM values. For the black curve, the contribution of each fit ( $n = 56$ ) is weighted according to the statistical reliability of RM (see Results) before being averaged. Cubic splines were also fit to all 56 neurons, and these fits were averaged (green curve).

depending on stimulus amplitude. If such movements were elicited by the stimuli used here, they would have reduced modulation in retinal disparity over the receptive field of the neuron. Because this reduction would only occur at low stimulus frequencies, it could lead to the appearance of low-frequency attenuation. However, examination of vergence records revealed no systematic responses to the disparity modulation in the stimulus. The measured amplitude of vergence modulation was on average only 15% of the modulation in stimulus disparity. Three observations suggest that this vergence variation primarily reflects artifacts [as also suggested by Read and Cumming (2003)]. First, the amplitude of the response at any one frequency was independent of the stimulus modulation frequency. Second, the extent of modulation observed in vertical vergence was similar to that seen for horizontal vergence, despite the fact that there was no modulation in the vertical disparity of the stimulus. Finally, we found no correlation between the relative amplitude of the vergence eye movements and the extent of low-frequency attenuation. Thus, it appears that the modest reduction in neuronal response amplitude at low stimulus frequencies is a real property of the response of some neurons to disparity modulation, not an artifact related to vergence eye movements.

To summarize the responses of the whole neuronal population, we averaged the Gaussian fits ( $n = 56$ ) describing the variation of RM as a function of temporal frequency (Fig. 3A). (In Figure 3A, the fits were not normalized before averaging. We also normalized each fit by its peak before calculating the average, and the resulting curve was very similar to that shown.) The temporal frequency cutoff for this averaged fit was 10.5 Hz (Fig. 3A, dotted line), similar to the mean of the temporal frequency cutoffs, and the peak was 2 Hz (Fig. 3A, arrow). The low-frequency roll off is modest. If the temporal integration of disparity signals in striate cortex is a limiting factor for psychophysical performance, these features might be reflected in the psychophysical sensitivity to disparity modulation.

### Comparison of neuronal and psychophysical sensitivity to disparity modulation

A previous study showed that the highest temporal frequency at which humans were able to detect motion in depth in RDS was 9

Hz and to detect flickering in depth 14 Hz (Norcia and Tyler, 1984). However, it is difficult to relate these measures of limiting frequency with the continuous sensitivity curve we show for the neuronal population (Fig. 3A). We therefore used a protocol very similar to that of Norcia and Tyler (1984) but measured disparity thresholds at each frequency. The reciprocal of these thresholds then yields a sensitivity curve for human observers that can be compared with the neuronal population (Fig. 3A, B).

Comparing the neuronal and psychophysical sensitivity curves, there are two similarities. For all four subjects, performance was best between 1.5 and 3 Hz, which corresponds very closely to the peak of the neuronal population (2 Hz). Second, despite differences between subjects in the absolute sensitivity to disparity, the temporal frequency high cutoff (defined as two-thirds peak performance) was almost

identical in all subjects (5.1–5.9 Hz; mean,  $5.5 \pm 0.4$  Hz SD). Although this is lower than the equivalent measure for the neuronal population (10.5 Hz), the neuronal response to disparity modulation is closer to human psychophysical performance than it is to the responses of V1 neurons to contrast modulation (Hawken et al., 1996). As the responses get smaller than our (arbitrary) cutoff criterion, neuronal and psychophysical data get closer: for a high-cut criterion defined as 10% peak performance, the mean for the neurons was 31 Hz, and the mean for the psychophysical performance was 24 Hz.

There are several reasons why the pooled activity of a neuronal population may yield poorer performance than the average shown in Figure 3A. First, for the single neuron studies, the disparity range for the modulation was matched to the monotonic region of the disparity response for each individual neuron. Many of the neurons activated by a particular choice of range for the disparity modulation will not be optimally modulated by this choice. For example, consider a neuron whose preferred disparity lies in the center of the disparity range. This disparity appears twice in each stimulus cycle, so the neuron will modulate its response at twice the stimulus frequency. The neuronal response magnitude will therefore reach a criterion level of attenuation at half the frequency we observe with our stimuli. Second, if signals from neurons with slightly different delays are pooled, this results in cancellation of responses at high frequencies, which would lower the temporal frequency cutoff of the population average.

Of course, other pooling models (e.g., using only the neurons with highest temporal resolution) would produce a result with higher temporal resolution than we show in Figure 3A. To explore the effects of different ways of pooling these neuronal responses, we generated a population average response in several different ways. The results are compared in Figure 3C. One possibility is that only those neurons able to modulate at higher temporal frequencies would contribute to a decision on disparity modulation. We therefore averaged fits of the 14 (25%) neurons with the highest cutoff values (orange curve). Inevitably, this produces larger responses at high frequencies, with a cutoff value of 18.7 Hz and a peak temporal frequency of 6.9 Hz. There are several reasons why this pooling rule might not be used by psychophysical observers. First, it requires knowledge of which neu-

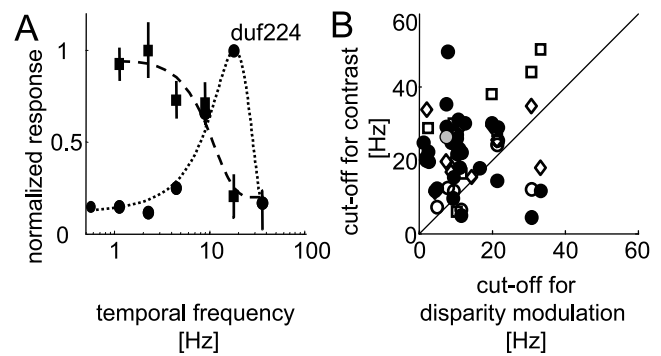
rons have the highest temporal response. Second, to deliver good performance across all frequencies, this rule requires using different pools of neurons for different stimuli.

A simpler approach would be to pool only those neurons whose responses modulate most strongly. Figure 3C (cyan curve) therefore also shows the average Gaussian fit of those 14 (25%) neurons with the highest values of RM. This response is very similar to the mean for the whole population. One problem with this approach is that it takes no account of the statistical reliability of any modulation. To do this, we weighted the contribution of each neuron according to the reliability of their modulation. We estimated this reliability from the ratio of the maximum value of RM for a cell over the square root of the pooled residual variance around the mean RM at each frequency. The black curve in Figure 3C plots the average of the weighted fits (high-cut temporal frequency for this average is 13.7 Hz). Finally, to avoid any systematic error attributable to the choice of Gaussian fits, we also fitted cubic splines to the RM data for each neuron and then took the mean of these spline fits (Fig. 3C, green curve, with a high-cut temporal frequency of 12.0 Hz). All of these analyses produce similar results, with a neuronal response function that extends to slightly higher frequencies than the psychophysical data.

A pooling model that relies on the neurons most responsive at higher frequencies would, of course, also predict higher resolution for monocular contrast flicker than simple averaging. Thus, to avoid our conclusions being limited to any particular pooling model, we examined whether the difference in temporal resolution observed psychophysically is also present in striate cortex. If the response modulation of disparity-selective neurons is the limiting factor in psychophysical detection of disparity modulation, then the discrepancy between disparity and contrast resolution should also be seen in V1. Human psychophysical sensitivity to contrast modulation is a bandpass function of frequency (Kelly, 1971a), similar to the responses of many V1 neurons (Hawken et al., 1996). More importantly, the highest temporal frequency that an observer can detect with contrast modulation is substantially higher than that for disparity modulation. Under optimal conditions (high luminance, large fields) human flicker sensitivity can reach 80 Hz (Kelly, 1961). For smaller stimulus sizes, similar to those used in recording experiments (2–4°) and otherwise optimal conditions, contrast sensitivity falls to two-thirds peak at ~30 Hz (de Lange, 1958; Kelly, 1971a). This suggests that a simple pooling rule, like that we used to generate Figure 3A, also reconciles the responses of single V1 neurons with the psychophysical resolution for contrast modulation. The reported temporal resolution of V1 neurons for contrast is significantly higher (by a factor of ~2.5) than the temporal frequency cutoff that we find here in response to disparity modulation.

We now consider three explanations for this discrepancy. First, there could be a difference between the population of disparity-selective cells and other V1 neurons. Alternatively, V1 neurons might modulate their outputs more sluggishly than the inputs from the LGN. The high temporal frequency response to luminance modulation would then be supported by a sustained, nonmodulated elevation in firing rates at the cortical level in response to a modulated input from LGN neurons. After rejecting these two possibilities, we demonstrate that the mechanism of computing binocular disparity might itself explain the difference and show that the mathematical nature of the computation predicts this difference.

With respect to the first possible explanation, only ~60% of V1 neurons are disparity selective (Poggio et al., 1977, 1985; Prince et al., 2002). If disparity-tuned neurons tended to prefer

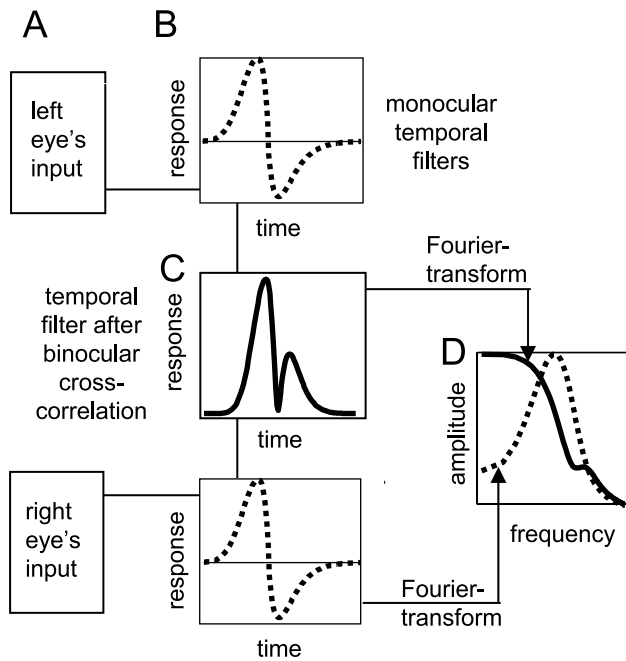


**Figure 4.** Temporal frequency cutoffs in response to contrast modulation are generally higher than in response to disparity modulation. **A** shows responses of one example neuron to disparity modulation (squares) and to drifting luminance gratings (circles). Axes show normalized responses [RM for disparity modulation, and firing rate (in spikes per second) for drifting gratings] as a function of temporal frequency (in Hertz). The temporal frequency high cuts are 7.5 and 26.4 Hz for disparity and contrast modulation, respectively. **B**, Filled symbols show the contrast temporal frequency cutoff values calculated from mean response rates; open symbols show cutoff values calculated from modulated responses. Note that, for counterphase-modulating stimuli, the ordinate refers to the frequency of the modulation in neuronal firing (i.e., twice the stimulus frequency because these are all complex cells). Temporal frequency cutoffs calculated from mean response rates to drifting luminance gratings were significantly higher than in response to disparity-modulating RDS (27 neurons; filled circles; neuron of **A** shown in gray): the geometric mean ratio is 2.3 ( $>1$ ;  $p < 0.001$ ). Ten cells showed modulation at the stimulus frequency in response to drifting gratings ( $f_1 > f_0$ , simple cells; open circles). For these cells, the cutoff frequency for the  $f_1$  component is plotted. For the remaining cells, the response modulation ( $f_2$  component) to counterphase-modulating stimuli, RDS (open diamonds;  $n = 11$ ) or gratings (open squares;  $n = 8$ ), was analyzed. For these 29 neurons, the geometric mean of the ratio in cutoff frequencies (all derived from modulated responses) is significantly higher than one (1.8;  $p < 0.001$ , by resampling).

lower temporal frequencies than the average V1 population, this discrepancy would simply be the consequence of our sampling only disparity-tuned neurons. To address this possibility, we directly compared the high-frequency cutoff in response to disparity modulation with that in response to drifting luminance gratings, in a subsample of 27 neurons.

In Figure 4A, the temporal frequency tuning curves in response to disparity modulation (squares, dashed line) and to drifting luminance gratings (circles, dotted line) are superimposed for one neuron (duf224). The high temporal frequency cutoff in response to disparity modulation (7.5 Hz) is substantially lower than that in response to drifting luminance gratings (26.6 Hz). Another prominent difference is that the neuronal tuning curves tend to be low pass for disparity modulation and bandpass for drifting luminance gratings. This shift to a low-pass response is easy to explain as a consequence of using dynamic RDSs (temporally broadband contrast stimulus) to measure temporal frequency tuning to disparity modulation: when the temporal modulation of disparity is zero and the disparity in the receptive field is constant, the dynamic RDS yields a sustained response (because its contrast energy is temporally broadband). Thus, the response does not drop to baseline at low frequencies for disparity modulation. Conversely, the response of many V1 neurons to drifting luminance gratings decreases at low frequencies, precisely because these stimuli are temporally narrowband for contrast.

In Figure 4B (filled circles, gray circle depicts the neuron shown in Fig. 4A), the temporal frequency high cut in response to drifting luminance gratings is plotted against the high cut in response to disparity modulation for 27 cells. (These 27 neurons are the subgroup of the 56 neurons in Figure 2 for which we also had



**Figure 5.** Schematic representation of the effect of bandpass filtering on the frequency response of the binocular cross-correlation. Input from the left and right eye (**A**) is passed through a temporal filter (**B**, dotted line). Appendix shows that the temporal kernel of the binocular cross-correlation (**C**, solid line) corresponds to the squared monocular kernel. **D** depicts the frequency response (bandpass) of the monocular kernel (dotted line) and of the binocular cross-correlation (solid line), which is low pass and has a lower cutoff frequency.

tuning data with drifting luminance gratings.) The vast majority of the cells respond at higher temporal frequencies to drifting luminance gratings than to disparity-modulated stimuli. Only three cells are reliably below the identity line (by resampling,  $p < 0.05$ ). The geometric mean of the ratio of grating cutoff over disparity cutoff is 2.3, which is significantly larger than 1.0 ( $p < 0.001$ , by resampling). The correlation is not significant ( $r = -0.28$ ). The mean temporal frequency high cut for the gratings is 21.7 Hz, similar to the values reported previously (Hawken et al., 1996). Thus, it appears that the temporal frequency tuning of individual neurons is different for the two stimuli. Again, to ensure that this conclusion was not limited to this particular high-cut criterion (temporal frequency yielding two-thirds of peak response), we repeated this comparison with high cuts defined at 10% of peak response. For this criterion, the geometric mean of the ratio of the temporal frequency cutoff to contrast modulation over the cutoff to disparity modulation was 2.0 ( $p < 0.001$ , by resampling).

However, the temporal frequency tuning curves for disparity modulation were based on RM, whereas the responses to drifting luminance gratings were obtained from mean firing rates. Suppose a cortical cell receives input from LGN neurons that can modulate their response up to higher frequencies than the cortical cell [as results by Hawken et al. (1996) suggest]. Thus, at higher temporal frequencies, the cortical neuron would still receive modulating input from the LGN, which may be sufficient to elevate the unmodulated firing rate (in the presence of a static nonlinearity relating input from the LGN to cortical output).

Thus, one explanation of our results might be that some input elements of the cells in Figure 4 have higher temporal frequency cutoffs than the output (as measured by the RM). Such an explanation should hold independently of which stimulus is used to drive response modulation. We therefore examined modulation

in firing rate elicited in contrast modulating stimuli. For simple cells ( $f_1:f_0 > 1$ ), we examined the  $f_1$  component in response to drifting gratings. To modulate the response rates of complex cells, we used counterphase-modulating stimuli (gratings or RDS; see Materials and Methods). The dominant response of complex cells to these stimuli is modulation at the second harmonic of the stimulus frequency (Movshon et al., 1978b; Hawken et al., 1996).

Figure 4B (open symbols) compares these measures of high-frequency cutoff for the response modulation of each neuron with those observed in response to disparity modulation. Note that the ordinate marks the frequency of the modulation of the cells: for the open squares and diamonds (counterphase-modulating stimuli), it therefore corresponds to twice the stimulus frequency. As for the filled circles, the cutoff frequency is systematically higher for contrast modulation, with only one cell significantly below the identity line. The geometric mean of the ratio of the temporal frequency cutoff in response to contrast modulation over that in response to disparity modulation is 1.8, significantly larger than 1 ( $p < 0.001$ , by resampling).

The comparison of temporal integration of disparity in RDS with temporal integration of contrast in gratings implicitly assumes that these functions are independent of the spatial structure of the monocular stimuli. However, it is possible that cells could show different temporal integration of contrast in response to RDS than to gratings. Changes in the extent of contrast normalization appear to change temporal responses (Holub and Morton-Gibson, 1981; Albrecht, 1995; Carandini et al., 1997). Therefore, if the RDS and grating stimuli elicited different magnitudes of this normalization, it could explain part of the difference we observed. To eliminate any concerns about spatial properties influencing temporal tuning, we therefore also restricted this analysis to neurons tested with counterphase-modulating RDS (see Materials and Methods), with similar results (the geometric mean of the ratio of the frequency cutoff in response to counterphase-modulating RDS over that in response to disparity modulation was 2.0;  $p < 0.01$ ;  $n = 11$ ).

This indicates that the neurons are able to modulate their response rates at higher frequencies in response to contrast modulation than in response to disparity modulation. This cannot be explained simply by suggesting that cortical neurons modulate their output more sluggishly than their inputs.

### The sluggish temporal response to disparity is attributable to the computation of binocular cross-correlation

We will now show that the low temporal resolution for disparity modulation can be explained by the interaction of two factors: first, the broadband signals from the left and right eyes pass through a monocular stage with bandpass temporal characteristics (Reid et al., 1997) for contrast stimuli; second, the monocular signals are brought together by computing a form of cross-correlation between left and right eye signals (Ohzawa et al., 1990; Fleet et al., 1996; Qian and Zhu, 1997; Anzai et al., 1999). This sequence of operations is by itself sufficient to ensure that the temporal response to disparity modulation must be slower than the response to contrast modulation.

To understand the behavior of such a model to our stimulus, we explored how the calculation of binocular correlation is affected by bandpass filtering of the broadband temporal sequence of monocular images received by each retina (Fig. 5). Note that the differences between the frequency response to correlation and that of the monocular filter are similar to those between the temporal frequency tuning to disparity modulation and contrast

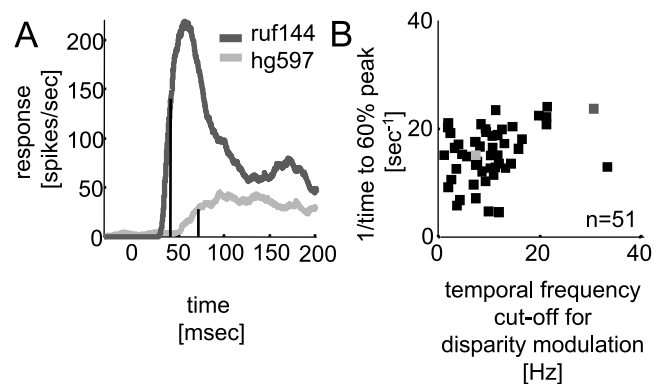
modulation (compare Figs. 4A, 5D). In Appendix, we show that, when broadband monocular inputs are temporally bandpass filtered, the frequency response to changes in binocular correlation is very different from the response of the monocular temporal kernel. The frequency response for variation in binocular correlation has a low-pass response [as observed for the majority of the neurons (Fig. 2C)]. For a similar monocular temporal filter in the two eyes, the frequency response of the correlation is determined by the Fourier transform of the square of that temporal kernel. Because of the squaring, the temporal frequency cutoff is therefore primarily determined by the envelope, i.e., the bandwidth, of the monocular temporal filters, regardless of their temporal structure within it.

Note that this is analogous to envelope detection in acoustics (Lawson and Uhlenbeck, 1950; Viemeister, 1979; Dau et al., 1999): for amplitude-modulated broadband noise, the envelope spectrum is a low-pass function whose cutoff frequency depends on the bandwidth of the frequency spectrum of the carrier.

Thus, the fact that the cutoff for the temporal frequency of disparity modulation is lower than that for contrast modulation results directly from the computation of binocular correlation by the neuron (Fig. 5). This follows if temporally broadband monocular input is bandpass filtered (see Appendix) before calculating cross-correlation. For a monocular low-pass filter, the squared filter would be narrower and would yield a higher temporal frequency cutoff, the opposite from our mean results. Consistent with this scheme, two of the three neurons in Figure 4 with cutoff values reliably higher to disparity modulation than to contrast modulation (measured in response to drifting luminance gratings) have low-pass responses to contrast modulation.

According to this explanation, it should be possible to predict the frequency response to disparity from the frequency response to luminance stimuli measured monocularly. However, several factors make this prediction difficult to examine in practice. First, the predicted cutoff frequency depends on both the peak frequency and the bandwidth of the luminance response. Indeed, it is quite sensitive to the exact form used for the linear temporal kernel (see Appendix). Our monocular measures of temporal frequency did not constrain estimates of the temporal kernel with sufficient reliability and detail. Second, the predicted relationship depends on the temporal kernel for both eyes, and we generally only measured temporal frequency tuning in the dominant eye (and for many disparity-selective neurons, monocular responses from the nondominant eye are too weak for such quantitative measures). Finally, the predicted relationship also changes if the output nonlinearity of the energy model (normally half-squaring) is altered. Given these difficulties, it is perhaps not surprising that we did not find a significant correlation between the high-frequency cutoff for disparity modulation predicted from our measures of contrast temporal frequency response and that observed ( $r = 0.25$ ;  $p = 0.3$ ;  $n = 19$ ; comparison restricted to those neurons for which we measured temporal frequency tuning to drifting gratings based on mean firing rates and for which the tuning curves could be fitted with Gaussian functions of linear temporal frequency).

Note that the explanation we propose does not require any of the known temporal nonlinearities in the early visual system (Holub and Morton-Gibson, 1981; Albrecht, 1995; Carandini et al., 1997). Even for pure linear temporal summation, the lower temporal frequency resolution is a necessary consequence in a system that computes cross-correlation of bandpass-filtered broadband monocular images. Temporal nonlinearities, such as



**Figure 6.** Temporal frequency cutoff to disparity modulation and rise time at response onset. **A** shows averaged SDFs of the onset in response to disparity-modulating RDS. Solid lines mark the rise time to 60% peak. The neuron (ruf144) with the shorter rise time (42 ms; dark gray) has the higher temporal frequency cutoff (35.3 Hz). Neuron hg597 (light gray) has a rise time of 72 ms and a temporal frequency cutoff of 8.1 Hz. (The cells are the same for which the SDFs in Figure 1 are shown.) The scatter plot in **B** compares the temporal frequency cutoffs in response to disparity modulation with the reciprocal of the time to 60% peak at the response onset ( $n = 51$ ; filled squares). The correlation is significant ( $r = 0.34$ ;  $p < 0.01$ ), as expected for the cross-correlation of bandpass-filtered images (see Results and Appendix).

those known for contrast (Carandini et al., 1997), may of course act in addition to this basic mechanism.

#### Additional examination of the temporal properties of the neurons: response onset and phase

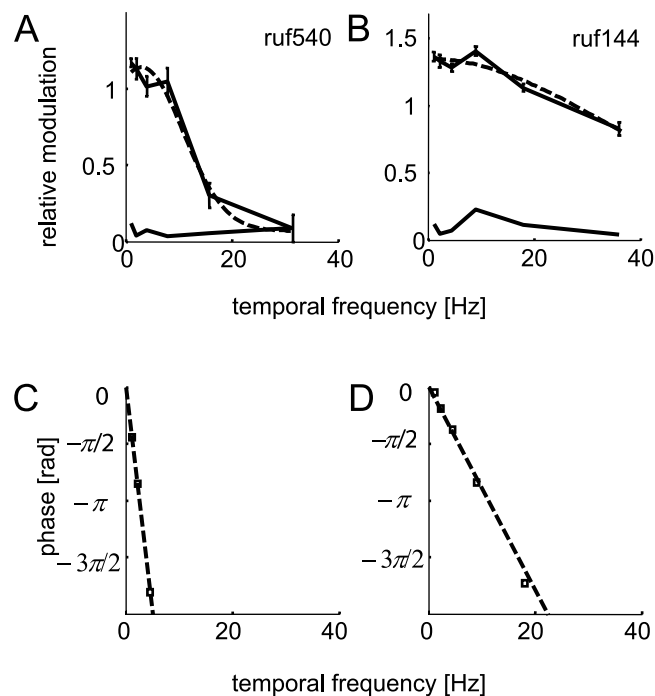
The above scheme (low-pass response to changes in binocular correlation) predicts a reciprocal relationship of the temporal frequency cutoff for disparity modulation with two other properties: (1) the onset rise time [which we quantify with the first point in time after stimulus onset when the response reached 60% its peak (Fig. 6A and Materials and Methods)]; and (2) the integration time deduced from phase lags. Figure 6B shows that neurons with faster initial onsets (shorter times to 60% peak response) were able to modulate at higher temporal frequency ( $r = 0.34$ ;  $p < 0.01$ ;  $n = 51$ ).

According to our explanation, which shows that responses to disparity modulation are low pass, the phase lag should also increase systematically with frequency. The slope of this relationship can be used to estimate the temporal integration time (Reid et al., 1992; Hawken et al., 1996). It corresponds to the sum of the conduction delay and the delays caused by the temporal filtering.

Figure 7 (bottom) plots phase as a function of temporal frequency for two cells. (For comparison, the variation of RM as a function of temporal frequency for the cells is shown in the top row.) To reduce the error caused by noise, we calculated phases only for RM values  $> 0.5$  peak RM. Note that the slope of the line relating phase and temporal frequency (Fig. 7C,D, dashed line) is steeper (i.e., that the cell has a longer temporal integration time) for the cell (ruf540) with the lower temporal frequency cutoff. We computed temporal integration times for all cells for which we had at least three phase values. The mean temporal integration was  $72 \pm 23$  ms SD.

The correlation between the reciprocal of the temporal integration time and the temporal frequency cutoff in response to disparity modulation is highly significant ( $r = 0.59$ ;  $p < 10^{-4}$ ;  $n = 37$ ) (Fig. 8). Note the curved relationship between the cutoff and the temporal integration time. This is expected because the integration time corresponds to the sum of the conduction delays and the time constant of the neuron, whereas the cutoff only





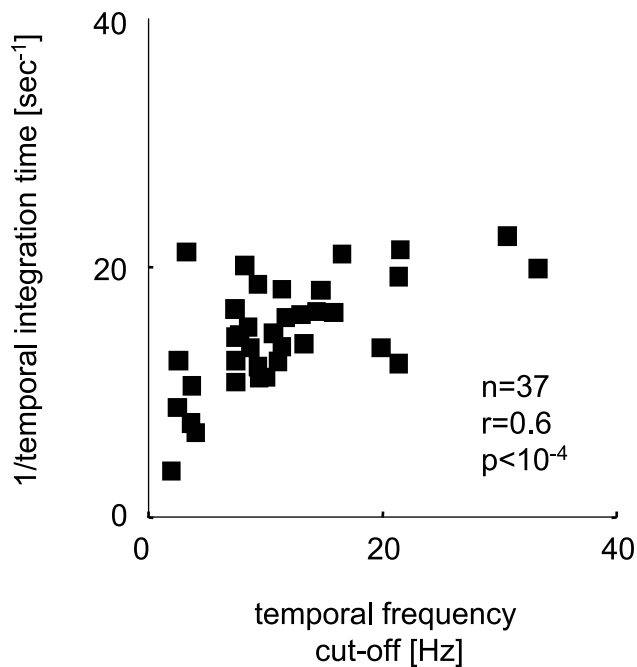
**Figure 7.** Phases of the neuronal responses. **A** and **B** depict two temporal frequency tuning functions in response to disparity modulation, plotting RM (ordinate) as a function of temporal frequency (abscissa). The temporal frequency cutoffs are 10.4 Hz (cell ruf540; **A**) and 35.3 Hz (cell ruf144; **B**). In **C** and **D**, the phase of the averaged SDF is plotted as a function of temporal frequency for the cells in **A** and **B**, respectively. The slope of the line relating phase and temporal frequency (dashed line) will be referred to as temporal integration time (206 and 44 ms for ruf540 and ruf144, respectively). Note the longer temporal integration time (steeper slope) for the neuron with the lower temporal frequency cutoff (ruf540; **A, C**).

depends on the time constant of the neuron. Thus, in Figure 8, as the time constant approaches 0, the cutoff frequency goes toward infinity, whereas the integration time asymptotes to one over the conduction delay.

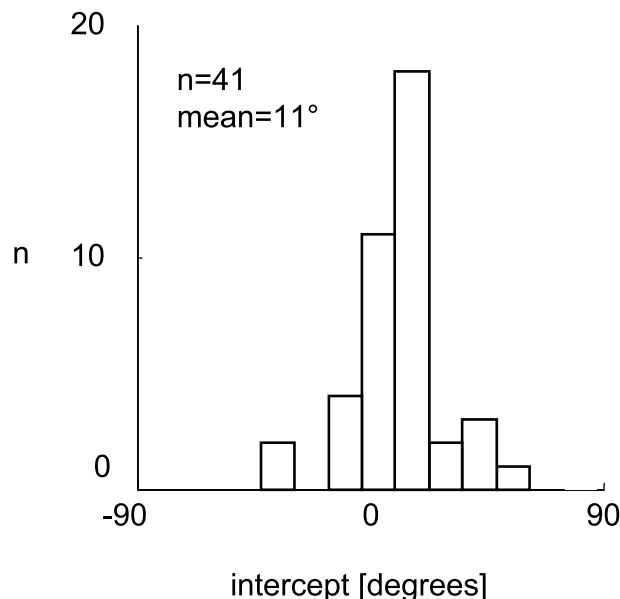
Taking the data from Figures 6 and 8 together, it appears that all aspects of the temporal response to disparity changes in RDSs are compatible with temporal bandpass filtering of broadband input, followed by measurement of the correlation between the filtered images. This results in responses to changes in correlation that are low-pass filtered (compare Fig. 5, Appendix).

**No specialization for signaling motion in depth**

Plotting phase as a function of temporal frequency also allowed us to examine whether the neurons were specialized for signaling motion in depth. A previous study found a small number of macaque V1 neurons sensitive to opposite directions of image motion in the two eyes (Poggio and Talbot, 1981) and suggested that some V1 neurons are tuned for motion in depth. However, as Maunsell and Van Essen (1983) pointed out, when using targets that move in depth, care must be taken that changes in mean disparity alone are not responsible for the tuning observed. These difficulties are all avoided in our analysis in which only the temporal frequency of modulation varies. Sensitivity for motion in depth should result in an additional 90° phase shift independent of stimulus frequency. (In other words, consider a stimulus that is a sine in disparity. A neuron that responds to the rate of change of disparity will thus respond to the first derivative of the stimulus: a cosine.) The intercept of the line relating phase and stimulus frequency would then be at values close to ±90°. In the examples



**Figure 8.** Temporal integration time and temporal frequency cutoff. Temporal frequency cutoffs were significantly correlated with the reciprocal temporal integration time (the slope of the line relating response phase and temporal frequency; see Fig. 7C,D) ( $n = 37; r = 0.59; p < 10^{-4}$ ). The mean temporal integration time is  $72 \pm 23$  ms SD.



**Figure 9.** The intercept of the line relating phase of the response and temporal frequency with the ordinate. The intercept (in degrees) with the ordinate in the phase plots (Fig. 7C,D) is shown in the frequency histogram for 41 cells. Most values are close to 0° (mean of  $10.6 \pm 19.9^\circ$  SD), suggesting that the neurons respond to the instantaneous disparity and not to motion in depth.

in Figure 7, this is not the case. Figure 9 shows the distribution of this intercept for 41 cells. The mean is  $11 \pm 19^\circ$  SD, which is indistinguishable from 0°: no cell has an intercept outside the range of ±65°. Furthermore, at low temporal frequencies, almost all neurons had small phase lags. Thus, V1 neurons appear insensitive to the temporal rate of change of disparity within the receptive field.

## Discussion

We compared neuronal and psychophysical responses to temporal modulation of disparity. Two similarities emerged. First, both types of response showed a comparable high-frequency limitation. Second, the frequency that produced maximal neuronal modulation produced the greatest psychophysical sensitivity. These findings suggest that temporal integration at the stage of V1 neurons may be responsible for the poor temporal resolution of stereopsis.

This result is interesting because the temporal response of V1 neurons also seems to match psychophysical detection of contrast modulation, despite the fact that the psychophysical ability to resolve contrast modulation (de Lange, 1958; Kelly, 1971a,b) is much better than the resolution for disparity modulation (Norcia and Tyler, 1984). This implies that modulation of the input to a cortical neuron by changes of disparity produces a different temporal response function from modulation of the same neuron by contrast. This is confirmed by our comparison of temporal frequency tuning for disparity against that for contrast (Fig. 4).

These results can be explained by modeling the computation of disparity in V1 as a cross-correlation of the bandpass-filtered monocular inputs. The temporal frequency cutoff for contrast modulation is then determined by the frequency response of the temporal kernel. Conversely, the temporal frequency cutoff for disparity modulation corresponds to the cutoff predicted approximately by the width of the bandpass filter. A similar behavior is well known in the auditory system for amplitude-modulated broadband noise (Lawson and Uhlenbeck, 1950; Viemeister, 1979; Dau et al., 1999): the envelope spectrum of bandpass noise is a low-pass function, which can be approximately predicted from the bandwidth of the noise carrier.

It has long been recognized that there are significant temporal nonlinearities in the early visual system (Holub and Morton-Gibson, 1981; Albrecht, 1995; Carandini et al., 1997), which could also contribute to the difference between temporal frequency tuning for contrast and disparity modulation. However, without taking any of these temporal nonlinearities into consideration, we show that the lower temporal resolution for disparity modulation is an inevitable result of computing binocular cross-correlation between temporally broadband images that are bandpass filtered monocularly, before the calculation of cross-correlation.

Our explanation also predicts correlations between two different aspects of the response: the rise time of the initial response and the slope of the relationship between phase lag and frequency. We observe exactly such correlations (Figs. 6B, 8).

Our comparison of psychophysical and neuronal resolution for temporal modulation of disparity is not rigorously based on signal detection theory, so we cannot unequivocally establish disparity-selective V1 neurons as the limiting step. Indeed, some features of our data suggest that V1 neurons may support somewhat better performance than the psychophysics shows. This may be an argument for additional temporal filtering of disparity signals outside V1 (although this may be no more than summing across V1 neurons with different response delays). However, we have established that a striking feature of human depth perception, the poorer temporal resolution for disparity than contrast, has a close parallel in the responses of V1 neurons, and this neuronal property can in turn be explained by a simple mechanism.

In a separate study (Nienborg et al., 2004), we recently demonstrated that the spatial structure of receptive fields in disparity-selective neurons can explain why the spatial resolution of stereopsis is so much poorer than that for contrast. Here we show that

the poor temporal resolution of stereopsis also results from the earliest steps in binocular combination. The fact that the computation of binocular cross-correlation occurs after both spatial and temporal bandpass filtering accounts for the relatively poor spatial and temporal resolution.

That this simple mathematical explanation might hold more generally is supported by evidence from motion detection: when motion perception is isolated from cues about position, detection thresholds of human observers are lowest for  $\sim 2$  Hz and rise steeply above 5 Hz (Nakayama and Tyler, 1981), almost identical to the values observed for detecting temporal disparity modulation (Norcia and Tyler, 1984) (Fig. 3). There is also a close similarity in spatial resolution (Tyler, 1974; Nakayama and Tyler, 1981; Rogers and Graham, 1982; Banks et al., 2004). Both of these features are expected because the inputs that are cross-correlated to construct motion detectors (Adelson and Bergen, 1985) are spatiotemporally similar to those for disparity detectors (Ohzawa et al., 1990).

One other psychophysical phenomenon can also be explained by our results. When human subjects judge the speed of motion in depth, performance is poor when the only cue is changing disparity in dynamic RDSs (Harris and Watamaniuk, 1996). This suggests that the subjects do not have access to explicit neuronal signals about the rate of change of disparity. Examining the phase of neuronal responses to disparity modulation (Fig. 9) suggests that V1 neurons likewise are not sensitive to the rate of change of disparity over time (motion in depth), although, of course, such information could be extracted from the responses of the whole neuronal population.

We have shown that the main aspects of the psychophysical response to disparity modulation are mirrored by the properties of disparity-selective neurons in V1. The difference can be explained by the cross-correlation of the spatiotemporally bandpass-filtered monocular input. Thus, it appears that the mechanism by which signals are generated in the primate striate cortex is responsible for the major temporal and spatial differences between the psychophysical processing of disparity and that of contrast.

## Appendix

### The effect of temporal filtering of the monocular images on the cross-correlation measured between the images

The binocular energy model results in disparity signals that are functionally equivalent to measures of the cross-correlation between monocular images after filtering. This analysis in terms of cross-correlation only applies strictly to the energy model in its original form, with a squaring output nonlinearity. Simulations confirm similar phenomena with other output exponents, or just half-wave rectification. We now consider the effect of a correlation between the monocular inputs that varies as a sinusoidal function of time (as in the RDS we used whose disparity varied sinusoidally over time). In the following, we will examine the effect of temporal filtering of the monocular images on the frequency response of this cross-correlation. We calculate the average correlation between all pixels of the two-dimensional monocular images. Given that we are only applying a temporal filter, we can consider each pixel individually and then average the result. Thus, we begin by considering a single pixel in each eye. We write  $I(t)$  for the luminance of the pixel at time  $t$ ; this may be  $-1, 0, 1$  (white, gray, or black), so the mean value is zero. There are no correlations across time in either monocular sequence, so the luminance of a pixel at time  $t$  is independent of its value at any other time  $t'$  [ $I(t)$  and  $I(t')$  are independent]. However, corre-

sponding pixels in the left and right images,  $I_L(t)$  and  $I_R(t)$ , are correlated at any moment in time, and the correlation between left and right images is changing from moment to moment: it is a function of time  $c(t)$ . We examine here how the correlation between the filtered time series behaves. This analysis considers only the behavior of the product of the left and right responses (cross terms), although the response of the energy model also has terms that depend on the left and right responses alone (Fleet et al., 1996). However, the monocular terms do not vary with disparity and so are less relevant to the analysis here. Furthermore, if one considers the difference in activity between two energy model units, which rearrange the same monocular receptive fields to produce different disparity selectivities, only the cross terms remain.

The pixel luminances  $I_L(t)$ ,  $I_R(t)$  are passed through temporal filters  $\rho_L$  and  $\rho_R$  to obtain the filtered time series  $L(t)$ ,  $R(t)$ :

$$L(t) = \int_{-\infty}^t dt' I_L(t') \rho_L(t' - t)$$

$$R(t) = \int_{-\infty}^t dt'' I_R(t'') \rho_R(t'' - t).$$

To treat the general case, it is necessary to evaluate  $I_L(t)$  and  $I_R(t)$  for different values of  $t$ , so  $t'$  and  $t''$  are used to differentiate time for the left and right eyes. Now consider the value of the product of the filtered time series:

$$L(t) R(t) = \int_{-\infty}^t dt' I_L(t') \rho_L(t' - t) \int_{-\infty}^t dt'' I_R(t'') \rho_R(t'' - t),$$

$$L(t) R(t) = \int_{-\infty}^t dt' \int_{-\infty}^t dt'' \rho_L(t' - t) \rho_R(t'' - t) I_L(t') I_R(t'').$$

We now average this quantity over all pixels in the image. This average over pixels will be written with angle brackets  $\langle \rangle$ .

The original images were uncorrelated from moment to moment (broadband white noise). Under these circumstances,  $\langle I_L(t') I_R(t'') \rangle = 0$  unless  $t' = t''$ . This gives a Dirac delta function (Bracewell, 1986), which means that  $\langle L(t) R(t) \rangle$  now becomes

$$\langle L(t) R(t) \rangle = \int_{-\infty}^t dt' \rho_L(t' - t) \rho_R(t' - t) \langle I_L(t') I_R(t') \rangle.$$

Note that the angle brackets  $\langle \rangle$  imply averaging over pixels, not over time.

The value of  $\langle I_L(t') I_R(t') \rangle$  depends on the instantaneous correlation between the images. This is  $c(t')$ . So,

$$\langle L(t) R(t) \rangle = \int_{-\infty}^t dt' \rho_L(t' - t) \rho_R(t' - t) c(t'),$$

or

$$\langle L(t) R(t) \rangle = \int_{-\infty}^t dt' \rho^2(t' - t) c(t'),$$

if the temporal kernels of the left and right receptive fields are identical.

This is the correlation time series convolved with the product of the left and right temporal receptive fields. Thus, the Fourier transform of  $\langle L(t) R(t) \rangle$  is the product of the Fourier transform of  $(\rho_L \rho_R)$  (or  $\rho^2$  if  $\rho_L = \rho_R$ ) and the Fourier transform of  $c(t)$ .

If the correlation time series is a sine wave,  $c(t')$ , then  $\langle L(t) R(t) \rangle$  will be a sine wave too. The amplitude of the sine wave will be the amplitude of the Fourier transform of  $(\rho_L \rho_R)$  or  $\rho^2$  at that frequency. Suppose that  $\rho$  is a temporal bandpass filter. The squared filter ( $\rho^2$ ) then is a low-pass filter and approximates the squared envelope of the filter, regardless of the temporal structure within it. As a consequence, the monocular bandpass filters give rise to a low-pass response to correlation changes, with lower cutoffs.

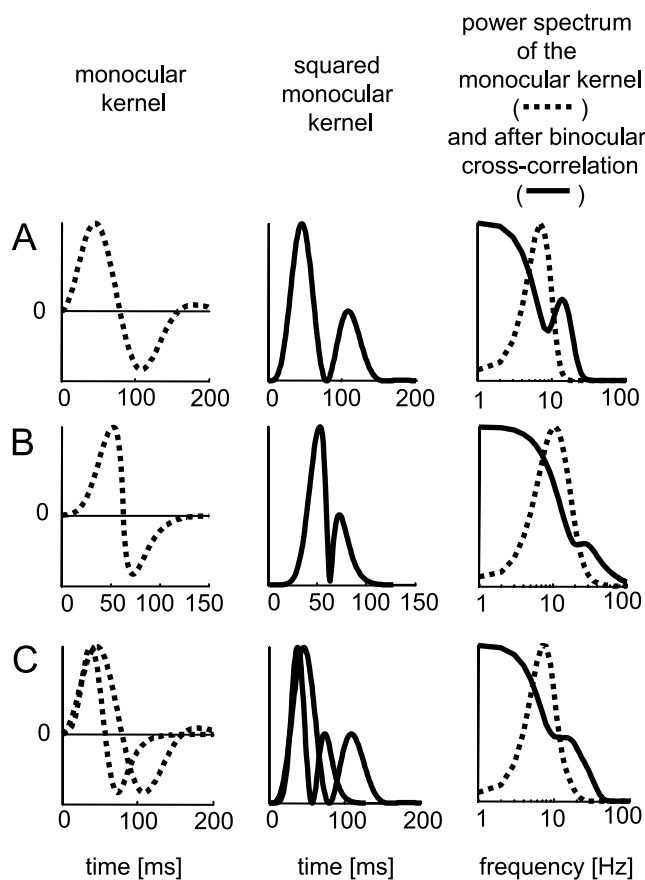
Mathematically, this is closely similar to the analysis of envelope detection in the auditory system. If an amplitude-modulating broadband noise stimulus is subjected to half-wave rectification, the same relationship holds between the frequency spectrum of the input and the frequency response of the envelope detector (Lawson and Uhlenbeck, 1950; Viemeister, 1979; Dau et al., 1999).

When comparing model responses with luminance gratings and with disparity modulation, it is important to note that, for the energy model, because of the squaring output nonlinearity, the response amplitude to contrast for any one temporal frequency corresponds to the square of that amplitude in the spectrum of the temporal filters (power spectrum of  $\rho_L$  and  $\rho_R$ ).

Manipulating different parameters of the monocular kernel therefore has the following effects. First, for a given envelope, increasing the peak frequency of the monocular filter will have little effect on the frequency response to changes of correlation. Second, if one keeps the peak frequency of the monocular filter constant, decreasing the bandwidth (making the envelope wider) will shift the frequency response to changes of correlation toward lower frequencies. Third, for a given kernel shape, rescaling it (i.e., compressing or expanding it on the time axis while keeping the relative size of the positive and negative lobes constant) will yield shifts in the same direction for the temporal frequency tuning for correlation and contrast. For a given kernel shape, this then results in correlation between the temporal frequency high cutoff for correlation and contrast across different scales of the kernel.

Note, however, that if only one subunit is used, the exact form of the responses to correlation modulation is quite sensitive to the shape of the temporal kernel used. In Figure 10, we show that, for the simplified case of only one subunit (identical temporal kernels in the left and right eye), the shape of the bandpass temporal kernel must be chosen with care to produce a smooth low-pass response to binocular correlation. However, when input is averaged over several subunits with slightly different temporal kernels, the smooth low-pass response to changes in binocular correlation becomes a general feature.

Figure 10 compares monocular and binocular responses for several temporal kernels. The kernel in row A is similar to a typical temporal impulse response for the LGN. Note that, because of the squaring, the frequency response for the correlation is low pass but, unlike real neurons, has a second peak at double the peak frequency of the monocular kernel. In row B, a very similar physiologically plausible kernel is used, but the transition between the positive and negative lobes is steeper. This results in a squared kernel that is more similar to a single Gaussian function. As a consequence, the frequency spectrum has only very little power at twice the frequency of the bandpass kernel, and the cutoff of the monocular kernel is approximately at double the



**Figure 10.** The role of the shape of the bandpass filter for the binocular energy model. In the first and second columns, the monocular kernel (dotted line) and the squared monocular kernel (solid line) are shown, respectively. In column three, the frequency response of the squared monocular kernel (solid line) is superimposed on the monocular frequency response (dotted line). Row **A**, The kernel used resembles that known for the LGN. The frequency response for the squared kernel is low pass but has a second peak at double the peak frequency of the monocular kernel. Row **B**, the kernel is modified to have a steeper transition between the positive and negative lobes, such that the squared kernel is close to a single asymmetrical Gaussian function. The frequency response of the squared kernel therefore has only a small second peak and a cutoff at approximately half the cutoff frequency of the monocular kernel. Row **C**, Input from two subunits with slightly different temporal kernels (peak frequencies are 7 and 10 Hz, respectively) converge onto a binocular cell. The frequency response of the sum of the squared monocular kernels has only very small peaks at double the peak frequencies of the monocular kernel and a cutoff at approximately half the cutoff frequency of the sum of the monocular kernels.

frequency of that of the cross-correlation (just as we found for the neurons).

Thus, the relationship between responses to contrast modulation and responses to correlation modulation is sensitive to the choice of temporal kernel, if a single monocular kernel is used. However, real neurons receive a large number of inputs. Figure 10, row **C**, shows that, when a model binocular neuron receives input from two subunits, the choice of kernel becomes less critical. Here two slightly different kernels are used, with peak frequencies at 7 and 10 Hz. In each subunit, the kernel in the right eye is identical to that in the left eye. In the frequency response of the sum of the subunits, the two second peaks at double the frequencies of the monocular kernels disappear almost entirely, resulting in a lower temporal frequency cutoff for the binocular correlation than for the sum of the temporal kernels. In this way, a variety of physiologically plausible temporal kernels can be reconciled with smooth low-pass responses to modulation of binocular correlation, like those we observed.

## References

- Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A* 2:284–299.
- Albrecht DG (1995) Visual cortex neurons in monkey and cat: effect of contrast on the spatial and temporal phase transfer functions. *Vis Neurosci* 12:1191–1210.
- Anzai A, Ohzawa I, Freeman RD (1999) Neural mechanisms for processing binocular information. II. Complex cells. *J Neurophysiol* 82:909–924.
- Banks MS, Gepshtein S, Landy MS (2004) Why is spatial stereoresolution so low? *J Neurosci* 24:2077–2089.
- Barlow HB, Blakemore C, Pettigrew JD (1967) The neural mechanism of binocular depth discrimination. *J Physiol (Lond)* 193:327–342.
- Bracewell RN (1986) *The Fourier transform and its applications*, Ed 2. Singapore: McGraw-Hill.
- Bredfeldt CE, Ringach DL (2002) Dynamics of spatial frequency tuning in macaque V1. *J Neurosci* 22:1976–1984.
- Carandini M, Heeger DJ, Movshon JA (1997) Linearity and normalization in simple cells of the macaque primary visual cortex. *J Neurosci* 17:8621–8644.
- Cumming BG, Judge SJ (1986) Disparity-induced and blur-induced convergence eye movement and accommodation in the monkey. *J Neurophysiol* 55:896–914.
- Cumming BG, Parker AJ (1997) Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature* 389:280–283.
- Cumming BG, Parker AJ (1999) Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity. *J Neurosci* 19:5602–5618.
- Cumming BG, Parker AJ (2000) Local disparity not perceived depth is signaled by binocular neurons in cortical area V1 of the macaque. *J Neurosci* 20:4758–4767.
- Dau T, Verhey J, Kohlrausch A (1999) Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers. *J Acoust Soc Am* 106:2752–2760.
- Davison AC, Hinkley DV (1997) *Bootstrap methods and their application*. Cambridge, UK: Cambridge UP.
- de Lange H (1958) Res into the dynamic nature of the human fovea-cortex systems with intermittent and modulated light. II. Phase shift in brightness and delay in color perception. *J Opt Soc Am A* 48:784–789.
- Fleet DJ, Wagner H, Heeger DJ (1996) Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision Res* 36:1839–1857.
- Harris JM, Watamaniuk SN (1996) Poor speed discrimination suggests that there is no specialized speed mechanism for cyclopean motion. *Vision Res* 36:2149–2157.
- Hawken MJ, Shapley RM, Grosof DH (1996) Temporal frequency selectivity in monkey visual cortex. *Vis Neurosci* 13:477–492.
- Holub RA, Morton-Gibson M (1981) Response of visual cortical neurons of the cat to moving sinusoidal gratings: response-contrast functions and spatiotemporal interactions. *J Neurophysiol* 46:1244–1259.
- Judge SJ, Richmond BJ, Chu FC (1980) Implantation of magnetic search coils for measurement of eye position: an improved method. *Vision Res* 20:535–538.
- Kelly DH (1961) Flicker fusion and harmonic analysis. *J Opt Soc Am* 51:917–918.
- Kelly DH (1971a) Theory of flicker and transient responses. I. Uniform fields. *J Opt Soc Am A* 61:537–546.
- Kelly DH (1971b) Theory of flicker and transient responses. II. counter-phase gratings. *J Opt Soc Am A* 61:632–640.
- Kelly DH, Boynton RM, Baron WS (1976) Primate flicker sensitivity: psychophysics and electrophysiology. *Science* 194:1077–1079.
- Lagarias JC, Reeds JA, Wright MH, Wright PE (1998) Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM J Optimization* 9:112–147.
- Lawson JL, Uhlenbeck GE (1950) Threshold signals. In: *Radiation laboratory series*. New York: McGraw-Hill.
- Maunsell JH, Van Essen DC (1983) Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J Neurophysiol* 49:1148–1167.
- Merrill EG, Ainsworth A (1972) Glass-coated platinum-plated tungsten microelectrodes. *Med Biol Eng* 10:662–672.
- Movshon JA, Thompson ID, Tolhurst DJ (1978a) Spatial summation in the

- receptive fields of simple cells in the cat's striate cortex. *J Physiol (Lond)* 283:53–77.
- Movshon JA, Thompson ID, Tolhurst DJ (1978b) Receptive field organization of complex cells in the cat's striate cortex. *J Physiol (Lond)* 283:79–99.
- Nakayama K, Tyler CW (1981) Psychophysical isolation of movement sensitivity by removal of familiar position cues. *Vision Res* 21:427–433.
- Nienborg H, Bridge H, Parker AJ, Cumming BG (2004) Receptive field size in V1 neurons limits acuity for perceiving disparity modulation. *J Neurosci* 24:2065–2076.
- Norcia AM, Tyler CW (1984) Temporal frequency limits for stereoscopic apparent motion processes. *Vision Res* 24:395–401.
- Ohzawa I, DeAngelis GC, Freeman RD (1990) Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science* 249:1037–1041.
- Pettigrew JD, Nikara T, Bishop PO (1968) Binocular interaction on single units in cat striate cortex: simultaneous stimulation by single moving slit with receptive fields in correspondence. *Exp Brain Res* 6:391–410.
- Poggio GF, Talbot WH (1981) Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey. *J Physiol (Lond)* 315:469–492.
- Poggio GF, Doty Jr RW, Talbot WH (1977) Foveal striate cortex of behaving monkey: single-neuron responses to square-wave gratings during fixation of gaze. *J Neurophysiol* 40:1369–1391.
- Poggio GF, Motter BC, Squatrito S, Trotter Y (1985) Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms. *Vision Res* 25:397–406.
- Prince SJ, Pointon AD, Cumming BG, Parker AJ (2000) The precision of single neuron responses in cortical area V1 during stereoscopic depth judgments. *J Neurosci* 20:3387–3400.
- Prince SJ, Pointon AD, Cumming BG, Parker AJ (2002) Quantitative analysis of the responses of V1 neurons to horizontal disparity in dynamic random-dot stereograms. *J Neurophysiol* 87:191–208.
- Qian N, Zhu Y (1997) Physiological computation of binocular disparity. *Vision Res* 37:1811–1827.
- Read JC, Cumming BG (2003) Measuring V1 receptive fields despite eye movements in awake monkeys. *J Neurophysiol* 90:946–960.
- Reid RC, Victor JD, Shapley RM (1992) Broadband temporal stimuli decrease the integration time of neurons in cat striate cortex. *Vis Neurosci* 9:39–45.
- Reid RC, Victor JD, Shapley RM (1997) The use of m-sequences in the analysis of visual neurons: linear receptive field properties. *Vis Neurosci* 14:1015–1027.
- Ringach DL, Hawken MJ, Shapley R (1997) Dynamics of orientation tuning in macaque primary visual cortex. *Nature* 387:281–284.
- Rogers B, Graham M (1982) Similarities between motion parallax and stereopsis in human depth perception. *Vision Res* 22:261–270.
- Tyler CW (1974) Depth perception in disparity gratings. *Nature* 251:140–142.
- Viemeister NF (1979) Temporal modulation transfer functions based upon modulation thresholds. *J Acoust Soc Am* 66:1364–1380.
- Williams PE, Mechler F, Gordon J, Shapley R, Hawken MJ (2004) Entrainment to video displays in primary visual cortex of macaque and humans. *J Neurosci* 24:8278–8288.