Detectability of sine- versus square-wave disparity gratings: A challenge for current models of depth perception

Fredrik Allenmark Institute of Neuroscience, Newcastle University, UK

Jenny C. A. Read Institute of Neuroscience, Newcastle University, UK

Stereo vision is an area in which we are increasingly able to construct detailed numerical models of the computations carried out by cerebral cortex. Piecewise-frontoparallel cross-correlation is one such model, closely based on the known physiology and able to explain important aspects of human stereo depth perception. Here, we show that it predicts important differences in the ability to detect disparity gratings with square-wave vs. sine-wave profiles. In particular, the model can detect square-wave gratings up to much higher disparity amplitudes than sine-wave gratings. We test this prediction in human subjects and find that it is not borne out. Rather there seems to be little or no difference between the detectability of square- and sine-wave disparity gratings for human subjects. We conclude that the model needs further refinement in order to capture this aspect of human stereo vision.

Keywords: binocular vision, stereo vision, correspondence problem, correlation

Citation: Allenmark, F., & Read, J. C. A. (2010). Detectability of sine- versus square-wave disparity gratings: A challenge for current models of depth perception. *Journal of Vision, 10*(8):17, 1–16, http://www.journalofvision.org/content/10/8/17, doi:10.1167/10.8.17.

Introduction

Stereopsis, the ability to estimate 3D depth based on binocular vision, is one of the best understood aspects of human perception. One hundred and fifty years of psychophysical experiments have documented in detail how binocular disparities between the eyes result in a depth percept (Howard & Rogers, 1995), while in the last two decades, physiological experiments have mapped how disparities drive the firing rates of individual neurons in visual cortex (Roe, Parker, Born, & DeAngelis, 2007). Stereo vision has thus emerged as a paradigm for relating perceptual experience to neuronal activity.

A recent, highly successful example has been the development of a computational model explaining the spatial resolution of stereopsis in terms of the properties of neurons in primary visual cortex (Banks, Gepshtein, & Landy, 2004; Filippini & Banks, 2009; Nienborg, Bridge, Parker, & Cumming, 2004). Stereo spatial resolution is traditionally assessed using sinusoidal "disparity gratings," corrugated surfaces that go back and forth in depth (Figure 1A). The upper frequency limit at which such disparity gratings can be perceived has been found to be around 3–4 cycles per degree (Banks et al., 2004; Bradshaw & Rogers, 1999; Filippini & Banks, 2009; Tyler, 1974), much lower than the corresponding limit for luminance gratings, which can be as high as 50–60 cpd under optimal luminance conditions (Campbell & Green, 1965). In a linked pair of papers, Banks et al. (2004) and Nienborg et al. (2004) explained this limit in terms of the receptive field size of disparity-selective neurons in primary visual cortex (V1).

Their analysis was based on the stereo energy model, in which disparity is encoded by a local cross-correlation between the two eyes' images (Banks et al., 2004; Filippini & Banks, 2009; Nienborg et al., 2004; Ohzawa, DeAngelis, & Freeman, 1990, 1997). In this model, interocular correlation is measured locally within a finite window corresponding to the neuronal receptive field, and the stereoresolution limit is determined by the smallest window size available. When the frequency is high enough that the disparity changes significantly within this window, the effective interocular correlation is reduced and the signal is lost in the noise (compare Figure 2B with Figure 2A). It is this which eventually limits the ability to resolve the grating. Banks, Cumming, and colleagues showed that the stereoresolution of human and monkey observers was remarkably consistent with the size of receptive fields in V1. Thus, the local cross-correlation model is a noteworthy example of how perceptual abilities can be successfully related to the properties of nerve cells recorded in cerebral cortex.

An important feature of this model is that the initial encoding of disparity is piecewise frontoparallel. That is, the model neurons respond best when the disparity within their receptive field is constant. This explains why the resolution for disparity gratings is so much lower than for luminance gratings. V1 receptive fields typically have several different ON or OFF subregions, which respond to

doi: 10.1167/10.8.17 Received January 26, 2010; published July 23, 2010 ISSN 1534-7362

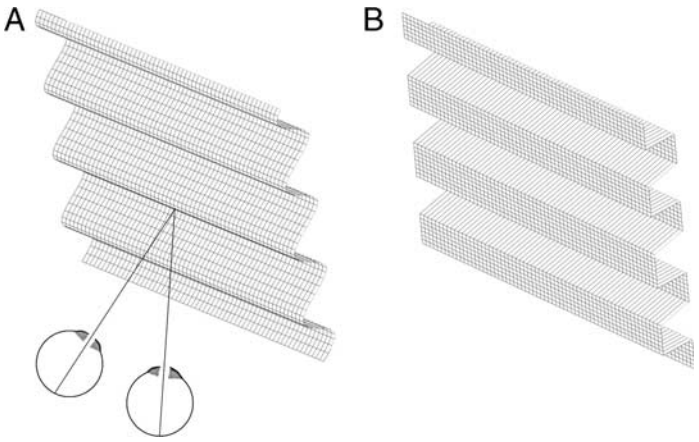

Figure 1. Physical surfaces implied by (A) sine-wave and (B) square-wave disparity gratings.

different luminance polarities. The limiting period for luminance gratings reflects the size of these subregions, not the receptive field as a whole. In contrast, in the stereo domain, V1 receptive fields appear to prefer uniform disparity (Nienborg et al., 2004).

Sine-wave disparity gratings (Figure 1A) are always a suboptimal stimulus for this population, since their disparity is never even locally constant. Square-wave disparity gratings, on the other hand, consist of regions of locally constant disparity (Figure 1B). When the grating's period exceeds the window size used for local cross-correlation, the disparity within the window is constant. Neurons with the optimal tuning (black ellipses in Figure 2) should thus experience an interocular correlation of near unity. Critically, this statement is true independent of the grating's amplitude (compare Figure 2C with Figure 2A, blue curves). In contrast, for sine-wave gratings, the range

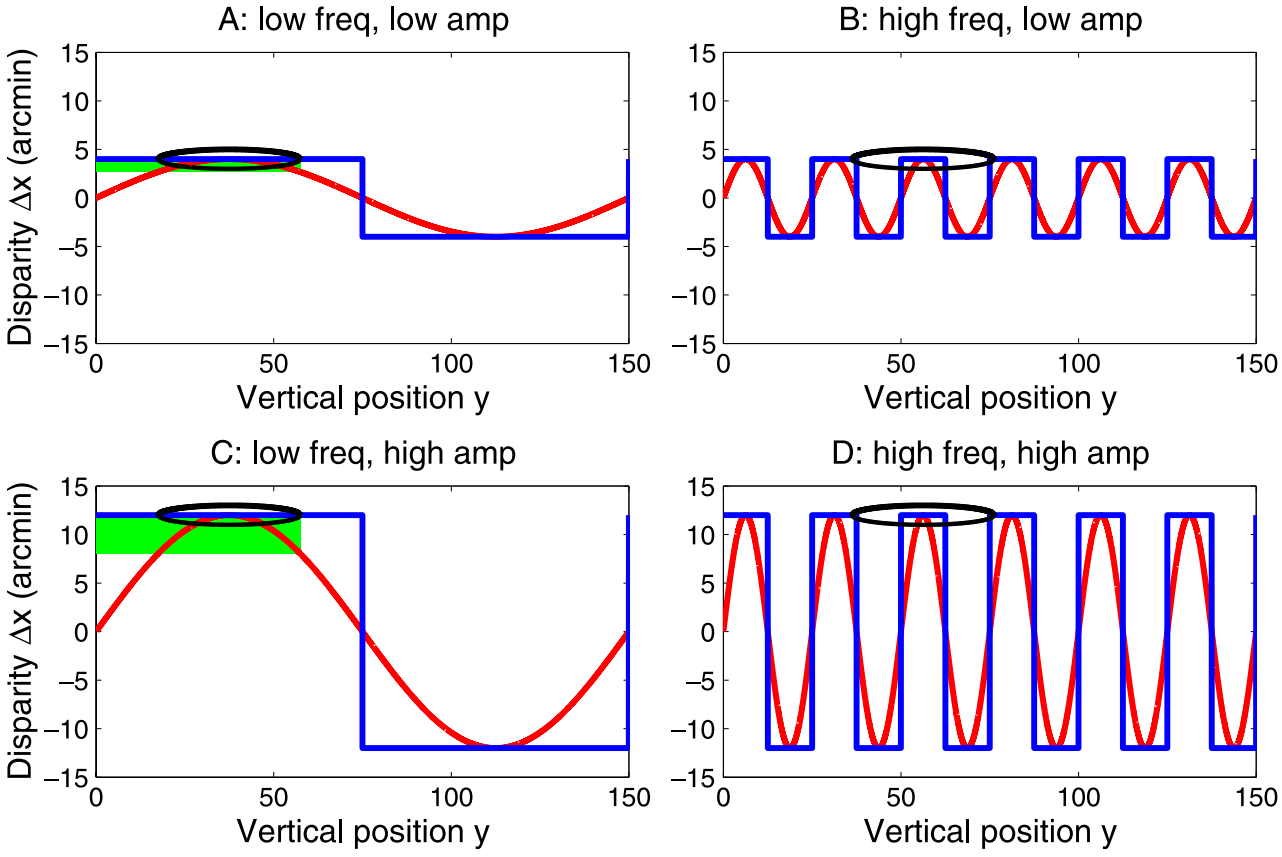

Figure 2. Sketch of sine- and square-wave gratings and a receptive field. The red and blue curves show the profile of (respectively) sine- and square-wave disparity gratings, with disparity varying as a function of vertical position in the visual field. The 4 panels show gratings with low (A, C) and high (B, D) spatial frequencies and with small (A, B) and large (C, D) amplitudes. The black ellipse shows the receptive field of a model neuron tuned to the largest disparity in the grating. At low frequencies (A, C), the period of the grating is large compared to the correlation window (black lines), and the grating can be resolved. At high frequencies (B, D), the period is small compared to the window, and the grating cannot be perceived. At low frequencies (A, C), the square wave presents only a single disparity within the local correlation window. This is not so for the sine wave. For low amplitudes, the range of disparities within the receptive field is small (green shaded region in A), but as the amplitude increases, this range increases (green shaded region in C), even for low-frequency sine waves.

of different disparities falling in a window depends on the amplitude of the grating (compare Figure 2C with Figure 2A, red curves and green-shaded regions).

From this qualitative argument, we expected that the piecewise-frontoparallel model should find it easier to detect square-wave disparity gratings than sine-wave gratings, especially at high frequencies and/or amplitudes. If this prediction were borne out in human observers, this would be a powerful confirmation of the model. In this paper, then, we first carry out computer simulations to establish whether the piecewise-frontoparallel model really does respond better to square-wave than sine-wave disparity gratings, and whether this is sensitive to the precise way in which the model is implemented. We next carry out psychophysical experiments to compare human performance to the predictions of the piecewise-frontoparallel model.

Methods

Psychophysics

Experimental setup

The experiments were performed using a mirror stereoscope. The stimuli were displayed on the left and right halves of a single LCD monitor with a physical display size of 41 × 25.5 cm and a resolution of 1440 × 900 pixels. The size of the images was 350 × 350 pixels. With the viewing distance of 308 cm, the images subtended 1.8° × 1.8° and each pixel subtended 0.3 arcmin. The mirrors were aligned to make the vergence distance 308 cm. The monitor was linearized (gamma-corrected) using a Minolta LS-100 photometer. White pixels were 240 cd/m^2 and black pixels were 0.26 cd/m^2.

Stimuli

Stimuli were presented using Matlab (The Mathworks, Natick, MA, USA; www.mathworks.com) with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). The grating stimuli used were random dot stereograms depicting horizontally oriented depth corrugations with either sine-wave or square-wave profiles (Figure 1). We varied the amplitude, frequency, and phase of the gratings. Amplitude is defined as half the peak-to-trough range of the waveform, (max − min)/2, except in the Frequency analysis section, where the amplitude of the fundamental is specified. The dots were 2 × 2 pixels, 0.6 × 0.6 arcmin, and were white on a black background. Anti-aliasing, implemented in-house in our own Matlab code, was used to place dots at subpixel locations. The long viewing distance (308 cm) and small pixel size (0.3 arcmin, less than the retinal cone spacing) were used to ensure that the range of disparities and frequencies perceived by human observers was not limited by the resolution of the display. For the highest grating frequencies used in this study (5.7 cpd), sine- and square-wave profiles could be readily perceived and distinguished from one another when the stimuli were viewed up close in anaglyph, although they became invisible as the observer walked further away. This demonstrates that the limits on grating detectability were contained in the observer's visual system, not the physical display.

Task

A two-interval forced-choice task was used, where one temporal interval contained a disparity grating and the other contained disparity noise (described below). The task was to report, by a button press, which interval contained the disparity grating. For three subjects, the length of the temporal intervals was 500 ms, with a 100-ms blank between the intervals. A fourth subject was allowed to view each of the intervals for as long as he wanted before making a choice. Experimental trials were organized in blocks, most of which consisted of 240–280 trials, where the frequency of the gratings was kept constant in each block. The two waveforms and the different phases were always interleaved in blocks of experimental trials and in most cases different amplitudes were interleaved as well.

On each trial, the disparity noise image was generated by assigning each dot a disparity drawn at random from the same distribution as the disparity grating presented in the other interval. Thus, in trials where the grating was a square wave with amplitude A, the disparity noise dots had disparity $+A$ or $-A$ with equal probability. On sine-wave trials, they had a disparity in the range $[-A, +A]$. In the grating stimuli, all dots at a given vertical position had the same disparity, but in the noise stimuli, disparity was picked without reference to vertical position, so dots in the same row would have different disparities.

Observers

The 4 observers were the two authors, one additional experienced psychophysical observer and one inexperienced observer.

Data analysis

A truncated probability density function of a gamma distribution was fit to the data for each frequency. This was simply a descriptive function without any theoretical significance. The Matlab function FIT, using non-linear least squares, was used to do the fitting.

Model

Stimuli and task

The same stimuli that were used in the psychophysics were also used in the modeling. The model had the same

task as the human subjects: in each trial it was presented with two image pairs, one containing a grating and one containing a noise pattern and it had to judge which one contained the grating.

Preprocessing

The model used here was based on the piecewise-frontoparallel local cross-correlation-based model of Banks et al. (2004). The left- and right-eye images were first preprocessed to simulate the effects of the eye's optics, and then passed to a cross-correlator.

The preprocessing consisted of convolving the images with the point-spread function of the well-focused eye:

$$h(x,y) = a * h_1(x,y) + (1-a) * h_2(x,y), \quad (1)$$

where

$$h_i(x,y) = (s_i\sqrt{2\pi})^{-2} e^{-0.5(x^2+y^2)/s_i^2}, \quad (2)$$

and a = 0.583, s_1 = 0.443 arcmin, and s_2 = 2.04 arcmin (Filippini & Banks, 2009; Geisler & Davila, 1985). The images were then scaled to make the distance between rows and columns 0.6 arcmin. This was done to make sure the resolution of the images was no higher than the spacing between cones at the fovea (Filippini & Banks, 2009; Geisler & Davila, 1985; Rossi & Roorda, 2009).

Cross-correlator

The preprocessed images were then passed to the cross-correlator. A window was moved along a vertical line in one eye's image. For each vertical position of that window, a second window in the other image at the same vertical position was moved across an interval of horizontal positions centered on the horizontal position of the first window. For each combination of window positions, the correlation between the contents of the windows was recorded. The correlation was defined as

$$C(y,\Delta x) = \frac{\mathrm{cov}(L_w, R_w)}{\sqrt{\mathrm{cov}(L_w, L_w)\mathrm{cov}(R_w, R_w)}}, \quad (3)$$

where L_w and R_w are the contents of the windows in the left and right images multiplied by the window function and cov is the covariance. The window functions used to obtain the main results presented here were Gaussians centered on the current window position (that is, $(\Delta x/2, y)$ in one eye and $(-\Delta x/2, y)$ in the other) and cut off two standard deviations from the center in each direction. The output from the cross-correlator was a two-dimensional image of correlation as a function of the horizontal disparity, Δx, between the windows as well as the vertical position of the windows, y (see Figure 3). The disparities used were in the range from −25 to 25 arcmin with a step of 0.6 arcmin (1 pixel in the scaled images). The step in the y-position was also 1 pixel in the scaled images.

Decision rule 1: Autocorrelation

Two different methods were used to make a decision on which interval contained the gratings based on the correlation images. The first was based on autocorrelation, and the second on template matching. The method based on autocorrelation started by finding the maximum

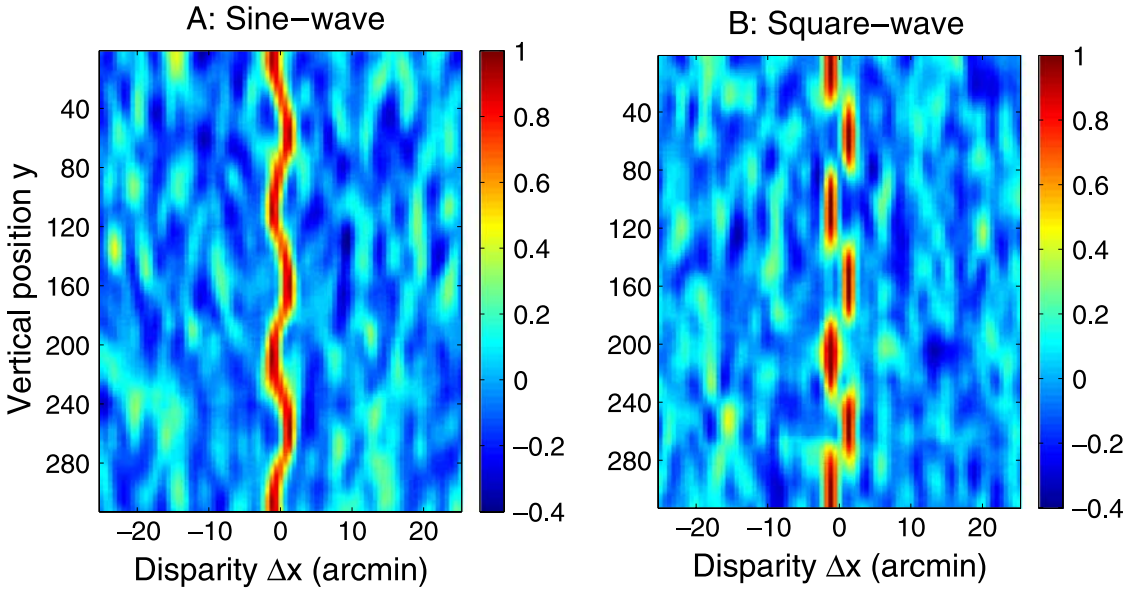

Figure 3. Examples of output from the cross-correlator for one sine-wave and one square-wave grating, both with a frequency of 1.9 cpd. A Gaussian window with 2 * σ = 6 arcmin was used.

correlation across all horizontal window positions, Δx, for each vertical window position, y, and recording the difference in horizontal position between the two windows as an estimate of the horizontal disparity at that vertical position:

$$\Delta x_{\text{est}}(y) = \arg\max(C(y, \Delta x)). \quad (4)$$

The autocorrelation of the resulting curve of estimated disparity as a function of vertical position, $\Delta x_{\text{est}}(y)$ was then calculated as

$$ac_n = \frac{\sum_{i=1}^{N-n}(\Delta x_{\text{est}}(y_i) - \mu)(\Delta x_{\text{est}}(y_{i+n}) - \mu)}{((N-n) * \sigma^2)}, \quad (5)$$

where μ is the mean and σ is the standard deviation of Δx_{est}. Two examples of what the autocorrelograms looked like are given in Figure 4. Finally, both a sine wave and a triangular wave, which are the autocorrelation functions of a sine wave and a square wave, respectively, with the same frequency used in the stimulus were fit to the autocorrelogram and the r^2 value of the best fit was recorded. For each pair of a wave and a noise pattern, making up a single trial, the image pair that got the highest r^2 value was guessed to contain the grating (Figure 5).

Decision rule 2: Template matching

This method used a set of templates of the correlator output for the disparity gratings (grating templates) as well as a set of templates of the correlator output for the two types of noise patterns (noise templates). The set of grating templates covered all frequencies, amplitudes, and phases used in the simulations as well as both waveforms. The set of noise templates covered all the amplitudes (the noise patterns were by their nature independent of frequency and phase). For each interval, the grating template and the noise template with the highest correlation to the correlator output were chosen. The correlations were calculated as follows:

$$C_n = \frac{\sum((CO(\Delta x, y) - \mu_{CO}))(T_n(\Delta x, y) - \mu_{T_n}))}{\sqrt{\sum(CO(\Delta x, y) - \mu_{CO})^2}\sqrt{\sum(T_n(\Delta x, y) - \mu_{T_n})^2}}, \quad (6)$$

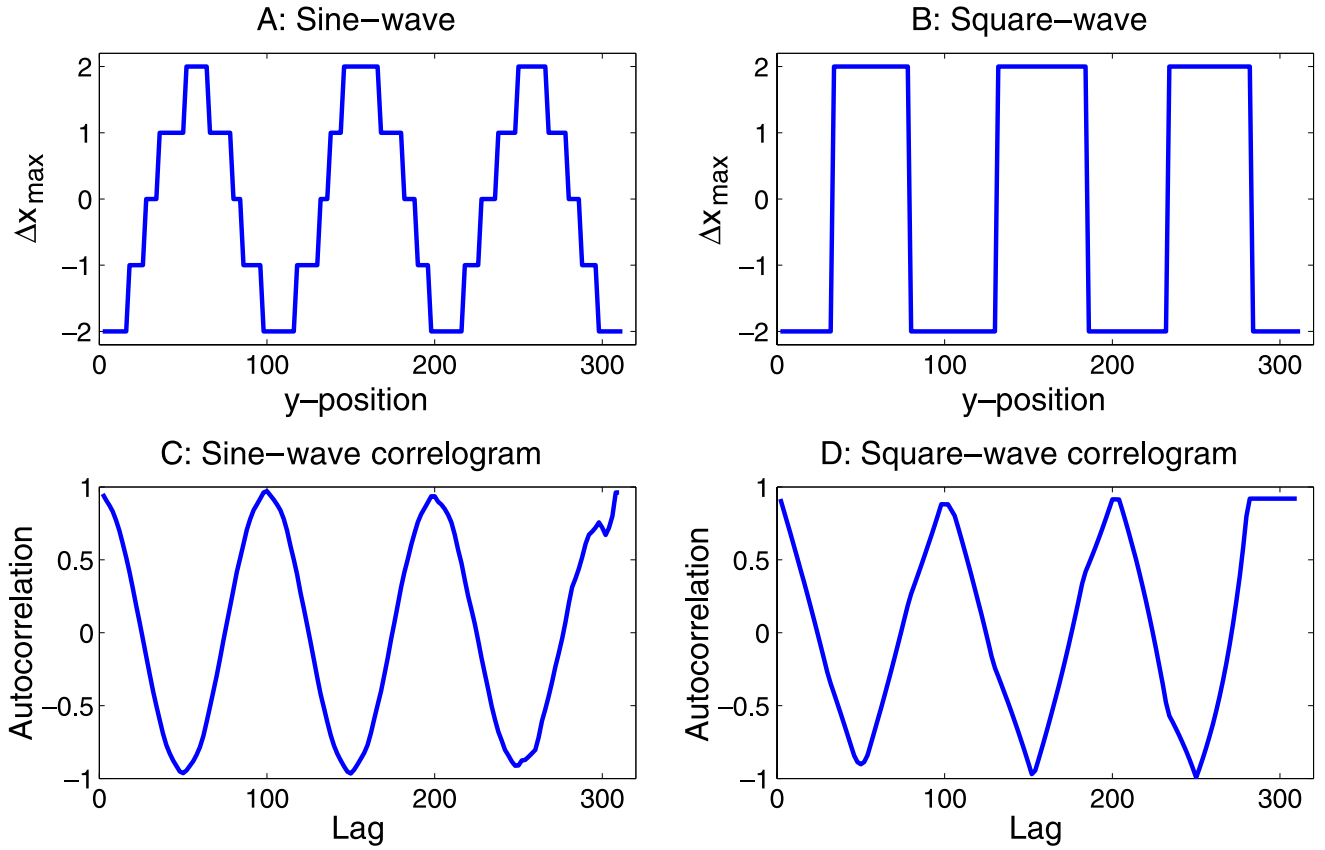

Figure 4. Examples of estimated disparity curves and their autocorrelograms for one square wave and one sine wave both with a frequency of 1.9 cpd. A Gaussian window with 2σ = 6 arcmin was used. The estimated disparity curve for the sine-wave grating is quantized because the model only included detectors tuned to integer pixel disparities.

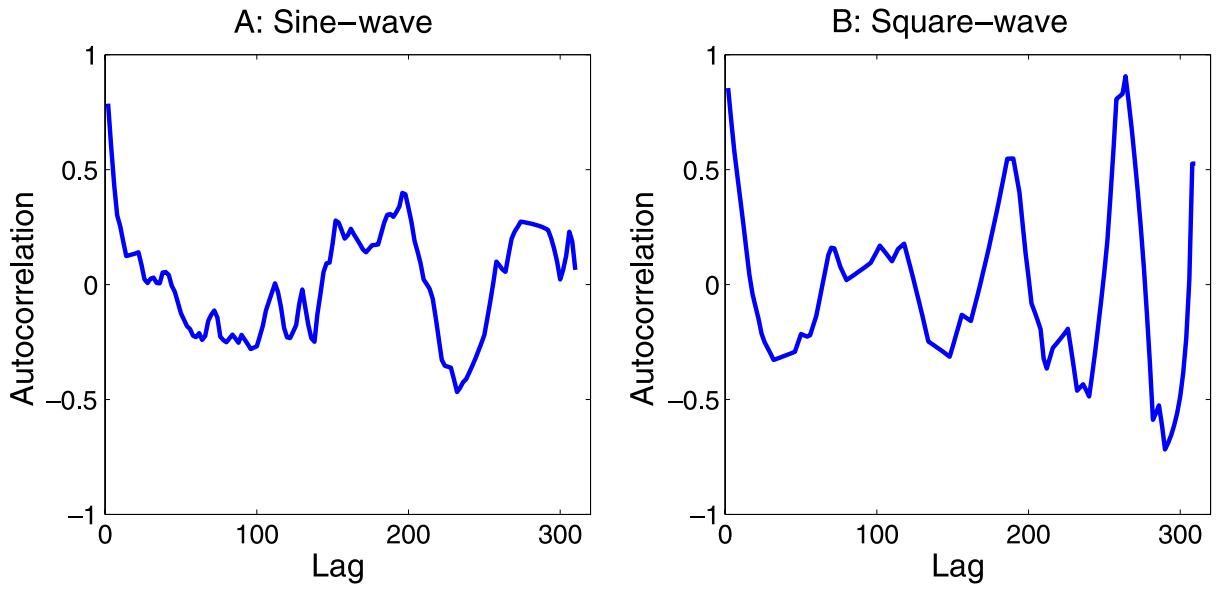

Figure 5. Example of autocorrelation for the corresponding noise patterns to one square wave and one sine wave. A Gaussian window with 2 * σ = 6 arcmin was used.

where CO is the correlator output, T_n is one of the templates, μ_{CO} and μ_{Tn} are the means over all disparities Δx and all y-positions of the correlator output and template T_n, respectively. All sums were performed over disparity and y-position. The interval for which the difference between the correlation to the grating template and the correlation to the noise template was the highest was guessed to contain the grating.

For each grating profile (sine vs. square wave), frequency, amplitude, and phase, the corresponding template was generated by presenting 100 different random dot stereograms to the cross-correlator, after the same preprocessing steps used in the main model. The resulting set of 100 correlation images were then used to calculate the average for each pixel (see Figure 6). The phase of the disparity gratings was varied in steps of 10° when generating the templates and when testing the model the phase was randomly chosen at each trial to be one of the 36 different phases represented in the set of templates. The template amplitudes were 0.3, 1.3, 2.5, 5.1, 7.6, 10.1, 15.2, and 20.2 arcmin. The template frequencies were 1.9, 2.5, 3.2, 3.8, 4.4, 5.1, 5.7, 6.3, 7.0, and 7.6 cpd. Thus, there were 5760 grating templates and 16 noise templates.

We have also examined a somewhat different template matching rule, where the correlator output was matched only to templates of the same frequency, where no noise

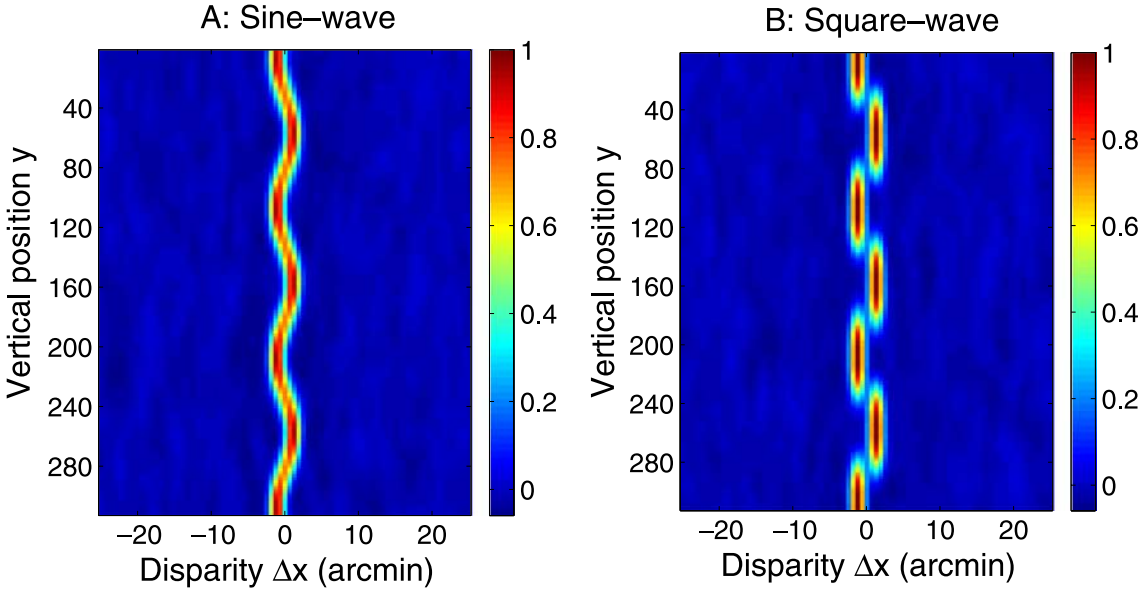

Figure 6. Examples of templates for (left) sine waves and (right) square waves with a frequency of 1.9 cpd.

templates were used and where the matching was based on sums of squared differences instead of correlation. This decision rule performed slightly worse in general, but the results were qualitatively very similar to the results with the decision rule described in this section.

Results

We begin by examining the behavior of the two correlation-based models, and then compare this to the performance of our human observers.

Model

Decision rule 1: Autocorrelation

Figure 7 shows the results for the model with Decision rule 1 (autocorrelation). The boxed panel summarizes the results by plotting maximum performance over all amplitudes against frequency. In this and all further graphs, the error bars show 95% confidence intervals, the red curves show data for sine waves, and the blue curves show data for square waves.

When the maximum performance over all amplitudes is plotted against frequency, there is very little difference between the curves for the two different waveforms (boxed panel (K) in Figure 7). However, when we examine how performance depends on disparity amplitude (Figures 7A–7J), a key difference emerges between the two waveforms. At the smallest amplitude tested, performance is near chance but rapidly rises to its peak value. For sine-wave disparity gratings, performance then declines again as the disparity amplitude increases further. For the square-wave gratings, in contrast, performance remains at its peak value as the amplitude increases.

This is readily explicable. The model is built from local correlation detectors tuned to constant disparity. They respond well if disparity remains roughly constant over their window, and they do not respond well when there are steep disparity gradients. For sine-wave gratings, increasing

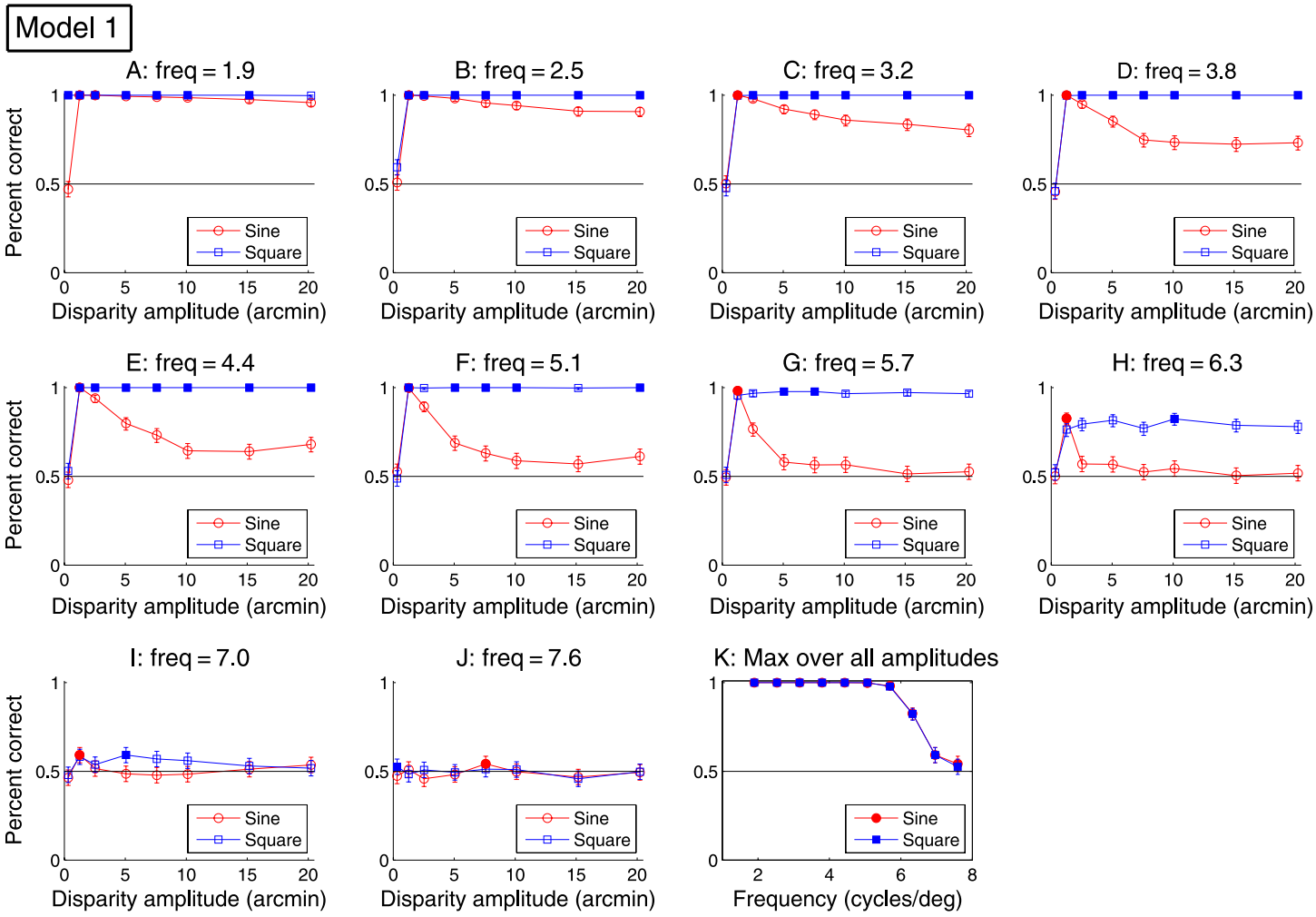

Figure 7. Performance as a function of amplitude and frequency for the model with the decision rule based on autocorrelation and a window with 2 * σ = 6 arcmin. The boxed plot (K) shows the maximum performance over all amplitudes for each frequency.

the disparity amplitude also increases the disparity gradient at every point (except the turning points), reducing performance. This reduction of performance with increasing disparity gradient was also found by Banks et al. (2004) and Filippini and Banks (2009) and was predicted by Kanade and Okutomi (1994) for a slightly different cross-correlation model. For square-wave gratings, the disparity gradient is zero everywhere except at the discontinuities, and this remains true as the amplitude increases. Thus, performance remains high, as long as the amplitude does not go outside the range of disparities to which the model is sensitive. The reason why performance is low for the lowest amplitude is because this amplitude, 0.3 arcmin, is lower than the step in the range of correlation detectors, which is 0.6 arcmin. The closest detectors are therefore at 0 and 0.6 arcmin, which are equally far from 0.3 arcmin and they will therefore be close to equally strongly activated by this disparity. The autocorrelation-based rule only uses the detector with the strongest response at each *y*-position and the detector for 0.6 arcmin can only be the most strongly activated one when the entire window or very close to the entire window is seeing 0.3 arcmin. This can only happen for the square waves, and it is only for the lowest frequency that it happens for a large enough range of *y*-values to allow detection.

Given that the model is built to respond to locally constant disparity, it is perhaps surprising that at low amplitudes (above 0.6 arcmin) it performs as well with sine waves as for square waves. Figure 3 shows that the peak cross-correlator output reached for sine-wave gratings does fluctuate across the cycle, being—unsurprisingly—lower where the disparity gradient is higher. However, recall that our model estimates disparity from the correlation detector reporting the largest response. Thus, so long as the peak is above the background noise, the correct disparity will still be identified. In addition, the decision rule (here, based on the autocorrelation of the estimated disparity profile) can still correctly identify which interval contains the grating, even if the estimated disparity is not accurate everywhere. We have examined the behavior of this model with different window sizes. Quantitatively, as the window size increases, performance naturally starts dropping at lower frequencies. Banks et al. found that decreasing the window size improves performance up to a limit that depends on the level of blur (Banks et al., 2004; Filippini & Banks, 2009). For optical levels of blur, they found the limiting window size to be about 6 arcmin, the value used in Figure 7. Window size does not affect the qualitative behavior of the model. In particular, we continue to find that (1) maximum performance as a function of frequency remains the same for both sine- and square-wave gratings (see the boxed panel in Figure 7); and (2) performance declines as a function of amplitude for sine-wave gratings but remains at its peak value for square-wave gratings (see Figure 7).

Decision rule 2: Template matching

It is important to be clear whether these features of the model performance reflect the low-level, correlation-based encoding of disparity, or whether they are specific to the particular decision rule chosen. In this section, we therefore present results from a more elaborate decision rule. This rule is based on matching the output of the correlation detector to any one stimulus, to a set of stored template responses to grating patterns. The template set includes responses to both sine- and square-wave gratings, and the decision rule uses whichever matches.

The results for this decision rule are shown in Figure 8, in the same format as in the previous section. The modeling results with the decision rule based on template matching are qualitatively very similar to the results with the autocorrelation-based decision rule. The main differences seem to be that, for a given window size, performance starts dropping at slightly lower frequencies and that for the lowest frequencies performance for the sine-waves remains high up to the highest amplitude tested. The reason for the higher performance for high amplitude sine waves may be that the template matching rule requires accurate disparity detection at a smaller percentage of *y*-positions to identify a grating; high correlation in small regions close to the peaks of the sine waves (see Figure 13) may be enough since the relevant template has the same pattern. The drop in performance for the lowest amplitude happens only to a lesser degree for the template matching rule than for the autocorrelation-based rule. This is because the template matching rule uses the outputs from all the correlation detectors and not just the one that has the strongest response at each *y*-position. However, critically, both decision rules show the same key features highlighted at the end of the previous section. In particular, as disparity amplitude increases, performance remains high for the square-wave gratings and declines for the sine wave. The alternative template matching approach mentioned in the Methods section also showed this behavior (results not shown). Thus, this key behavior is not dependent on any particular decision rule. As explained in the previous section, we attribute it to the properties of the initial disparity encoding performed by correlation detectors tuned to uniform disparity.

Psychophysics

We now examine the performance of human subjects, in order to compare it with the predictions of the model. Figure 9 shows the data for four subjects. The boxed panels summarize the results by plotting the maximum performance over all amplitudes for each frequency.

In striking agreement with the predictions of the correlation-based model of Banks et al., we find that the best performance reached at a given frequency is the same for both waveforms. The boxed panels (FLRX) in Figure 9

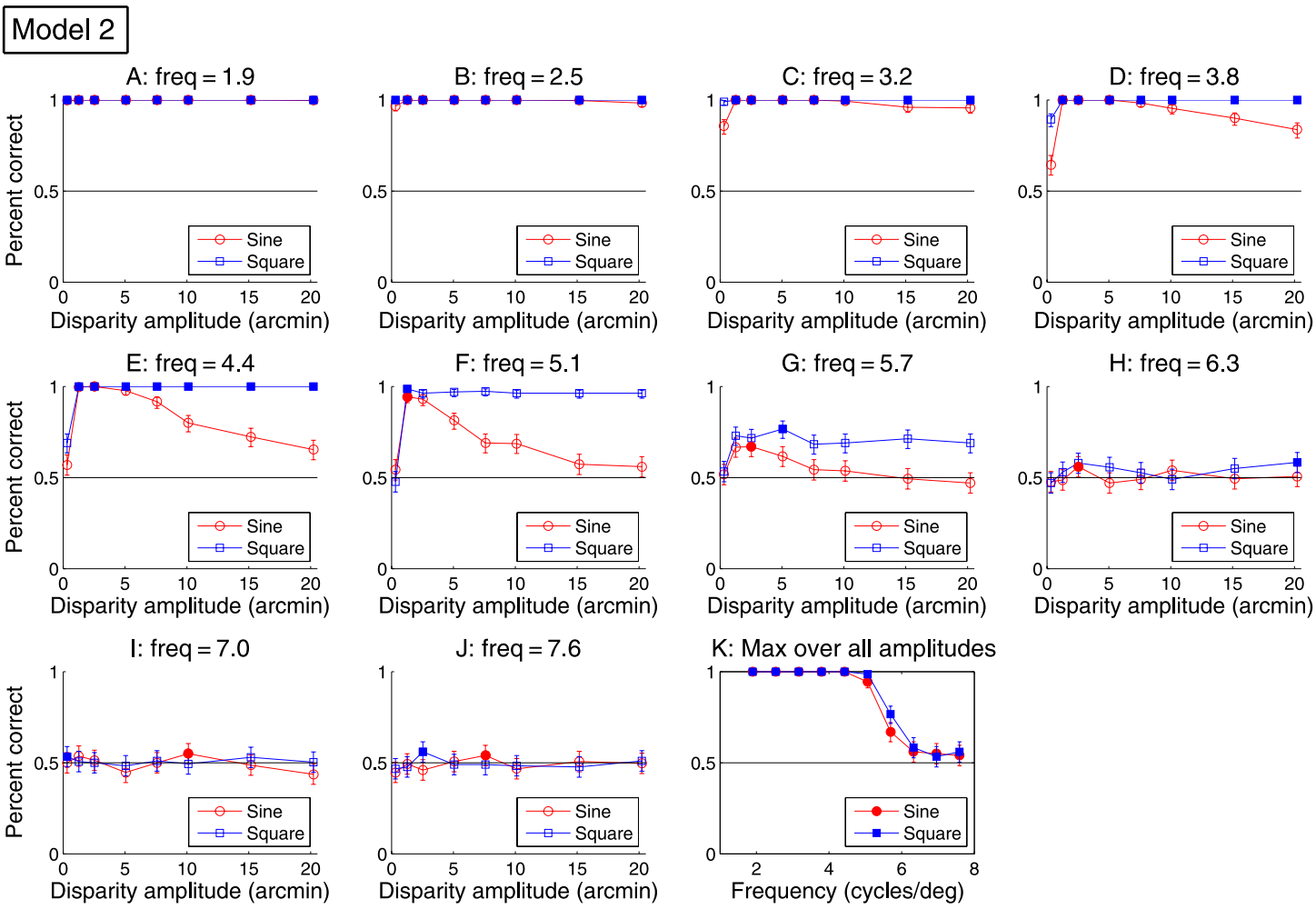

Figure 8. Performance as a function of amplitude and frequency for the model with the decision rule based on template matching and a window with 2σ = 6 arcmin. The boxed plot (H) shows the maximum performance over all amplitudes for each frequency.

show this best performance obtained at any disparity amplitude, plotted as a function of the frequency. None of our 4 subjects shows any significant difference in performance between sine- and square-wave gratings.

However, when we examine the graphs showing performance against disparity amplitude for individual frequencies, we find a significant departure from the model predictions. After initially rising to a peak value, performance then declines as disparity amplitude increases further. The model shows this decline only for sine-wave, not for square-wave gratings. However, for humans, the rate of decline is extremely similar for both sine-wave and square-wave gratings. Where significant differences do exist (e.g., subject PFA, discussed in more detail below), performance is better for the sine grating, not the square wave as predicted by the model.

Ultimately, of course, the performance of any realistic system must decline, as the disparity of the stimulus moves beyond the range to which its detectors respond. This effect was not included within our model (previous section), which contained an equal number of detectors for all disparities used. However, we do not believe that this omission can account for the difference between model and human performance we observe. We could force the model's performance down for large-amplitude square-wave gratings by reducing the range of disparity detectors. However, the resulting reduction in performance would not be specific to square-wave gratings but also affect sine gratings. It thus could not reconcile the model with human performance. It would also be unrealistic, because the disparity amplitudes used here are very small, well below Dmax (Glennerster, 1998; Read & Eagle, 2000) and perfectly detectable in other contexts. This is clear from our own data. Disparity amplitudes that our subjects find easy at low frequencies become impossible at higher frequencies. For example, at a frequency of 1.9 cpd, subject ISP performs at virtually 100% out to amplitudes as large as 10 arcmin, the largest examined. Yet at a frequency of 5.1 cpd, he is at chance for this amplitude, for both sine- and square-wave gratings. This cannot be because he lacks neuronal mechanisms capable of encoding disparities of 10 arcmin,

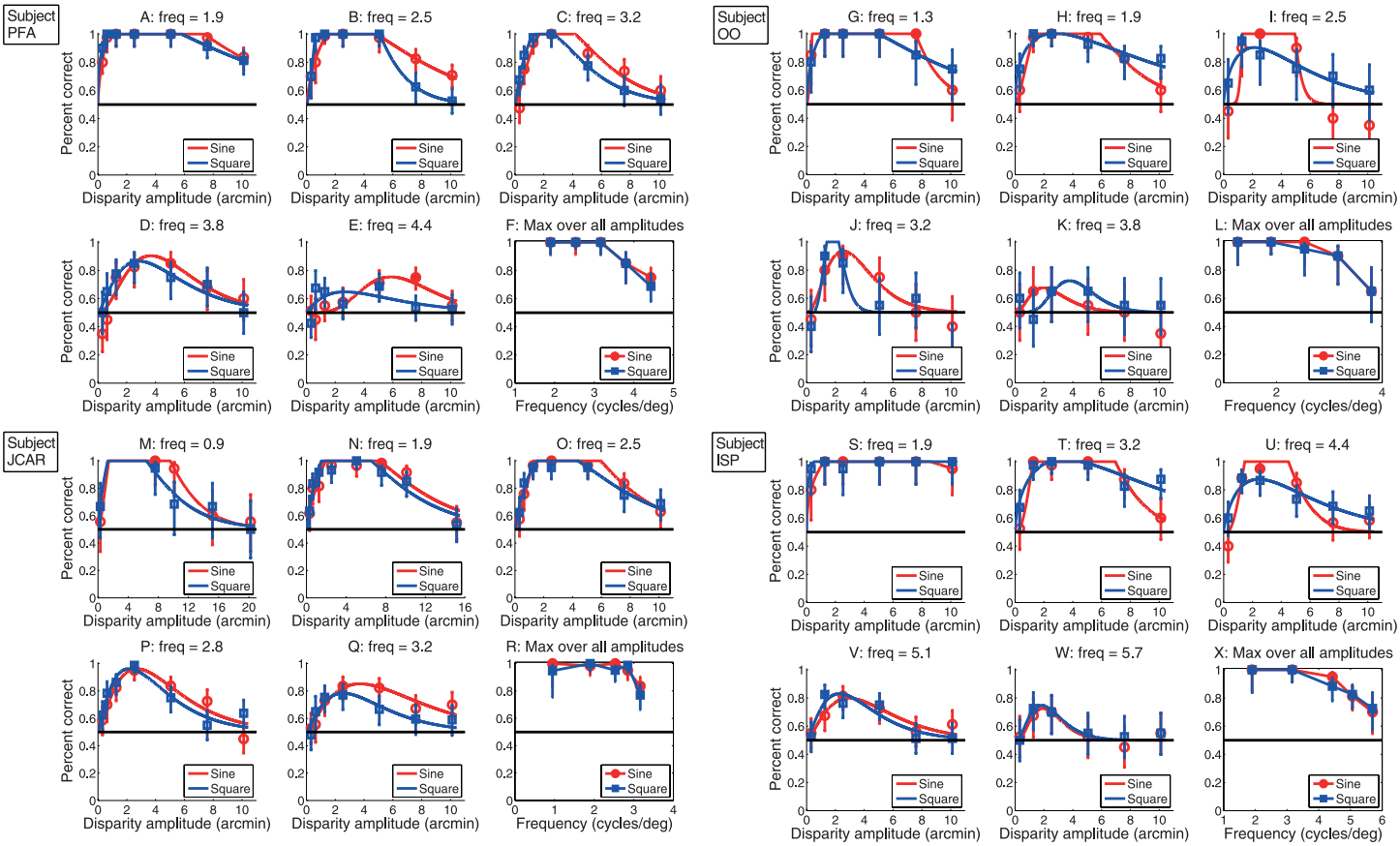

Figure 9. Performance as a function of amplitude and frequency for each subject. The squares and circles are data points and the lines are fits. The boxed plots (F, L, R, X) show, for each frequency, the maximum performance obtained at any amplitude.

since he perceived 10 arcmin perfectly at the lower frequency. Equally, it cannot be because 5.1 cpd is too high a frequency compared to the window size of his correlation detectors, because he reaches 80% correct for both grating profiles when the amplitude is smaller, 2–3 arcmin. His poor performance can only be due to the particular combination of frequency and amplitude.

Frequency analysis

This suggests that the correlation-based model may fail to capture some aspects of human depth perception. We now examine another influential approach to human perception, the Fourier or frequency-based analysis pioneered in the luminance domain by Campbell and Robson (1968), and later applied to disparity (Cobo-Lewis & Yeh, 1994; Grove & Regan, 2002; Schumer & Ganz, 1979; Tyler, 1975b).

In Fourier analysis, a square-wave grating can be decomposed into a sum of sine-wave gratings: a sine wave of the same period as the square wave but with $4/\pi$ its peak-to-trough range, plus successive lower amplitude sine waves. As the grating period decreases to the limit of detectable frequencies, a point is reached where the fundamental frequency is still above threshold, but the third harmonic is already below threshold. Sine- and square-wave gratings thus become indistinguishable. Most of our data fall within this domain, since for most subjects the highest frequency tested was just at the threshold of discriminability, whereas the lowest frequency tested was more than one-third of this value. This means that even at the lowest frequency tested, the third harmonic distinguishing the square-wave from the sine-wave grating would be nearly undetectable if presented alone. Thus if the linear theory is correct, if we plot performance as a function of the amplitude of the fundamental, instead of the whole-waveform amplitude used so far in this paper, performance should become the same for square-wave and sine-wave gratings.

This is examined in Figure 10. This figure shows the same data as Figure 9 but now plotted as a function of the amplitude of the fundamental. The sine (red) data are thus unchanged, while the square (blue) data and fits are shifted to the right by a factor of $4/\pi$. To assess whether this manipulation brings performance for the two waveforms closer together, we used the curves fitted to each set of data. For each frequency, we computed the integral of the absolute difference between the curves for the sine waves and the square waves, first for the original data and then for the adjusted data. If this integral was smaller for

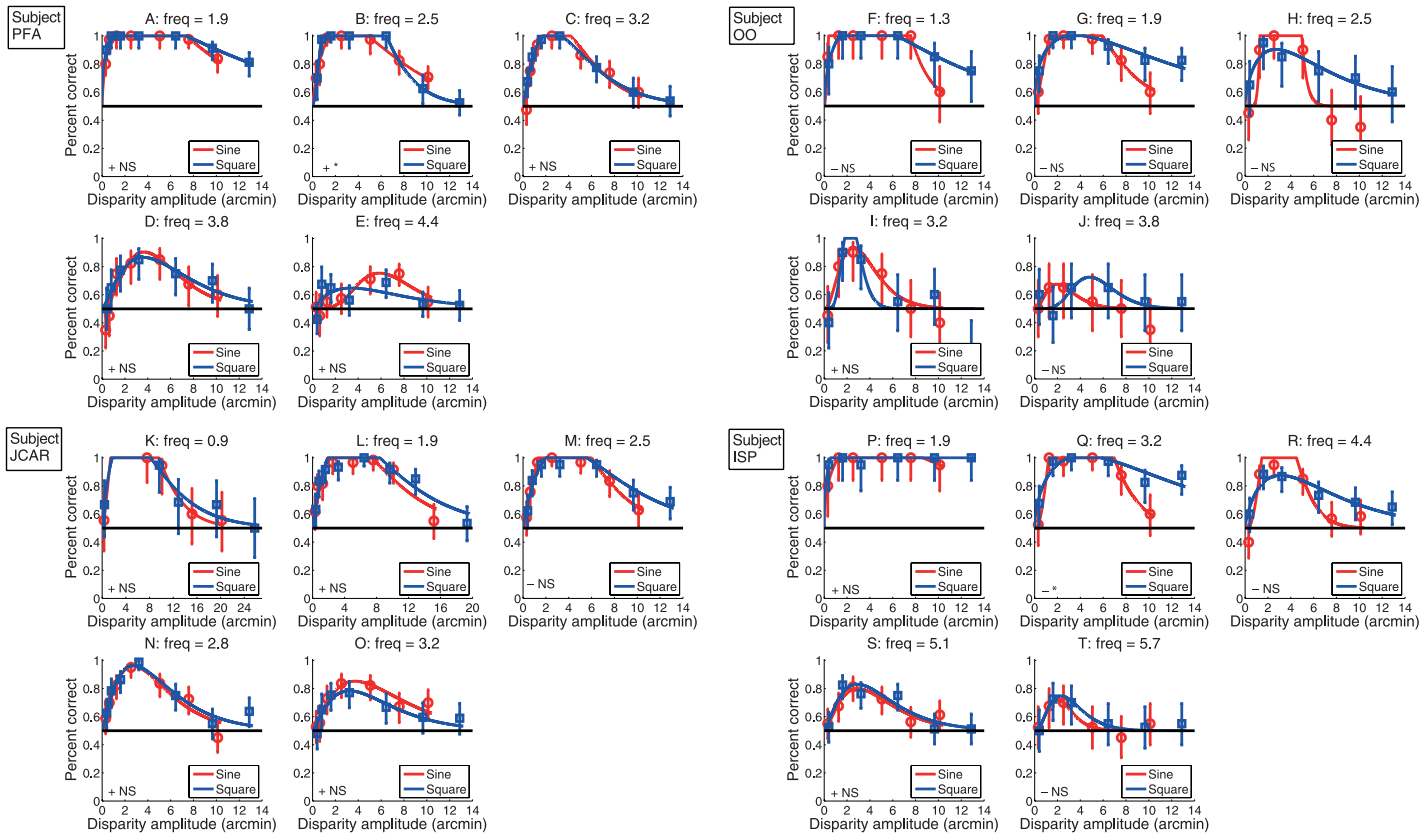

Figure 10. Performance as a function of amplitude and frequency for each subject, with the square-wave data plotted against the amplitude of the fundamental frequency component of the waves. The symbols at the bottom left of each panel show whether this has improved (+) or worsened (−) agreement between the sine- and square-wave results, and whether this is significant at the 5% level (*) or not (NS).

the adjusted data, this indicated that the shift to fundamental amplitude had brought the results closer together. This is indicated with a + symbol at the bottom left of the panels in Figure 10; a − symbol indicates that the shift to fundamental amplitude brought the fits further apart. Bootstrap resampling was used to estimate the significance of any change. The asterisks in Figure 10 indicate $p < 0.05$ (two-tailed test), while NS indicate that the adjustment had no significant effect either way.

For subject PFA, who performed the most repetitions per condition, plotting performance as a function of fundamental amplitude brings the curves closer together at every frequency. Plotted as a function of peak-to-trough amplitude (Figure 9), PFA often performed slightly better for the sine waves. When the data are adjusted so performance is a function of the fundamental amplitude (Figure 10), this effect is almost totally abolished, and the two sets of data overlap almost perfectly. However, this improvement was significant for only frequency, 2.5 cpd. For the other subjects, there is little evidence of any systematic effect one way or the other. Thus, our results provide little support for the linear Fourier analysis of disparity. Subjects perform very similarly for high-frequency sine-wave and square-wave gratings, but their performance does not seem to be set by the amplitude of the fundamental.

Disparity gradient limit

Several previous studies have suggested that stereopsis may be limited by the disparity gradient, rather than disparity per se (Banks et al., 2004; Burt & Julesz, 1980; Filippini & Banks, 2009; Kanade & Okutomi, 1994; McKee & Verghese, 2002; Tyler, 1975a). To examine this, in Figure 11 we plot the performance of all subjects on sine-wave gratings of all frequencies plotted against the amplitude of the gratings (ACEG) as well as the maximum disparity gradient in the gratings (the product of frequency and amplitude, BDFH). In order to test whether plotting against disparity gradient brings the curves closer together, we used the fits after extending them to end at the same point and cutting them to only include the portion of the curve after the peak. The standard deviation of the set of *y*-positions that the different curves passed through was computed at each *x*-position and the mean of this standard deviation over all *x*-positions was used as a measure of how closely the curves were superimposed. Bootstrap resampling was used to estimate the significance

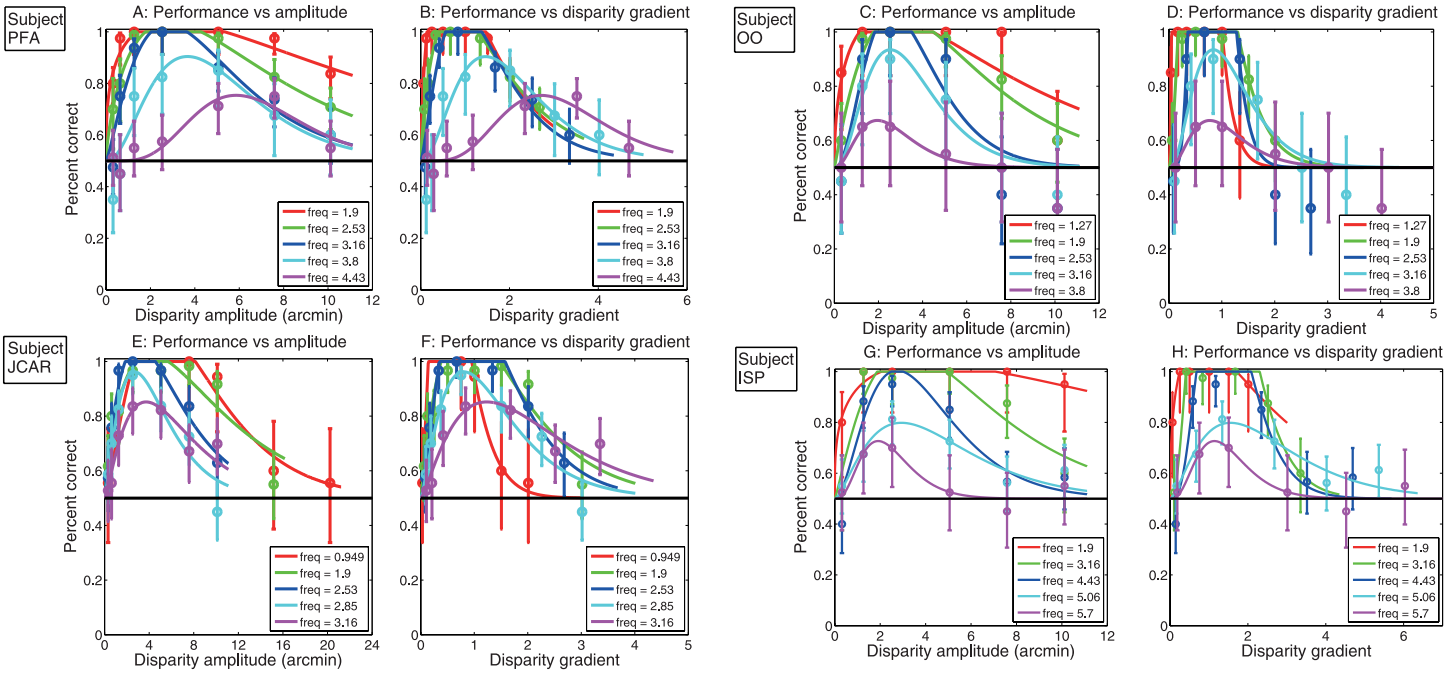

Figure 11. Performance plotted against amplitude (A, C, E, G) and maximum disparity gradient (B, D, F, H) for sine waves of all frequencies, for each of the 4 subjects.

of any difference between the two ways of plotting the data. The curves were found to be significantly more superimposed ($p < 0.05$) when plotting against disparity gradient for three out of four subjects (PFA, OO, and ISP). For the fourth subject, no significant difference either way was found. Thus, our data are consistent with the idea that performance at high amplitudes is limited by the highest perceivable disparity gradient.

Figure 12 shows the same plots for the model with the autocorrelation decision rule (similar results were obtained for the model with the template matching decision rule, not shown). Again, performance on sine waves of different frequencies is plotted against either the amplitude (A) or the maximum disparity gradient (B) of the grating. To test whether the model results superimposed better when plotted against disparity gradient, we

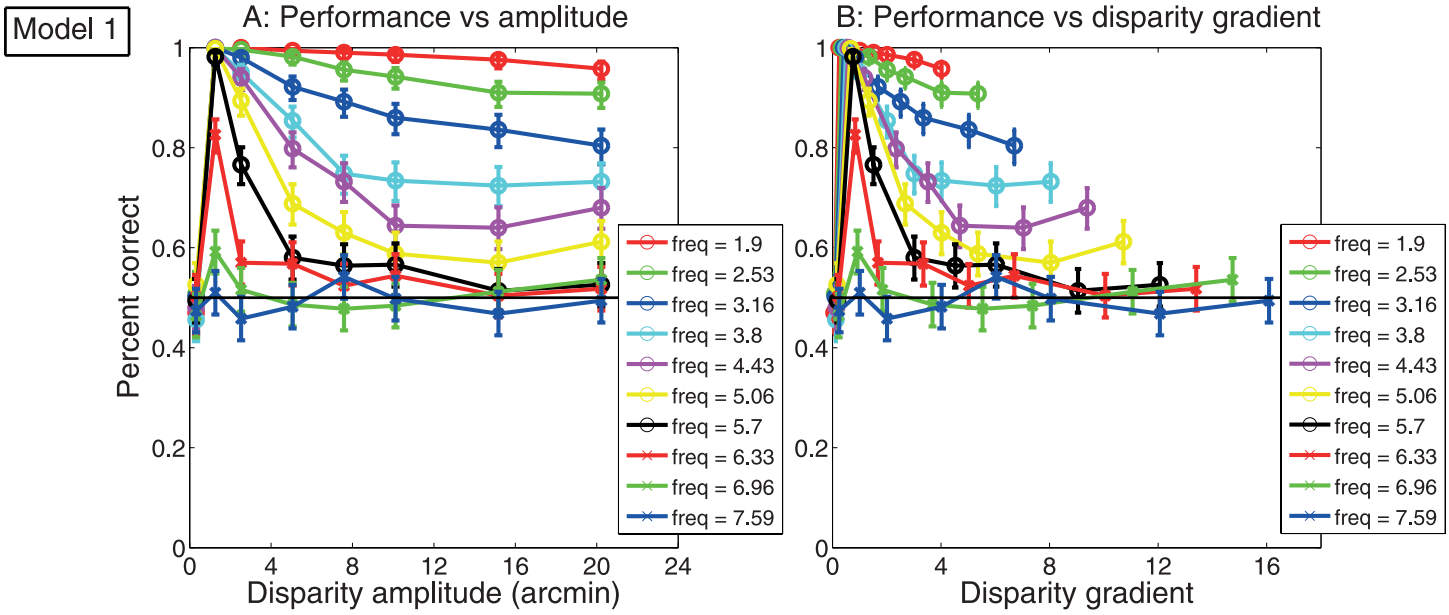

Figure 12. Performance plotted against (left) amplitude and (right) maximum disparity gradient for sine waves of all frequencies for the model with the decision rule based on autocorrelation.

used the parts of the curves from the peak to the last data point for the lowest frequency. As for the human data, the standard deviation of the set of y-positions that the different curves passed through was computed at each x-position and the mean of this standard deviation over all x-positions was used as a measure of how closely the curves were superimposed. Bootstrap resampling was used to estimate the significance of any difference between the two ways of plotting the data. No significant difference was found for the results with either of the decision rules. Thus, for the model results, the curves do not superimpose any better when the data are plotted against disparity gradient. Rather, the performance of the model depends separately on frequency and amplitude, and not simply on disparity gradient (amplitude × frequency). This is not surprising given that the model has no mechanisms that specifically detect disparity gradient. The observed dependence of frequency and amplitude may be because the correlation output from the first stage of the model has the highest correlation in the regions close to the flat parts of the sine wave (see Figure 13). Thus, performance may be limited by the size of the regions that are flat enough to generate high correlation, rather than by the maximum disparity gradient in the stimulus.

Discussion

In recent years, many models of human stereopsis have proposed that the initial encoding of disparity occurs in the primary visual cortex, V1, by disparity-selective neurons whose major properties are captured by the stereo energy model (Cumming & DeAngelis, 2001; Ohzawa et al., 1990; Qian, 1994; Qian & Zhu, 1997; Read, 2005). The neurophysiological evidence suggests that V1 neurons respond optimally to disparity that is constant across their receptive field (Nienborg et al., 2004). In higher brain areas, neurons that respond best to particular patterns of varying disparity are found (Janssen, Vogels, & Orban, 1999; Nguyenkim & DeAngelis, 2003; Sakata et al., 1999; Sugihara, Murakami, Shenoy, Andersen, & Komatsu, 2002). However, current models propose that these higher level neurons are built by combining the outputs of uniform-disparity V1 neurons (Bredfeldt & Cumming, 2006; Bredfeldt, Read, & Cumming, 2009; Thomas, Cumming, & Parker, 2002). Thus, Banks et al. (2004) and Filippini and Banks (2009) have argued that the initial piecewise-frontoparallel encoding of disparity imposes a fundamental limit on stereo resolution. In this

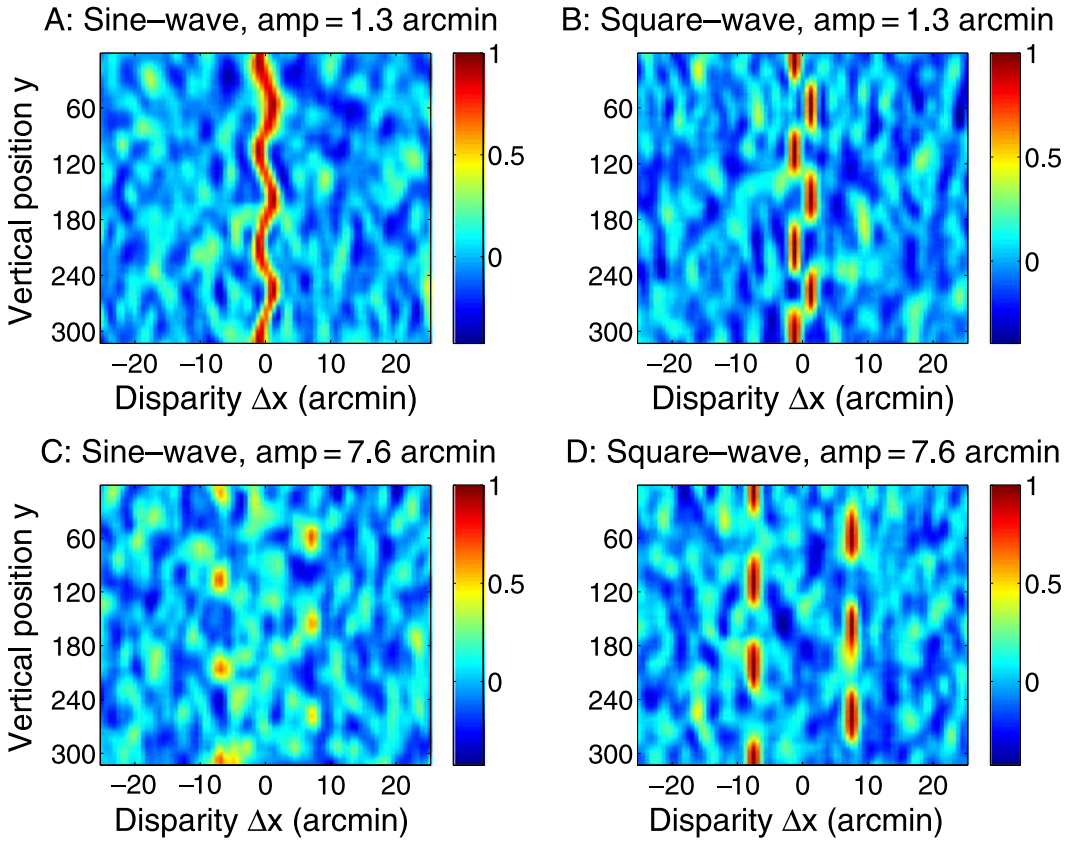

Figure 13. Examples of output from the cross-correlator for square waves and sine waves at low and high amplitudes. The quality of the correlation image remains high for the high-amplitude square wave but drops for the sine wave, with high correlation only near the peaks. These results are for a frequency of 1.9 cpd and a Gaussian window with 2σ = 6 arcmin.

view, the high-frequency limit for perceiving disparity gratings is imposed right down in V1, by the receptive field size of disparity-selective neurons.

This piecewise-frontoparallel theory of disparity encoding is quite different from the Fourier or frequency-based analysis pioneered in the luminance domain by Campbell and Robson (1968), and later extended to disparity (Cobo-Lewis & Yeh, 1994; Grove & Regan, 2002; Schumer & Ganz, 1979; Tyler, 1975b). In that picture, the quantity of interest (disparity or luminance) is initially encoded by a set of frequency channels. The basic "unit" in which the quantity is represented is the sine wave (or a local version of it, like a Gabor), not a constant-value patch as in the piecewise-frontoparallel theory. In linear Fourier theory, square-wave and sine-wave gratings with the same fundamental amplitudes should become equally detectable at high frequencies, once the third harmonic of the square wave has passed above the frequency threshold. In contrast, if the piecewise-frontoparallel theory is correct, it should be easier to perceive a square-wave disparity grating than a sine-wave grating, because the square-wave grating consists of locally frontoparallel regions of disparity, and so should drive V1 neurons more strongly. It is of course possible that there are frequency channels that are built by combining the outputs of uniform-disparity V1 neurons. If so, the initial frontoparallel representation of disparity will still set limits on performance, even though the later processing needed to construct the frequency channels may limit performance further.

We tested the behavior of the piecewise-frontoparallel model by running simulations. We verified that the model does indeed find it easier to detect square-wave gratings, which are piecewise-frontoparallel, than sine-wave gratings, which everywhere have a non-zero disparity gradient. In particular, for square-wave gratings the model was able to perform well out to high amplitudes (limited only by the range of preferred disparities included within the model neuronal population), whereas for sine-wave gratings, performance declined at high amplitudes. This behavior is what we expected given the structure of the model, cf. Figure 2. We confirmed that it does not depend critically on the particular details of the model implementation; for example, we obtained the same behavior with two quite different decision rules. Rather, it reflects the initial stage of local cross-correlation. Figure 13 shows the output of this stage for both sine- and square-wave gratings, at low and high amplitudes, for a relatively low frequency, 1.9 cpd. At low amplitudes, the piecewise-frontoparallel model can successfully track the disparity of both grating profiles (Figures 13A and 13B). In contrast, at high amplitudes (Figures 13C and 13D), only the very peaks of the sine-wave grating remain visible (where the disparity gradient is briefly zero), while the square-wave grating remains just as clear as at low amplitude. Thus, our simulations confirm our intuitions about the behavior of models based on piecewise-frontoparallel disparity encoding.

However, to our surprise, our psychophysical results were quite different. There was no evidence that performance was ever significantly better for square-wave than sine-wave gratings. Like the model, the maximum performance possible at a given frequency was indistinguishable for the two waveforms. However, after initially rising to a peak, human performance declines as a function of amplitude for both sine- and square waves. This is quite different from the behavior of the piecewise-frontoparallel model, where performance declines only for sine-wave gratings and remains high for the square-wave gratings out to large amplitudes. The decline in the performance of human observers occurs for disparity amplitudes that are clearly detectable at lower frequencies. This shows that the poor performance is caused by the frequency of disparity alternation, not the intrinsic detectability of the disparities present in the stimulus.

Thus like Banks et al., we find that the piecewise-frontoparallel model based on local cross-correlation does an excellent job of capturing human performance on sine-wave gratings. However, the discrepancy with square-wave gratings indicates that the model is incomplete as a model of human stereo vision.

A limitation of this model is that it only includes the initial encoding of disparity in V1, not the higher level neurons that respond to varying disparity (Janssen et al., 1999; Nguyenkim & DeAngelis, 2003; Sakata et al., 1999; Sugihara et al., 2002). Prominent among these are the class of disparity-edge detectors in V2 (Bredfeldt & Cumming, 2006; von der Heydt, Zhou, & Friedman, 2000). There is considerable psychophysical evidence suggesting that "edges" or discontinuities in disparity are particularly salient for stereo vision (Andrews, Glennerster, & Parker, 2001; Gillam, Blackburn, & Brooks, 2007; Serrano-Pedraza, Phillipson, & Read, 2010), presumably reflecting the activation of these neuronal disparity-edge detectors. Square-wave gratings contain sharp disparity edges, whereas sine-wave gratings do not. This is probably why disparity thresholds are consistently better for square-wave than for sine-wave gratings at low frequencies (below 2 cpd; Serrano-Pedraza & Read, 2009). Thus, the model's failure to include known mechanisms of edge detection should, if anything, bring square-wave performance *closer* to sine wave. This deficiency, therefore, also cannot explain the discrepancy between model and human results.

Our psychophysical results did not provide compelling evidence that disparity is encoded within a set of independent frequency channels. A linear frequency analysis would suggest that, at high frequencies, performance on the two types of grating should become more similar when the amplitude of the grating was expressed as the amplitude of the fundamental, rather than as half the

peak-to-trough distance. This was the case for only one of our four subjects. In contrast, our results were more clearly consistent with previous work indicating that disparity gradient is critical to perception (Burt & Julesz, 1980; McKee & Verghese, 2002; Tyler, 1975a).

If neither the piecewise-frontoparallel model, nor a linear frequency analysis, seems capable of fully explaining our results, how should we proceed in order to achieve an accurate model of human stereo depth perception? It may be necessary to invoke further processing happening after the cross-correlation stage. Alternatively, it may be possible to modify the cross-correlation model so as to reconcile it with our results. For example, our current model contains equal numbers of sensors with different disparity tuning, whereas V1 neurons are tuned predominantly to near-zero disparities (Prince, Cumming, & Parker, 2002). It also assumes that the "window" size used for cross-correlation is constant, whereas V1 neurons tuned to larger disparities tend to have larger receptive fields (Prince et al., 2002), reflecting the size/disparity correlation deduced from psychophysical results (McKee & Verghese, 2002; Smallman & MacLeod, 1994; Tsirlin, Allison, & Wilcox, 2008; Tyler, 1975a). Incorporating such sophistications into our model may help it account for human performance with sine-wave and square-wave gratings.

Conclusion

Piecewise-frontoparallel local cross-correlation successfully captures many aspects of human stereo vision. However, at least as currently implemented, it predicts that humans should be better at detecting square-wave disparity gratings than sine-wave gratings, when the frequency and amplitude of the gratings are high. In fact, humans perform almost equally well on both grating profiles. In particular, human performance declines as a function of amplitude for both square- and sine-wave gratings, whereas the model predicts a region where performance is independent of amplitude for square-wave gratings. We conclude that the model needs to be refined in order to capture this aspect of human depth perception.

Acknowledgments

The authors would like to thank Sergei Gepshtein and Martin Banks for helpful discussions. This research was supported by a Newcastle University Institute of Neuroscience PhD studentship to FA and by Royal Society University Research Fellowship UF041260 and MRC New Investigator Award 80154 to JCAR.

Commercial relationships: none.
Corresponding author: Fredrik Allenmark.
Email: fredrik.allenmark@ncl.ac.uk.
Address: Henry Wellcome Building, Newcastle upon Tyne, NE2 4HH, UK.

References

Andrews, T. J., Glennerster, A., & Parker, A. J. (2001). Stereoacuity thresholds in the presence of a reference surface. *Vision Research, 41,* 3051–3061.

Banks, M. S., Gepshtein, S., & Landy, M. S. (2004). Why is spatial stereoresolution so low? *Journal of Neuroscience, 24,* 2077–2089.

Bradshaw, M. F., & Rogers, B. J. (1999). Sensitivity to horizontal and vertical corrugations defined by binocular disparity. *Vision Research, 39,* 3049–3056.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436.

Bredfeldt, C. E., & Cumming, B. G. (2006). A simple account of cyclopean edge responses in macaque V2. *Journal of Neuroscience, 26,* 7581–7596.

Bredfeldt, C. E., Read, J. C., & Cumming, B. G. (2009). A quantitative explanation of responses to disparity-defined edges in macaque V2. *Journal of Neurophysiology, 101,* 701–713.

Burt, P., & Julesz, B. (1980). A disparity gradient limit for binocular fusion. *Science, 208,* 615–617.

Campbell, F. W., & Green, D. G. (1965). Optical and retinal factors affecting visual resolution. *The Journal of Physiology, 181,* 576–593.

Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *The Journal of Physiology, 197,* 551–566.

Cobo-Lewis, A. B., & Yeh, Y. Y. (1994). Selectivity of cyclopean masking for the spatial frequency of binocular disparity modulation. *Vision Research, 34,* 607–620.

Cumming, B. G., & DeAngelis, G. C. (2001). The physiology of stereopsis. *Annual Review of Neuroscience, 24,* 203–238.

Filippini, H. R., & Banks, M. S. (2009). Limits of stereopsis explained by local cross-correlation. *Journal of Vision, 9*(1):8, 1–18, http://www.journalofvision.org/content/9/1/8, doi:10.1167/9.1.8. [PubMed] [Article]

Geisler, W. S., & Davila, K. D. (1985). Ideal discriminators in spatial vision: Two-point stimuli. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 2,* 1483–1497.

Gillam, B., Blackburn, S., & Brooks, K. (2007). Hinge versus twist: The effects of "reference surfaces" and

discontinuities on stereoscopic slant perception. *Perception, 36,* 596–616.

Glennerster, A. (1998). Dmax for stereopsis and motion in random dot displays. *Vision Research, 38,* 925–935.

Grove, P. M., & Regan, D. (2002). Spatial frequency discrimination in cyclopean vision. *Vision Research, 42,* 1837–1846.

Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis*. New York: Oxford University Press.

Janssen, P., Vogels, R., & Orban, G. A. (1999). Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes. *Proceedings of the National Academy of Sciences of the United States of America, 96,* 8217–8222.

Kanade, T., & Okutomi, M. (1994). A stereo matching algorithm with an adaptive window—Theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 16,* 920–932.

McKee, S. P., & Verghese, P. (2002). Stereo transparency and the disparity gradient limit. *Vision Research, 42,* 1963–1977.

Nguyenkim, J. D., & DeAngelis, G. C. (2003). Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons. *Journal of Neuroscience, 23,* 7117–7128.

Nienborg, H., Bridge, H., Parker, A. J., & Cumming, B. G. (2004). Receptive field size in V1 neurons limits acuity for perceiving disparity modulation. *Journal of Neuroscience, 24,* 2065–2076.

Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science, 249,* 1037–1041.

Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1997). Encoding of binocular disparity by complex cells in the cat's visual cortex. *Journal of Neurophysiology, 77,* 2879–2909.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.

Prince, S. J., Cumming, B. G., & Parker, A. J. (2002). Range and mechanism of encoding of horizontal disparity in macaque V1. *Journal of Neurophysiology, 87,* 209–221.

Qian, N. (1994). Computing stereo disparity and motion with known binocular cell properties. *Neural Computation, 6,* 390–404.

Qian, N., & Zhu, Y. D. (1997). Physiological computation of binocular disparity. *Vision Research, 37,* 1811–1827.

Read, J. (2005). Early computational processing in binocular vision and depth perception. *Progress in Biophysics & Molecular Biology, 87,* 77–108.

Read, J. C., & Eagle, R. A. (2000). Reversed stereo depth and motion direction with anti-correlated stimuli. *Vision Research, 40,* 3345–3358.

Roe, A. W., Parker, A. J., Born, R. T., & DeAngelis, G. C. (2007). Disparity channels in early vision. *Journal of Neuroscience, 27,* 11820–11831.

Rossi, E. A., & Roorda, A. (2009). The relationship between visual resolution and cone spacing in the human fovea. *Nature Neuroscience, 13,* 156–157.

Sakata, H., Taira, M., Kusunoki, M., Murata, A., Tsutsui, K., Tanaka, Y., et al. (1999). Neural representation of three-dimensional features of manipulation objects with stereopsis. *Experimental Brain Research, 128,* 160–169.

Schumer, R., & Ganz, L. (1979). Independent stereoscopic channels for different extents of spatial pooling. *Vision Research, 19,* 1303–1314.

Serrano-Pedraza, I., Phillipson, G. P., & Read, J. C. A. (2010). A specialization for vertical disparity discontinuities. *Journal of Vision, 10*(3):2, 21–25, http://www.journalofvision.org/content/10/3/2, doi:10.1167/10.3.2. [PubMed] [Article]

Serrano-Pedraza, I., & Read, J. C. A. (2009). Horizontal/vertical anisotropy in sensitivity to relative disparity depends on stimulus depth structure. *Perception, 38,* 155.

Smallman, H. S., & MacLeod, D. I. (1994). Size–disparity correlation in stereopsis at contrast threshold. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 11,* 2169–2183.

Sugihara, H., Murakami, I., Shenoy, K. V., Andersen, R. A., & Komatsu, H. (2002). Response of MSTd neurons to simulated 3D orientation of rotating planes. *Journal of Neurophysiology, 87,* 273–285.

Thomas, O. M., Cumming, B. G., & Parker, A. J. (2002). A specialization for relative disparity in V2. *Nature Neuroscience, 5,* 472–478.

Tsirlin, I., Allison, R. S., & Wilcox, L. M. (2008). Stereoscopic transparency: Constraints on the perception of multiple surfaces. *Journal of Vision, 8*(5):5, 1–10, http://www.journalofvision.org/content/8/5/5, doi:10.1167/8.5.5. [PubMed] [Article]

Tyler, C. W. (1974). Depth perception in disparity gratings. *Nature, 251,* 140–142.

Tyler, C. W. (1975a). Spatial organization of binocular disparity sensitivity. *Vision Research, 15,* 583–590.

Tyler, C. W. (1975b). Stereoscopic tilt and size after-effects. *Perception, 4,* 187–192.

von der Heydt, R., Zhou, H., & Friedman, H. S. (2000). Representation of stereoscopic edges in monkey visual cortex. *Vision Research, 40,* 1955–1967.