Combining 1-D components to extract pattern information: It is about more than component similarity

Christian Quaia Laboratory of Sensorimotor Research, National Eye Institute, NIH, DHHS, Bethesda, MD, USA

Lance M. Optican Laboratory of Sensorimotor Research, National Eye Institute, NIH, DHHS, Bethesda, MD, USA

Bruce G. Cumming Laboratory of Sensorimotor Research, National Eye Institute, NIH, DHHS, Bethesda, MD, USA

At least under some conditions, plaid stimuli are processed by combining information first extracted in orientation and scale-selective channels. The rules that govern this combination across channels are only partially understood. Although the available data suggests that only components having similar spatial frequency and contrast are combined, the extent to which this holds has not been firmly established. To address this question, we measured, in human subjects, the short-latency reflexive vergence eye movements induced by stereo plaids in which spatial frequency and contrast of the components are independently varied. We found that, although similarity in component spatial frequency and contrast matter, they interact in a nonseparable way. One way in which this relationship might arise is if the internal estimate of contrast is not a faithful representation of stimulus contrast but is instead spatial frequency–dependent (with higher spatial frequencies being boosted). We propose that such weighting might have been put in place by a mechanism that, in an effort of achieve contrast constancy and/or coding efficiency, regulates the gain of detectors in early visual cortex to equalize their long-term average response to natural images.

Introduction

It has long been established that visual information is initially processed in scale-selective channels (N. Graham, 1989). As early as the retina, neurons are only sensitive to signals that fall within a limited spatial frequency (SF) bandwidth. In striate cortex, neurons also become selective for stimulus orientation, further partitioning the input into orientation channels (R. De Valois & De Valois, 1988). This decomposition of the visual input, akin to a localized Fourier (or wavelet) decomposition, gives rise to an efficient coding scheme (Field, 1999). However, because information from separate channels often needs to be selectively combined, it also poses a challenge for subsequent processing.

Moving "plaids," obtained by summing two drifting one-dimensional (1-D) sinusoidal gratings having different orientations, are the stimuli most often used to study this process: With plaids, the speed and direction of motion of the pattern is extracted by appropriately combining information across different channels. A similar process is required to compute the depth of stereo plaids, which are obtained by summing two static 1-D sinusoidal gratings with different binocular disparity (Farell, 1998). In both cases, if the two 1-D components are sufficiently different in spatial frequency or contrast, a single coherent percept no longer results, and two separate gratings are perceived (transparency) as if information from the two channels is no longer combined (Adelson & Movshon, 1982; Farell, 1998).

We recently showed (Quaia, Sheliga, Optican, & Cumming, 2013) that the reflexive, short-latency, disparity vergence mechanism discovered by Busettini, Masson, and Miles (1996) responds to the 2-D pattern disparity of stereo plaids, sometimes even when the plaid is not perceptually coherent (e.g., with square-wave plaids). Importantly, our findings supported Farell's (1998) proposal that the 2-D pattern disparity signal is computed by combining 1-D disparities, first extracted within orientation-selective channels in area V1 (Hubel & Wiesel, 1968), according to the intersection-of-constraints (IOC) rule (Adelson & Movshon,

Citation: Quaia, C., Optican, L. M., & Cumming, B. G. (2017). Combining 1-D components to extract pattern information: It is about more than component similarity. *Journal of Vision*, *17*(3):21, 1–13, doi:10.1167/17.3.21.

doi: 10.1167/17.3.21 Received November 23, 2016; published March 29, 2017 ISSN 1534-7362

1982; Fennema & Thompson, 1979): Disparity-selective 2-D feature detectors are neither necessary nor sufficient. Therefore, these eye movements provide a powerful tool to explore the rules governing the combination of components over a wide range of stimuli.

Paralleling findings on the perceptual coherence of stereo plaids, we previously reported that when the two components have similar contrast they induce stronger eye movements than when their contrast is different. We argued that a mechanism of cross-orientation inhibition (DeAngelis, Robson, Ohzawa, & Freeman, 1992; Morrone, Burr, & Maffei, 1982), which would precede the stage at which the two 1-D disparity signals are combined to extract pattern disparity, could be responsible for this behavior. We also observed that a robust vergence response requires the two gratings to have similar SF as if only 1-D signals of similar SF are combined. The full-width-at-half-height (FWHH) bandwidth of this operation was approximately two octaves, which is within the range of SF bandwidth observed in monkeys in V1 (R. De Valois & De Valois, 1988) and in middle temporal (MT) cortex (Priebe, Lisberger, & Movshon, 2006; Wang & Movshon, 2016). We argued that these data are compatible with a second model stage, replicated at all scales, in which only oriented filters having the same spatial scale are combined.

Because this two-stage model predicts that the effects of frequency and contrast are separable, stimuli that differ in both contrast and spatial frequency provide a stronger test of our hypothesis. We measured the disparity vergence responses (DVRs) to such stimuli in three human subjects and found that separability does not hold: When the contrast of one component is increased, the pattern computation breaks down unless the SF of this component is lowered. This interaction between stimulus contrast and SF implies that our proposed model needs to be considerably revised. We show that our results can be reproduced if (a) the internal estimates of component contrast do not faithfully encode actual contrast but are instead biased in a SF-dependent manner (with high SFs being boosted relative to low SFs at least within the range of SFs tested in our experiment) and (b) the combination spans multiple channels. We suggest that this SF-dependent weighting might result from a mechanism that regulates the gain of each SF channel in an effort to achieve contrast constancy (i.e., to correctly estimate stimulus contrast irrespective of SF; Blakemore, Muncey, & Ridley, 1973; Bowker, 1983; Bryngdahl, 1966; Georgeson & Sullivan, 1975; Kulikowski, 1976; Stephens & Banks, 1985; Swanson, Georgeson, & Wilson, 1988; Watanabe, Mori, Nagata, & Hiwatashi, 1968) and/or to whiten the spectral representation of natural images (Atick & Redlich, 1992; Barlow, 2001; Brady & Field, 1995, 2000; Field, 1987; Field & Brady, 1997; Simoncelli & Olshausen, 2001) but fails to do so uniformly across the entire SF spectrum.

Whether the processing of moving plaids is based on similar computations remains to be seen. We have recently reported (Quaia, Optican, & Cumming, 2016) that, at least as far as the eye movement responses to unikinetic plaids are concerned, the mechanisms used for extracting pattern velocity from moving plaids might be quite different from those used for extracting pattern disparity from stereo plaids. Nevertheless, if the nonseparability reported here reflects processes occurring in V1, as we suggest, it should presumably affect all later computations based on these signals, thus including the extraction of pattern motion from moving plaids.

Methods

The methodology used to acquire the data presented here is identical to that used in our previously published study (Quaia et al., 2013). Accordingly, it will only be described briefly here.

We used the eye search coil technique (Collewijn, van der Mark, & Jansen, 1975) to record binocular horizontal and vertical eye position from three male subjects (one author). All subjects had normal or corrected-to-normal visual acuity and normal stereoacuity. All protocols were approved by the Institutional Review Board concerned with the use of human subjects. Our research was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki), and informed consent was obtained for experimentation with human subjects.

Experiments were carried out in a dark room, and the subject's head was stabilized using chin and head rests. In each trial, the subject fixated a central cross superimposed on a gray background (20.8 cd/m^2). After 800–1100 ms, a stimulus suddenly appeared and remained visible for 200 ms. Subsequently, the screen turned gray (20.8 cd/m^2), signaling the end of the trial. Stimuli were unidisparity plaids formed by summing a vertical sinusoidal grating having ±90° of phase disparity and an oblique (±45°) sinusoidal grating having zero disparity. The stimuli were static although we have no reason to believe that dynamic stimuli (i.e., stimuli in which the monocular phase of the gratings is changed randomly in each frame while keeping the binocular phase difference constant) would have resulted in different responses. All stimuli spanned 42° horizontally and 32° vertically and were presented dichoptically on two CRT monitors (part of a Wheatstone mirror stereoscope). They had a mean luminance of 20.8 cd/m^2; the contrast and spatial frequency of

each of the two gratings were manipulated independently.

All measures reported here are based on vergence velocity, which was obtained by differentiating eye position traces to obtain velocity signals for each eye and by subtracting the velocity from the two eyes. Trials with saccadic intrusions or poor fixation were discarded using an automatic procedure for the detection of outliers. Average vergence velocity traces, time-locked to stimulus onset, were then computed from the remaining trials. To quantify the strength of the response, we computed the mean vertical vergence velocity in a time window (75–120 ms for CQ, 80–120 ms for the other two subjects) that extended from the shortest onset latency of the vertical vergence response to twice the shortest onset latency of the horizontal vergence response (which determines the end of the open-loop period). We call this measure the pattern disparity response (PDR). All results and analyses reported are quite insensitive to 10- to 20-ms changes to window beginning and/or duration. Idiosyncratic components of the response related to fixation disengagement were removed by computing the difference between the vergence response to stimuli having opposite disparity (i.e., stimuli in which the images seen by the two eyes are swapped). Nonparametric bootstrap methods were used for all statistical analyses.

Two descriptive models were developed (see Results) and were fit to the data using a multiobjective genetic algorithm (Goldberg, 1989; van Soest & Casius, 2003) under the optimization software package modeFRONTIER (Esteco, Trieste, Italy). The algorithm was initialized using a Sobol quasirandom sequence; each generation had 400 individuals, and evolution spanned 50 generations. The model had nine parameters, and nine error measures were defined, which were then consolidated into two objectives to be minimized (see Results). Each error measure was computed as the mean of the deviation between model prediction and each data point, divided by the standard error of the mean (SEM) of the data point (i.e., the mean Z score) across the data points that contributed to the error measures. Next, we divided these measures into two groups: one with the six error measures capturing deviations for location or magnitude of a peak and the other with the error measures related to tuning bandwidth. We then defined two optimization objectives, O_{peak} and O_{bw}, as the maximum value within each group of error measures. The algorithm thus sought to minimize simultaneously the worse error within each group. The outcome of the optimization was then a set of designs (i.e., parameter sets) on the Pareto front (i.e., designs for which no other design could be found with a lower value for both objectives). Finally, we selected as the "best" design the one with the lowest value for the weighted sum of the two objectives with a weight of 0.7 for O_{peak} and a weight of 0.3 for O_{bw} (thus weighting more the peak measures).

Results

Unidisparity plaids (Quaia et al., 2013) are ideal stimuli to study pattern disparity computations. They are obtained by summing two gratings (typically sinusoidal or 1-D noise): one vertical with near or far disparity and one oblique (±45°) with zero disparity (i.e., in the plane of fixation). We previously showed that when a unidisparity plaid is suddenly presented to a human subject, a very short latency (<65 ms) vergence response is elicited. Initially, the response is purely horizontal with a sign that is determined by the disparity of the vertical grating (near disparities induce convergence, far disparities induce divergence). After 10–15 ms, a vertical vergence component of the response also emerges. The sign of this delayed component is a function of the sign of the disparity of the vertical grating and the orientation of the oblique grating in accordance with the IOC rule (Adelson & Movshon, 1982; Fennema & Thompson, 1979). This late component is the signature of the pattern disparity computation and the focus of our study. We quantify this component by computing the average vertical vergence velocity in a time window (see Methods for details); we call this measure the PDR. We found that when the two gratings have the same contrast, the PDR is maximal when the two gratings have similar SF and decreases as their difference increases (Quaia et al., 2013). The relationship between PDR and relative SF is well captured by a log-Gaussian curve with a FWHH bandwidth of approximately two octaves. However, we now report that this relationship varies as a function of the relative contrast of the two gratings.

In the experiments presented here, the SF of the vertical sinusoidal grating, which had a phase disparity of ±90°, was always 0.354 c/°, and its Michelson contrast was always 20%. The SF of the zero-disparity oblique sinusoidal grating was varied between 0.088 c/° and 1.414 c/° in half-octave steps; its Michelson contrast was varied between 7.07% and 56.6%, again in half-octave steps.

In Figure 1A, we show seven tuning curves, one for each contrast of the oblique grating (see legend) as a function of the SF of the oblique grating. The location of the peak (peak SF) of each tuning curve varies as a function of the contrast of the oblique grating: Decreasing (increasing) the contrast of the oblique grating results in an increase (decrease) of the peak SF. The relationship between contrast and peak SF is captured quite well by a linear fit on a log–log scale (Figure 1B). In all three subjects, the slope is

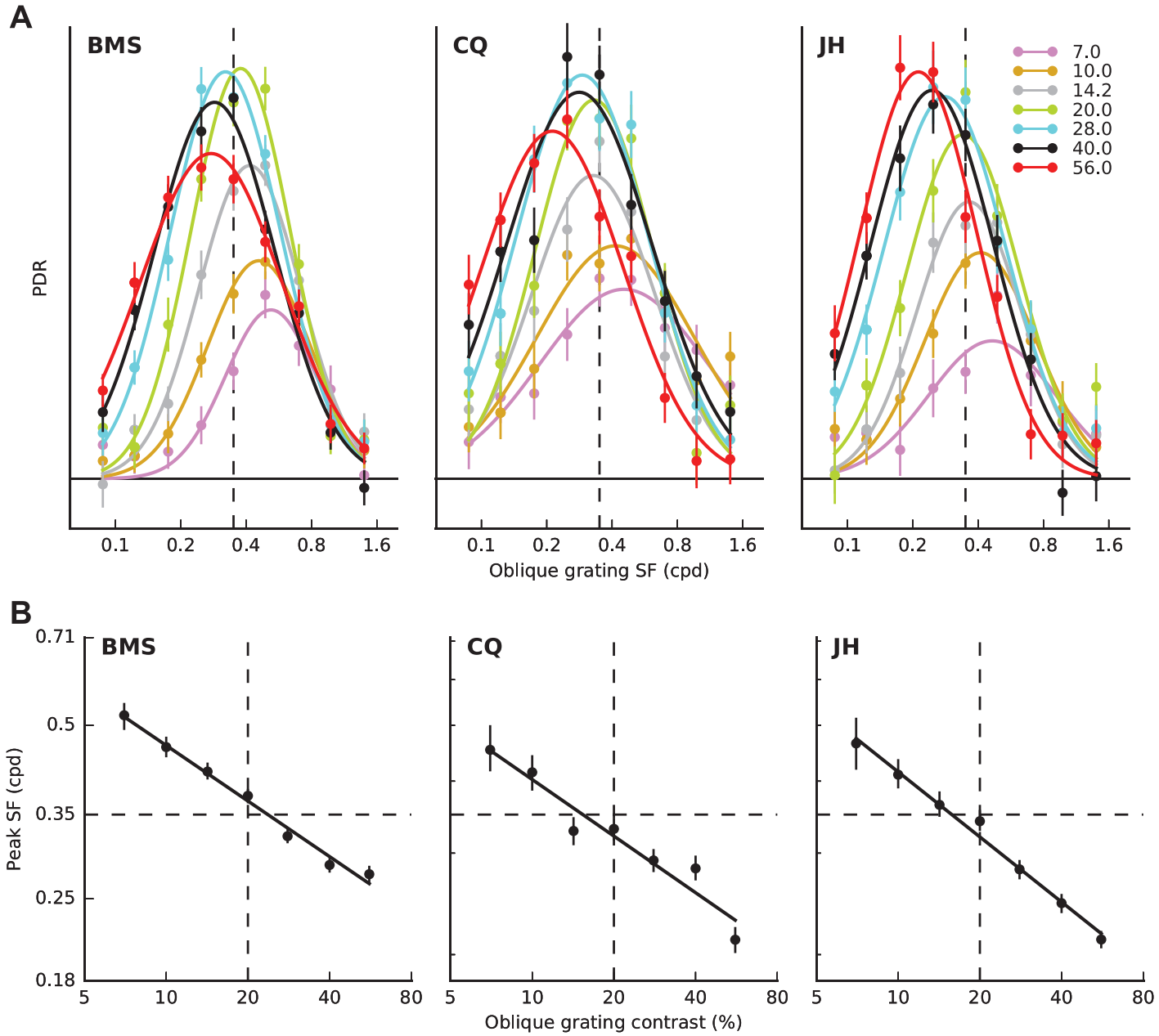

Figure 1. PDR to unidisparity plaids with unequal SF and contrast. The SF of the oblique grating associated with the strongest response varies as a function of the contrast of the oblique component. (A) PDR in each of the three subjects. Each curve (log-Gaussian fit to the data) shows the magnitude of the PDR (±*SEM*) for different values of the SF of the oblique grating for one contrast of the oblique grating (see legend). SF and contrast of the vertical grating are always the same. Dashed line indicates SF of the vertical grating. Horizontal line indicates zero response. The scale is different in the three subjects so that the axis top corresponds to 0.9 °/s for BMS, 0.32 °/s for CQ, and 0.58 °/s for JH. (B) The peak of each curve fit in panel A shifts to lower SFs as the contrast of the oblique grating increases. Linear fit in log–log coordinates is shown. Dashed lines indicate SF and contrast of vertical grating.

significantly ($p < 0.001$) smaller than zero: JH: −0.37 ± 0.04, BMS: −0.32 ± 0.02, CQ: −0.32 ± 0.04.

Modeling the PDR

Across our previously described (Quaia et al., 2013) and current experiments, there are several factors that consistently contribute to determining the strength of the PDR induced by unidisparity plaids. The major factors are the relative SF and the relative contrast of the two gratings, which, as we have shown here, interact. However, a role is also played by the absolute contrast of the two gratings as responses generally increase with overall contrast. Furthermore, the absolute SF of the grating also matters as only a limited

range of SFs is effective in inducing PDRs—and DVRs in general (Miura et al., 2008; Sheliga, FitzGibbon, & Miles, 2006).

To capture all of these effects, we developed a descriptive model, which uses as input the contrast and SF of each of the two gratings and produces as output the PDR. First, the Michelson contrast c_O of the oblique grating (the one with zero disparity) is converted into an effective contrast e_O, using the following Naka-Rushton equation:

$$e_O = \frac{c_O^n}{c_O^n + c_{50}^n}. \quad (1)$$

The Michelson contrast c_V of the vertical grating (the one with disparity) is similarly converted into an effective contrast e_V, using the following equation:

$$e_V = \frac{(c_V + c_b)^n}{(c_V + c_b)^n + c_{50}^n}. \quad (2)$$

The parameter c_b (where b stands for boost) is used here to account for the observation that when the two gratings have the same SF, the strongest PDR is measured not when they have the same contrast but rather when the oblique grating has a somewhat larger contrast (Quaia et al., 2013). We tentatively attributed this phenomenon to an internal boost of the vertical grating signal, applied by the brain to compensate for the fact that, because this grating has a fairly large disparity (1°), presumably it would not be in focus in natural viewing, reducing its apparent contrast. Multiplying c_V, or directly the entire function, by a number slightly larger than one would most likely be similarly effective (the few data points sensitive to this parameter poorly constrain the mechanism involved). Note that in our experiments all stimuli are physically presented in the plane of fixation and are thus in focus; disparity is artificially introduced using a mirror stereoscope.

Our previous results could then be accounted for by simply taking the sum of the two effective contrasts and multiplying it by four Gaussian functions:

$$\mathrm{PDR} = (e_V + e_O)\, e^{-\frac{(e_V - e_O)^2}{2\alpha^2}} e^{-\frac{(\log_{10}(SF_V) - \log_{10}(SF_O))^2}{2\beta^2}} e^{-\frac{(\log_{10}(SF_V) - \log_{10}(\mu))^2}{2\gamma^2}} e^{-\frac{(\log_{10}(SF_O) - \log_{10}(\mu))^2}{2\gamma^2}}, \quad (3)$$

where SF_V is the SF of the vertical grating, and SF_O is the SF of the oblique grating. The dispersion parameters α and β determine how rapidly the PDR decreases as differences in contrast and SF (respectively) are introduced between the two gratings. μ is the SF associated with the strongest pattern vergence response when the two gratings have the same SF and effective contrast, and the dispersion parameter γ accounts for how the PDR decreases as the SF of either grating deviates from this value.

Because Equation 3 is separable in SF and contrast, this model does not account for the dependency of the PDR on the joint value of SF and contrast and thus needs to be modified to reproduce the data reported here. As noted above, the pattern computation mechanism does not have access to the actual contrast of the components (c_V and c_O) but must instead rely on an internal estimate of it (e_V and e_O). In Equations 1 and 2, we assumed that this estimate is not dependent on SF, but this is not necessarily true. For example, it is well known that at detection threshold this internal estimate of contrast is far from faithful as demonstrated by the strong dependency of the contrast sensitivity function on SF, luminance, and spatial location (Campbell & Green, 1965; Daitch & Green, 1969; Georgeson & Sullivan, 1975; Hilz & Cavonius, 1974). At higher contrast, this estimate becomes less dependent on SF, but absolute contrast constancy is not achieved (Georgeson & Sullivan, 1975). It is thus conceivable that the internal estimate of contrast is biased in a SF-dependent manner, and this might account for our results. We present here two different ways to introduce this bias in internal contrast estimation. In the first one, which we dub model E (for early), we simply scale the effective contrast of each component by a factor k that increases with its SF separately for each component (i.e., high SFs are boosted relative to low SFs). Assuming that the scaling factor k increases linearly, in a log–log scale, with SF, we thus defined

$$x_{VE} = e_V\, e^{k \log(SF_V/2.0)}$$
$$x_{OE} = e_O\, e^{k \log(SF_O/2.0)}$$

In this formulation, the scaling factor is unitary when the SF is 2 c/°; because all the SFs used in our experiment are lower, in our simulations, the scaling factor is always smaller than one. The PDR in the overall model E is then estimated by replacing e_V and e_O with x_{VE} and x_{OE} in Equation 3.

In the second one, which we call model L (for late), we instead introduced reciprocal inhibition between the two effective contrasts and modulated the strength of this inhibition using the difference in SF between the two gratings so that the grating with the higher SF inhibits the one with the lower SF more strongly than vice versa. Assuming that the strength of inhibition varies linearly with the difference in $\log_{10}(SF)$, we thus defined

$$x_{VL} = \lfloor e_V - e_O[T + k(\log_{10}(SF_V) - \log_{10}(SF_O))] \rfloor$$
$$x_{OL} = \lfloor e_O - e_V[T + k(\log_{10}(SF_O) - \log_{10}(SF_V))] \rfloor$$

where $\lfloor\ \rfloor$ indicates that negative values are set to zero. The PDR in model L is then estimated by replacing e_V and e_O with x_{VL} and x_{OL} in Equation 3. There is one final parameter, present in both models, to scale the

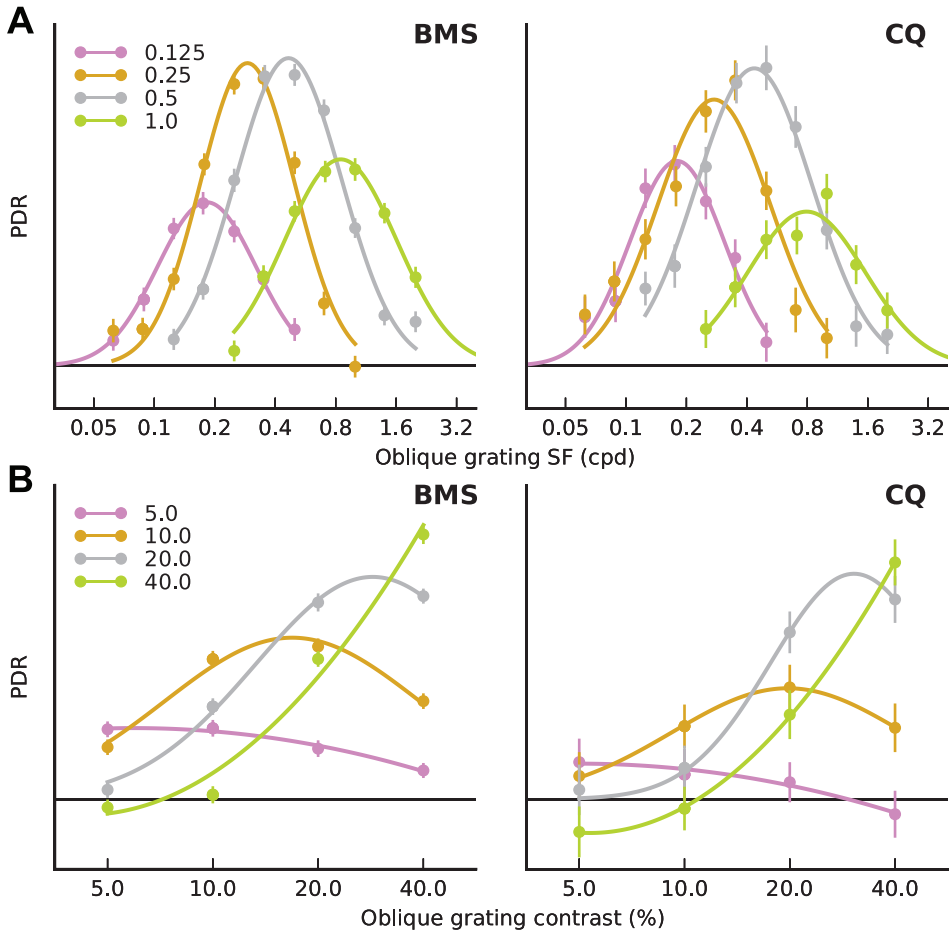

Figure 2. PDR to unidisparity plaids with unequal SF or contrast. (A) PDR in two subjects when the SF of the two gratings is varied and the contrast is the same (32%). Each curve shows the magnitude of the PDR (±*SEM*) for different values of the SF of the oblique grating, for one SF of the vertical grating (see legend). Horizontal line indicates zero response. The scale is different in the two subjects so that the axis top corresponds to 0.8 °/s for BMS and 0.4 °/s for CQ. (B) PDR in two subjects when the contrast of the two gratings is varied and the SF is the same (0.25 c/°). Each curve shows the magnitude of the PDR (±*SEM*) for different values of the contrast of the oblique grating for one contrast of the vertical grating (see legend). Horizontal line indicates zero response. The scale is different in the two subjects so that the axis top corresponds to 0.9 °/s for BMS and 0.35 °/s for CQ.

output of the model to best match the amplitude of the eye movements. This can be seen as a visuomotor gain and is not part of the optimization (for each set of parameters, the scaling value that results in the best fit to the data is computed through regression).

The models thus have either eight (model E) or nine (model L) parameters, many of which are well constrained only if the models are also fit to the data from our previous experiments. Two data sets are particularly useful: how the PDR varies as a function of SF_O for different values of SF_V (Figure 2A) and how the PDR varies as a function of c_O for different values of c_V (Figure 2B). Note that these data are available only for two of the three subjects as a different third subject participated in our previous study.

Running an optimization algorithm with eight or nine parameters to directly minimize a single error function on such a large number of data points (111 for BMS and CQ, 63 for JH) is almost guaranteed to result in a poor outcome (because of the large number of local minima across the design space). Instead, we extracted the parameters of the fits showed in Figures 1A and 2 and optimized the model parameters to reproduce those fits. The data in Figure 1A are well fit by log-Gaussian functions, from which we extracted the location of the peak (already showed in Figure 1B), the dispersion (measured in octaves), and the height of the peak. The dependence of these parameters on the contrast of the oblique grating is shown in the top row in Figures 3 and 4 and in Figure 5. The data in Figure 2A were also well fit by log-Gaussian functions from which we extracted the same parameters. Their dependency on the SF of the vertical grating is shown in Figures 3 and 4, middle rows. The contrast data (Figure 2B) is more difficult to handle. Because of the limited range tested, well-constrained log-Gaussian fits

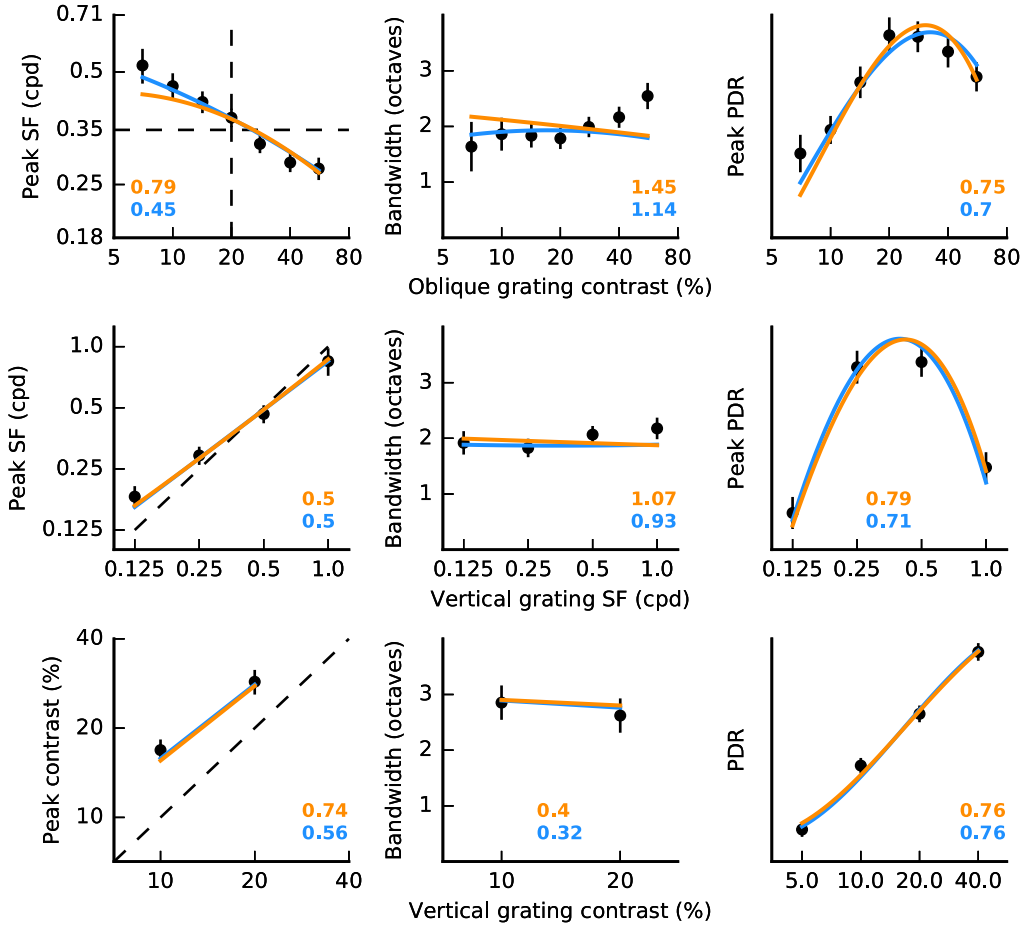

Figure 3. Model fits for subject BMS. Data (black circles, mean ± *SEM*) and model fits (orange lines for model E, blue lines for model L) to the experimental results presented in Figure 1A (top row), Figure 2A (middle row), and Figure 2B (bottom row). The numbers in each panel indicate, for each model, the error (average *z* score) on that data set, a measure of goodness of fit (lower is better).

are only available for two of the curves (c_V of 10% and 20%, yellow and gray curves), and thus, we extracted location of the peak and width from those fits. Their dependency on the contrast of the vertical grating is shown in the bottom row, left and center columns, in Figures 3 and 4. To cover a wider range of contrasts, instead of fitting the maximum response for those two fits, we simply fit the model directly to the PDR produced when the two gratings have the same contrast (four data points, shown in the bottom row, right column, in Figures 3 and 4).

By relying on these summary data, we considerably reduced the number of data points to be fit (41 for BMS and CQ, 21 for JH). Furthermore, we could now easily extract nine error measures, one for each panel in Figures 3 and 4, which were then used as minimization targets in a multiobjective algorithm (see Methods). Because of the more limited data available, in subject JH the optimization was further constrained by setting the values of c_b, α, and γ to the average values obtained for the other two subjects (these parameters are only minimally constrained by the data available for this subject).

The colored lines (orange: model E; blue: model L) in Figures 3 through 5 show the results of the optimization for our three subjects (model parameters used can be found in Tables 1 and 2 for models E and L, respectively). The models performed remarkably well for all subjects and in most conditions with only very few conditions in which the discrepancy between model and data was larger than 1 *SEM* of the data (indicated by the error bars in the figures). The late model generally performed better than the early model, but this model also has one more parameter and so it might be expected to do so (even though this is not strictly necessary because the two models are not nested). There are ways to quantify which might be the "better" model (e.g., by computing Akaike's information criterion or other goodness-of-fit measures that account for the number of parameters in the model), but our goal here is only to highlight that the same

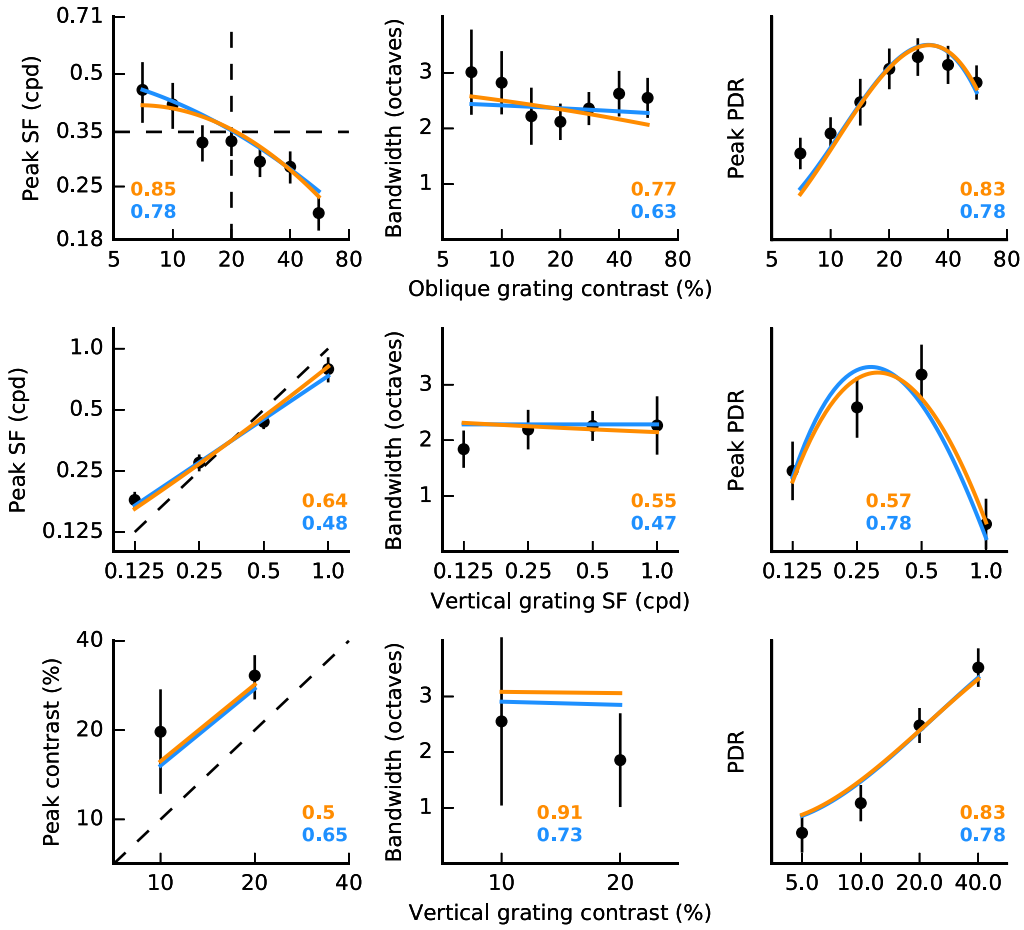

Figure 4. Model fits for subject CQ. Same format as in Figure 3.

overall result can be obtained invoking quite different mechanisms. Trends in tuning width (central column in the figures) were the least well captured, but they were also the least consistent across subjects, and hence, we did not feel that it was warranted to add additional parameters to the models to better account for them. Note that with any multiobjective optimization algorithm there is not a single way to identify an overall best design. We used a procedure that less heavily weighted tuning width (see Methods). In general, we found that all designs producing good fits had very narrowly clustered values for most parameters (c_{50}, n, c_b, α, β, μ, and γ). However, in model L, parameters k and T were more variable and generally negatively correlated (i.e., to some extent, they could be traded off against each other). This might explain, at least in part, the relatively large scatter for these parameters across subjects.

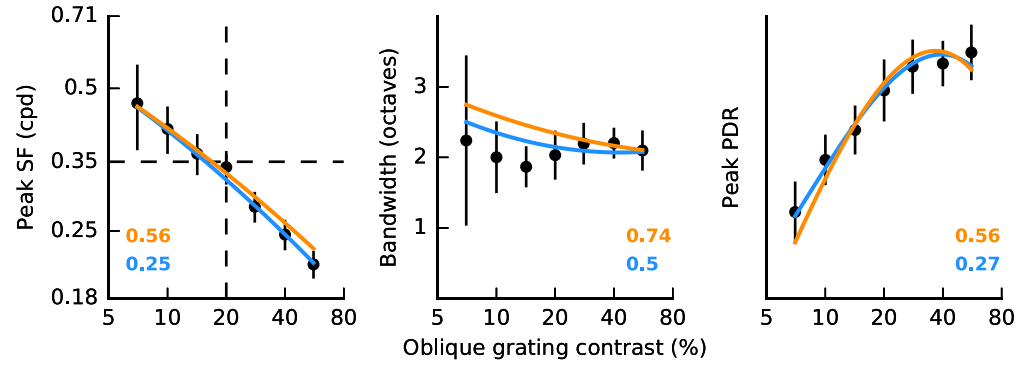

Figure 5. Model fits for subject JH. Data (black circles) and model fits (orange lines for model E, blue lines for model L) to the experimental results presented in Figure 1.

Subject	c_{50}	n	c_b	α	β	μ	γ	k
BMS	17	1.2	3.0	0.14	0.32	0.34	0.54	0.24
CQ	21	1.2	3.0	0.14	0.39	0.26	0.58	0.26
JH	12	1.2	3.0*	0.14*	0.45	0.2	0.56*	0.28

Table 1. Model E fits parameters. *Notes*: Parameters with an asterisk were not optimized but set to the mean value for the other subjects.

Subject	c_{50}	n	c_b	α	β	μ	γ	k	T
BMS	16	1.3	3.0	0.27	0.32	0.39	0.53	0.45	0.10
CQ	22	1.2	3.0	0.28	0.40	0.31	0.54	0.32	0.25
JH	16	0.8	3.0*	0.275*	0.43	0.20	0.535*	0.50	0.50

Table 2. Model L fits parameters. *Notes*: Parameters with an asterisk were not optimized but set to the mean value for the other subjects.

Discussion

We recorded vergence eye movements in humans to further refine our understanding of how components are combined to extract pattern disparity from stereo plaids. We found that the SF and contrast of the two gratings jointly determine the magnitude of the vergence responses in a nonseparable manner and presented two simple mathematical models that well describe these findings (as well as our previous results). These nonseparable responses further reveal that the pattern disparity mechanism combines filters spanning a broader range of SFs than we had previously reported. We had concluded, based on the data shown in Figure 2A and in the central panel of Figures 3 and 4, that the range of SFs over which components are effectively combined (i.e., the relative FWHH SF bandwidth of the mechanism) spanned approximately two octaves. When combinations across different values of contrast (i.e., different tuning curves in Figure 1) are considered, it becomes clear that we underestimated this measure. The FWHH bandwidth can thus be estimated from the data in Figure 1 by taking the ratio between the outside half-height points of the tuning functions associated with the lowest and highest contrast of the oblique grating. The relative bandwidth is then significantly larger than we previously reported: three octaves in BMS, 3.87 octaves in CQ, and 3.3 octaves in JH. This appears to be broader than the average width of a SF channel (Anderson & Burr, 1989; R. De Valois & De Valois, 1988), but it is certainly not outside the range normally observed in monkey striate cortex neurons (Wang & Movshon, 2016). Nevertheless, the nonseparability of the effects of contrast and SF implies that, in extracting pattern disparity, information must be pooled across different scale channels (somehow weighted according to their preferred SF). This is because the output of any single channel is a separable function of SF and contrast, so changes in peak frequency as a function of SF, like those shown in Figure 1, cannot arise from combining two single channels (one for each orientation). For each orientation, neurons tuned to multiple SFs must participate in the extraction of plaid disparity.

The nonseparable interaction of SF and contrast could emerge in many different ways. We explored two descriptive models although other formulations are certainly possible. Both models assume that components are combined based on the similarity of their SFs and the internal estimate of their contrasts. The combination is itself separable, but the internal contrast estimates vary with SF, thus bringing about the nonseparable interaction between SF and physical contrast. Combining components based on their similarity makes sense because, obviously, components should be combined (according to the IOC rule) only if they belong to the same object, and it is unlikely that components with very different contrasts or SF belong to the same object. This argument is not original: It has been used before to explain why differences in contrast or SF result in a breakdown in perceptual coherence for drifting plaids (Adelson & Movshon, 1982). Under these hypotheses, a strong sensitivity of the PDR to differences between the brain's estimates of the SF and contrast of the 1-D components is to be expected.

We had previously reported that when the two components have the same SF, the strongest response is observed when the zero-disparity grating has a higher contrast (Figure 2B). We argued that this might result from an internal boosting of the response of the filter that responds to the grating with disparity. Such boosting may in fact produce a more accurate estimate of the contrast of that component because, under natural conditions, a grating with such a large disparity would be out of focus. We thus suggested that, as a result of the optical properties of the eye, stimulus disparity might be taken into account when estimating contrast. With the two models that we present here, we now consider another factor that might play a part in how the brain computes its internal estimate of component contrast: SF. In the two models, the dependency of the internal estimate of contrast on SF arises in different ways. If we were to interpret the models in the context of a multichannel scheme, in the first ("early") model the contrast gain varies with the preferred SF of the channel (channels tuned to lower SF having lower gain) independently for the two gratings (i.e., orientations). In the second ("late") model, mutual inhibition between the 1-D filters that detect the two gratings—with stronger inhibition from channels tuned to high SFs to those tuned to low SFs—is responsible for the nonseparability. Both models fit the data well but do not speak to the reason for

introducing a SF dependency in computing the internal estimate of contrast.

One possible explanation, given the properties of early visual processes and of visual stimuli in nature, is that the interaction between SF and contrast may be a by-product of a mechanism that equalizes the long-term average response of neurons tuned to different spatial frequencies, the so-called response equalization hypothesis (D. Graham, Chandler, & Field, 2006). The idea that such a mechanism might operate at the level of cortical (or even subcortical) visual neurons has been put forward both as a means of achieving contrast constancy (i.e., to correctly estimate stimulus contrast irrespective of SF) and as a way of guaranteeing that neuronal resources are efficiently used given the properties of the stimuli to which we are naturally exposed.

It is well known that, because of optical and neural constraints, the sensitivity of the visual system (i.e., the inverse of the contrast at detection threshold) varies widely with SF, luminance, and spatial location (Campbell & Green, 1965; Daitch & Green, 1969; Georgeson & Sullivan, 1975; Hilz & Cavonius, 1974). Because sensitivity depends on the ratio between signal and noise, not much can be done to remove these dependencies. At higher contrast, when the noise becomes a small fraction of the signal, contrast estimates can, however, be made less dependent on these nuisance parameters (i.e., contrast constancy can be achieved) by simply appropriately adjusting the gain. At moderate-to-high contrasts, human vision comes close to achieving this property (Blakemore et al., 1973; Bowker, 1983; Bryngdahl, 1966; Georgeson & Sullivan, 1975; Swanson et al., 1988; Watanabe et al., 1968) although not before the 12th week of life (Stephens & Banks, 1985). Georgeson and Sullivan (1975) proposed that, to achieve contrast constancy, the gain of each channel could simply be adjusted so that, over a long period of time, each channel is equally active on average. Based on additional findings, they further suggested that this gain regulation would be applied to orientation-sensitive channels and thus at the cortical level.

However, the extent to which contrast constancy is achieved varies across the SF spectrum, being better for high (above 2 c/°) than for low SFs (Georgeson & Sullivan, 1975). We suggest that this partial failure at low SFs might be responsible for the interaction between SF and contrast that we unveiled in our experiments (in which, notably, all the stimuli had a SF lower than 2 c/°). To understand why this might happen, one needs to consider the biological substrate of the SF channels considered by the contrast constancy theory and the distribution of contrast across SFs in natural images—factors that are the basis for the coding efficiency theory of response equalization. It is well known (R. De Valois & De Valois, 1988) that V1 neurons respond well only to a limited band of SFs and that their SF tuning curve is generally well described as a log-Gaussian function (i.e., symmetric in log-SF space, not in linear SF space). Assuming a constant bandwidth (in log-SF space), if stimulated with a white noise pattern (i.e., a stimulus with a flat spectrum in linear SF space), neurons tuned to high SFs would be stimulated much more strongly than neurons tuned to low SFs. However, in the natural world, the spectrum of SFs is not flat and follows a 1/f distribution: On average, contrast energy decreases as SF increases. Neurons with a log-Gaussian tuning function and fixed bandwidth for SF are a good match for this type of energy distribution: If presented with natural images, neurons of this type tuned to different SFs would, on average, be equally stimulated (Brady & Field, 1995; Field, 1987). Put another way, such neurons effectively boost high SFs, which are relatively underrepresented in natural images, resulting in a whitening of the spectrum and an efficient use of resources (Atick & Redlich, 1992; Barlow, 2001; Brady & Field, 1995, 2000; Field, 1987; Field & Brady, 1997; D. Graham et al., 2006; Simoncelli & Olshausen, 2001). The log-Gaussian tuning of early visual neurons can then be seen as a developmental adaptation to the statistics of the visual environment, one of the many that have been proposed (Boots, Nundy, & Purves, 2007), which can be brought about by a mechanism of response equalization (easily implemented neurally as an unsupervised learning rule).

To see why this scheme might fail at low SFs, one needs to consider that the SF bandwidth of cortical neurons is constant only in channels tuned to high SFs. For channels tuned to less than about 2 c/°, bandwidth increases as the preferred SF decreases (Anderson & Burr, 1985, 1987, 1989; R. De Valois, Albrecht, & Thorell, 1982; Fiorentini, Pirchio, & Spinelli, 1983; Furchner, Thomas, & Campbell, 1977; Kulikowski & Bishop, 1981; Stromeyer, Klein, Dawson, & Spillmann, 1982; Wilson, McFarlane, & Phillips, 1983). This probably reflects a physical limitation: To keep the bandwidth constant, the size of the detectors would have to keep increasing as the preferred SF is reduced, which would require unrealistically large receptive fields for low SFs. Accordingly, when presented with natural images, low SF channels would be stimulated (on average) more strongly than high SF channels. A learning rule that adjusted gain based on long-term average activity would then reduce the gain of low SF channels with the gain inversely proportional to the bandwidth of the channel. We propose that the stronger weight given to high SF components (relative to low SF components) in our models reflect these unequal gains.

This hypothesis maps naturally into one of the two models we proposed above (the early model). Furthermore, with this model it is straightforward to test this hypothesis quantitatively: The scaling factor k in our model should match the dependency of bandwidth on SF when expressed as a slope in log–log coordinates, and indeed it does. In our model, the best fitting values for this parameter (see Table 1) are 0.24 (BMS), 0.26 (CQ), and 0.28 (JH); Anderson and Burr (1989) have reported that, in their two subjects, the bandwidth decreases at a rate of 0.26 octaves per octave of spatial frequency for SFs up to 2 c/° and is constant for higher SFs. This quantitative agreement makes it tempting to interpret our experiments as a confirmation of the response equalization rule proposed by Georgeson and Sullivan (1975) for achieving contrast constancy. However, because we obtain equally good fits with an alternative model, this is not a strong test of their hypothesis.

Even the interactions between spatial frequencies present in our "late" model can be interpreted within the same framework. In that model, it is mutual inhibition between different scales that accounts for the nonseparability we observed, but the overall effect on contrast gain control is similar. There is considerable evidence to support such cross-scale reciprocal inhibition in both humans and experimental animals (Albrecht & De Valois, 1981; K. De Valois & Tootell, 1983; McCourt & Foley, 1985; Movshon, Thompson, & Tolhurst, 1978; Tolhurst, 1972). Notably, an asymmetry, with high SFs exerting stronger suppression than low SFs, has been reported (Albrecht & De Valois, 1981; K. De Valois & Tootell, 1983) just as required by the model. Importantly, achieving contrast constancy has been proposed as a potential role for such asymmetric high-on-low SF inhibition (McCourt & Foley, 1985).

Regardless of the exact mechanism, the joint dependency on SF and contrast that we reported here might reflect the brain's attempt to correctly infer the contrast of stimuli in the visual world across the entire SF spectrum despite the limited size of neuronal receptive fields.

Conclusions

The idea that the components of sinusoidal plaid stimuli are simply combined according to their similarity in SF and physical contrast holds only as a first order approximation. When the effects of SF variations are measured at different contrasts, it becomes apparent that the pattern response component of the DVRs is a nonseparable function of contrast and SF. We show that this can be explained if the brain's internal estimate of physical contrast is affected by SF. Although puzzling at first sight, such a dependency between contrast and SF might be simply the side effect of a mechanism tasked with correctly and efficiently inferring contrast energy in natural images given the constraints on wiring in the early stages of the visual system.

Keywords: pattern disparity, *vergence eye movements*, *contrast constancy*, *natural images*, *whitening*

Acknowledgments

The authors thank Dr. Boris M. Sheliga for assistance in collecting the data and two reviewers for their constructive suggestions and comments. This work was supported by the NEI Intramural Research Program.

Commercial relationships: none.
Corresponding author: Christian Quaia.
Email: quaiac@nei.nih.gov.
Address: Laboratory of Sensorimotor Research, National Eye Institute, NIH, DHHS, Bethesda, MD, USA.

References

Adelson, E., & Movshon, J. (1982, Dec 9). Phenomenal coherence of moving visual patterns. *Nature*, *300*(5892), 523–525.

Albrecht, D., & De Valois, R. (1981). Striate cortex responses to periodic patterns with and without the fundamental harmonics. *Journal of Physiology*, *319*, 497–514.

Anderson, S., & Burr, D. (1985). Spatial and temporal selectivity of the human motion detection system. *Vision Research*, *25*, 1147–1154.

Anderson, S., & Burr, D. (1987). Receptive field size of human motion detection units. *Vision Research*, *27*, 621–635.

Anderson, S., & Burr, D. (1989). Receptive field properties of human motion detector units inferred from spatial frequency masking. *Vision Research*, *29*, 1343–1358.

Atick, J., & Redlich, A. (1992). What does the retina know about natural scenes. *Neural Computation*, *4*, 196–210.

Barlow, H. (2001). Redundancy reduction revisited. *Network*, *12*, 241–253.

Blakemore, C., Muncey, J., & Ridley, R. (1973).

Stimulus specificity in the human visual system. *Vision Research*, *13*, 1915–1931.

Boots, B., Nundy, S., & Purves, D. (2007). Evolution of visually guided behavior in artificial agents. *Network*, *18*, 11–34.

Bowker, D. (1983). Suprathreshold spatiotemporal response characteristics of the human visual system. *Journal of the Optical Society of America*, *73*, 436–440.

Brady, N., & Field, D. (1995). What's constant in contrast constancy? The effects of scaling on the perceived contrast of bandpass patterns. *Vision Research*, *35*, 739–756.

Brady, N., & Field, D. (2000). Local contrast in natural images: Normalisation and coding efficiency. *Perception*, *29*, 1041–1055.

Bryngdahl, O. (1966). Characteristics of the visual system. Psychophysical measurements of the response to spatial sine-wave stimuli in the photopic region. *Journal of the Optical Society of America*, *56*, 811–821.

Busettini, C., Masson, G., & Miles, F. (1996, Mar 28). A role for stereoscopic depth cues in the rapid visual stabilization of the eyes. *Nature*, *380*(6572), 342–345.

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

Collewijn, H., van der Mark, F., & Jansen, T. (1975). Precise recording of human eye movements. *Vision Research*, *15*, 447–450.

Daitch, J., & Green, D. (1969). Contrast sensitivity of the human peripheral retina. *Vision Research*, *9*, 947–952.

DeAngelis, G., Robson, J., Ohzawa, I., & Freeman, R. (1992). Organization of suppression in receptive fields of neurons in cat visual cortex. *Journal of Neurophysiology*, *68*, 144–163.

De Valois, K., & Tootell, R. (1983). Spatial-frequency-specific inhibition in cat striate cortex cells. *Journal of Physiology*, *336*, 359–376.

De Valois, R., Albrecht, D., & Thorell, L. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, *22*, 545–559.

De Valois, R., & De Valois, K. (1988). *Spatial vision*. Oxford, England: Oxford University Press.

Farell, B. (1998, Oct 15). Two-dimensional matches from one-dimensional stimulus components in human stereopsis. *Nature*, *395*(6703), 689–693.

Fennema, C., & Thompson, W. (1979). Velocity determination in scenes containing several moving images. *Computer Graphics and Image Processing*, *9*, 301–315.

Field, D. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, *4*, 2379–2394.

Field, D. (1999). Wavelets, vision and the statistics of natural scenes. *Philosophical Transactions of the Royal Society A Mathematical Physician and Engineering Sciences*, *357*, 2527–2542.

Field, D., & Brady, N. (1997). Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Research*, *37*, 3367–3383.

Fiorentini, A., Pirchio, M., & Spinelli, D. (1983). Electrophysiological evidence for spatial frequency selective mechanisms in adults and infants. *Vision Research*, *23*, 119–127.

Furchner, C., Thomas, J., & Campbell, F. (1977). Detection and discrimination of simple and complex patterns at low spatial frequencies. *Vision Research*, *17*, 827–836.

Georgeson, M., & Sullivan, G. (1975). Contrast constancy: Deblurring in human vision by spatial frequency channels. *Journal of Physiology*, *252*, 627–656.

Goldberg, D. (1989). *Genetic algorithms in search, optimization, and machine learning*. Reading, MA: Addison-Wesley.

Graham, D., Chandler, D., & Field, D. (2006). Can the theory of "whitening" explain the center-surround properties of retinal ganglion cell receptive fields? *Vision Research*, *46*, 2901–2913.

Graham, N. (1989). *Visual pattern analyzers*. New York: Oxford University Press.

Hilz, R., & Cavonius, C. (1974). Functional organization of the peripheral retina: Sensitivity to periodic stimuli. *Vision Research*, *14*, 1333–1337.

Hubel, D., & Wiesel, T. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, *195*, 215–243.

Kulikowski, J. (1976). Effective contrast constancy and linearity of contrast sensation. *Vision Research*, *16*, 1419–1431.

Kulikowski, J., & Bishop, P. (1981). Linear analysis of the responses of simple cells in the cat visual cortex. *Experimental Brain Research*, *44*, 386–400.

McCourt, M., & Foley, J. (1985). Spatial frequency interference on grating-induction. *Vision Research*, *25*, 1507–1518.

Miura, K., Sugita, Y., Matsuura, K., Inaba, N., Kawano, K., & Miles, F. (2008). The initial

disparity vergence elicited with single and dual grating stimuli in monkeys: Evidence for disparity energy sensing and nonlinear interactions. *Journal of Neurophysiology*, *100*, 2907–2918.

Morrone, M., Burr, D., & Maffei, L. (1982). Functional implications of cross-orientation inhibition of cortical visual cells. I. Neurophysiological evidence. *Proceedings of the Royal Society of London Series B-Biological Sciences*, *216*, 335–354.

Movshon, J., Thompson, I., & Tolhurst, D. (1978). Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. *Journal of Physiology*, *283*, 101–120.

Priebe, N., Lisberger, S., & Movshon, J. (2006). Tuning for spatiotemporal frequency and speed in directionally selective neurons of macaque striate cortex. *Journal of Neuroscience*, *26*, 2941–2950.

Quaia, C., Optican, L., & Cumming, B. (2016). A motion-from-form mechanism contributes to extracting pattern motion from plaids. *Journal of Neuroscience*, *36*, 3903–3918.

Quaia, C., Sheliga, B., Optican, L., & Cumming, B. (2013). Temporal evolution of pattern disparity processing in humans. *Journal of Neuroscience*, *33*, 3465–3476.

Sheliga, B., FitzGibbon, E., & Miles, F. (2006). Short-latency disparity vergence eye movements: A response to disparity energy. *Vision Research*, *46*, 3723–3740.

Simoncelli, E., & Olshausen, B. (2001). Natural image statistics and neural representation. *Annual Reviews of Neuroscience*, *24*, 1193–1216.

Stephens, B., & Banks, M. (1985). The development of contrast constancy. *Journal of Experimental Child Psychology*, *40*, 528–547.

Stromeyer, C., Klein, S., Dawson, B., & Spillmann, L. (1982). Low spatial-frequency channels in human vision: Adaptation and masking. *Vision Research*, *22*, 225–233.

Swanson, W., Georgeson, M., & Wilson, H. (1988). Comparison of contrast responses across spatial mechanisms. *Vision Research*, *28*, 457–459.

Tolhurst, D. (1972). Adaptation to square-wave gratings: Inhibition between spatial frequency channels in the human visual system. *Journal of Physiology*, *226*, 231–248.

van Soest, A., & Casius, L. (2003). The merits of a parallel genetic algorithm in solving hard optimization problems. *Journal of Biomechanical Engineering*, *125*, 141–146.

Wang, H., & Movshon, J. (2016). Properties of pattern and component direction-selective cells in area MT of the macaque. *Journal of Neurophysiology*, *115*, 2705–2720.

Watanabe, A., Mori, T., Nagata, S., & Hiwatashi, K. (1968). Spatial sine-wave responses of the human visual system. *Vision Research*, *8*, 1245–1263.

Wilson, H., McFarlane, D., & Phillips, G. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research*, *23*, 873–882.