Spatially extensive summation of contrast energy is revealed by contrast detection of micro-pattern textures

Tim S. Meese   School of Life and Health Sciences, Aston University, Birmingham, UK

Vision must analyze the retinal image over both small and large areas to represent fine-scale spatial details and extensive textures. The long-range neuronal convergence that this implies might lead us to expect that contrast sensitivity should improve markedly with the contrast area of the image. But this is at odds with the orthodox view that contrast sensitivity is determined merely by probability summation over local independent detectors. To address this puzzle, I aimed to assess the summation of luminance contrast without the confounding influence of area-dependent internal noise. I measured contrast detection thresholds for novel *Battenberg* stimuli that had identical overall dimensions (to clamp the aggregation of noise) but were constructed from either dense or sparse arrays of micro-patterns. The results unveiled a three-stage visual hierarchy of contrast summation involving (i) spatial filtering, (ii) long-range summation of coherent textures, and (iii) pooling across orthogonal textures. Linear summation over local energy detectors was spatially extensive (as much as 16 cycles) at Stage 2, but the resulting model is also consistent with earlier classical results of contrast summation (J. G. Robson & N. Graham, 1981), where co-aggregation of internal noise has obscured these long-range interactions.

Keywords: human vision, psychophysics, energy model, texture perception, spatial summation, area summation

Introduction

Most visual environments contain fine detailed information that requires analysis over tiny distances across the retina. For example, a blade of grass (2 mm wide) seen from a distance of 57 cm produces a retinal image that is only 60 μm wide (0.2 deg). The human primary visual cortex contains visual neurons with receptive fields that are well suited to this scale of analysis, and these mechanisms (filter elements) are easily capable of resolving the blades of grass. But how might an entire sports field be represented? One possibility is that it is encoded by mechanisms that operate at a coarser scale of analysis, such that the sports field falls within a single large receptive field. An obvious method by which this could be achieved is through neuronal convergence, where higher-order texture mechanisms pool over many lower-order filter elements (Bergen & Adelson, 1988; Lennie, 1998; Motoyoshi, Nishida, Sharan, & Adelson, 2007; Pollen, Przybyszewski, Rubin, & Foote, 2002; Victor & Conte, 2005; Wilson & Wilkinson, 1998). In this scheme, the properties of the filter elements (that respond to luminance contrast) and their variation across the retina characterize the texture mechanisms, which could encode various attributes of the patterns including their form and regularity (Bergen & Adelson, 1988; Wilson & Wilkinson, 1998), their depth gradients (Meese & Holmes, 2004), and other parameters (Kingdom & Keeble, 1996;

Li & Zaidi, 2000). However, the orthodox interpretation of typical psychophysical detection studies does not fit well with this idea. When contrast detection threshold is measured as a function of the area of a sine-wave grating, larger areas offer a benefit no greater than that predicted by probability summation (PS) among many small filter elements, each perturbed by independent additive noise (Anderson & Burr, 1991; Foley, Varadharajan, Koh, & Farias, 2007; Howell & Hess, 1978; Meese & Williams, 2000; Meese, Hess, & Williams, 2005; Robson & Graham, 1981) (Figure 1a). Thus, the standard psychophysical model of contrast detection makes no provision for detecting the sports field (large luminance contrast area) *per se*; aggregation at threshold is limited to individual blades of grass (local luminance contrast elements).

To address this conundrum, we must first consider some formal properties of several models of signal summation (for mathematical derivations, see Appendix A). For contrast detection, the PS model is usually implemented by Minkowski summation (Bonneh & Sagi, 1998; Meese & Williams, 2000; Quick, 1974; Robson & Graham, 1981; Watson, 1979; Watson & Ahumada, 2005) over $i = 1$ to n independent detecting mechanisms as follows:

$$\text{resp}_{\text{overall}} = \left[\sum_{i=1:n} |r_i|^{mink} \right]^{1/mink}, \qquad (1)$$

where, r_i *is* the response of the *i*th mechanism in the pool, $\text{resp}_{\text{overall}}$ is the observer's decision variable (Figure 1f),

and the Minkowski exponent (*mink*) is typically about 4 (Anderson & Burr, 1991; Bonneh & Sagi, 1998; Robson & Graham, 1981; Meese & Williams, 2000; Meese et al., 2005; Tyler & Chen, 2000). For a sine-wave grating, this predicts that contrast sensitivity should improve in proportion to the fourth-root of its area ($area^{1/4}$). Another way of expressing the benefit of increasing stimulus area is as a summation ratio (SR), where SR = *thresh*(*area*/2)/*thresh*(*area*), or 20 times the $\log_{10}$ of this when expressed

(a) Probability summation

(b) Linear summation

(c) Energy model

(d) Ideal summation

(e) Combination model

in dB. The variable *thresh*() is the contrast at detection threshold for each of two stimuli, one of which has twice the area of the other. Thus, in Figure 1, these stimuli would excite either $n/2$ or n input lines. Assuming equally sensitive input lines this gives an SR ~1.5 dB (a factor of ~1.2) for PS (and *mink* = 4), consistent with human data (Bonneh & Sagi, 1998; Foley et al., 2007; Meese & Williams, 2000; Meese et al., 2005; Meese & Hess, 2007; Meese & Summers, 2007; Robson & Graham, 1981). This is markedly less than the 6 dB (factor of 2) prediction made by the linear summing model (Figure 1b), where the signals combine linearly against a background of fixed internal noise. Note that in this model (and the next), summation extends over a fixed number of signal lines, some of which might not always carry signal.

Another possibility, for which there is growing evidence (Goris, Wagemans, & Wichmann, 2008; Lu & Dosher, 2008; Meese & Summers, 2009), is that the signals first pass through accelerating contrast transducers (point-wise nonlinearities). When the transducer has a contrast exponent of 2 (a square-law), this forms the basis of the energy model (Green & Swets, 1966; Kersten, 1984; Kukkonen, Rovamo, Tiippana, & Näsänen, 1993; Manahilov, Simpson, & McCulloch, 2001; Watson, Barlow, & Robson, 1983) (Figure 1c). However, this model also predicts too much summation (SR = 3 dB; a factor of $\sqrt{2}$) compared to empirical results, at least in the central visual field (Foley et al., 2007; Manahilov et al., 2001; Meese et al., 2005; Meese & Hess, 2007).

Consider next the ideal summation model (Figure 1d). This model is similar to the linear summation model but the observer is able to perform selective pooling, choosing the range of signals over which summation takes place, thereby matching the summation region to the stimulus. In this case, the observer can perform the ideal strategy of bypassing the summation of internal noise associated with the irrelevant inputs (Campbell & Green, 1965; Meese & Holmes, 2004; Tyler & Chen, 2000). With this
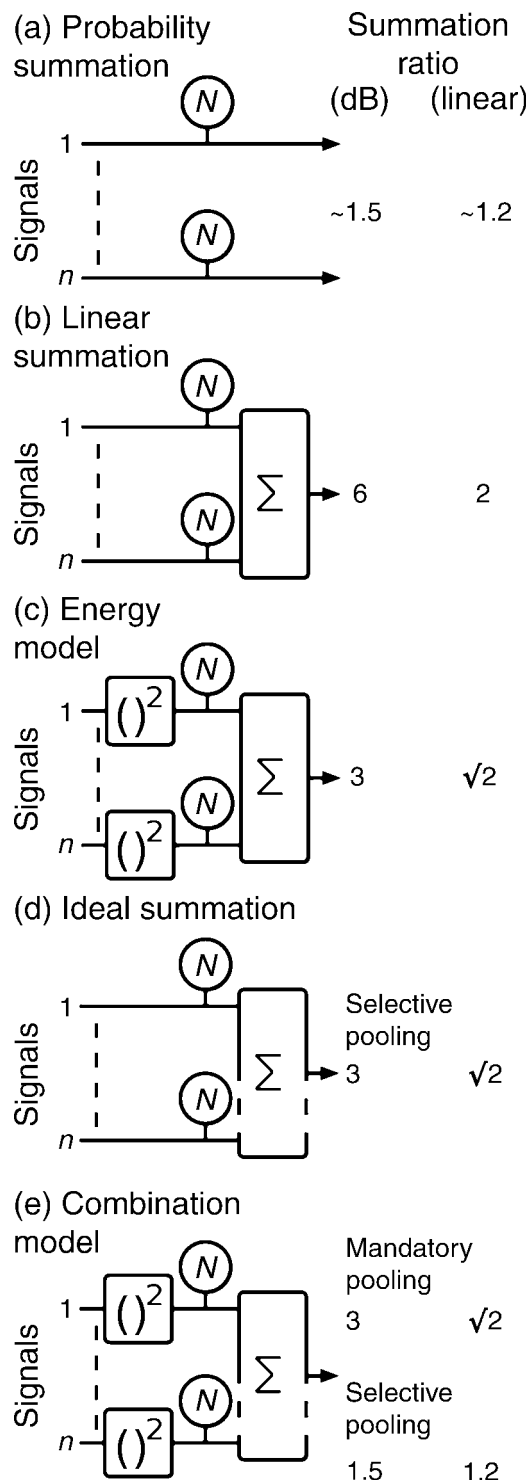
Figure 1. Various model architectures and their summation ratios (SR). In each case, the SR is for the situation where the number of signals is doubled from $n/2$ to n, for all even n. (a) Independent signals are perturbed by independent noise (N = zero mean, unit variance, additive Gaussian noise) and are combined probabilistically. An SR of 1.5 dB is consistent with the widely used fourth-root summation rule (*mink* = 4 in Minkowski summation). (b) Mandatory linear summation. (c) Signals are squared and followed by additive noise before mandatory summation to calculate energy. (d) The linear summation model but without restriction to mandatory summation. In this model, the observer selects only the relevant input lines, permitting ideal summation of signal and noise. (e) The combination model for which the SR depends upon whether summation is selective or mandatory. It behaves like the energy model when it is mandatory and the PS model when it is selective. Architecturally, this model combines features from the energy model (c) and the ideal summation model (d).

arrangement, sensitivity will always be greater than or equal to that achieved with the linear summation model (Figure 1b), but we must keep in mind that *summation* is a relative measure. Thus, the benefit of doubling the number of signals is not as great as in the linear summation model because performance is being compared with a more efficient (less noisy) starting point. The ideal model predicts SR = 3 dB (because it selectively sums both signal and noise), the same prediction as for the energy model. Thus, although linear summation (Figure 1b), energy measures (Figure 1c), and ideal summation (Figure 1d) might each be relevant to our interest in the aggregation of visual contrast texture, none appears consistent with the classical fourth-root empirical result at threshold.

Finally, consider the *combination* model in Figure 1e. This involves the square-law transducer, as in the energy model, and a facility for selective pooling, as in the ideal summation model. The cascade of these effects means that the model predicts a fourth-root result (SR = 1.5 dB), exactly the same as the PS model (Meese & Summers, 2007). This offers a solution to the conundrum above: perhaps neuronal convergence for contrast detection does take place over area after all but offers a performance benefit at threshold no better than probability summation. Since the completely different architectures in Figures 1a and 1e make identical predictions, how might we tell them apart?

To do this, I used micro-patterns like those in Figure 2a to construct the novel stimuli in Figure 3, which I call *Battenbergs* (see figure caption). This stimulus arrangement has two advantages over the conventional sine-wave grating. First, it allows the contrast area to be manipulated without changing the overall stimulus dimensions. This is a desirable property as I now explain. Let's assume that the observer does not have templates matched to these peculiar stimulus configurations (Meese & Summers, 2007; Näsänen, Kukkonen, & Rovamo, 1994). Let's also assume that summation is uniform over signal and gap regions for Battenbergs ($j \geq 1$; Figure 3). It then follows that the signal-to-noise ratio (SNR) is given by:

$$SNR \propto nC^p / \sqrt{(2n)}, \qquad (2)$$

where n is the number of signal elements and C is the signal contrast. From this, it is easy to show that contrast sensitivity increases with n for any positive nonlinearity p. Thus, even though the proposed strategy aggregates noise from gap regions that contain no signal, the system still benefits from aggregating over as large a stimulus area as possible. For the combination model, this strategy holds the overall variance of internal noise constant (the summation region is the same for each of the stimuli in Figure 3a), which means that it behaves exactly like the energy model (Figure 1c) and the predicted level of summation exceeds that for PS (Figures 1a and 1e). Of course, if the assumptions above are incorrect and the system is able to perform selective pooling by rejecting the gap regions from aggregation, then the experiment will have
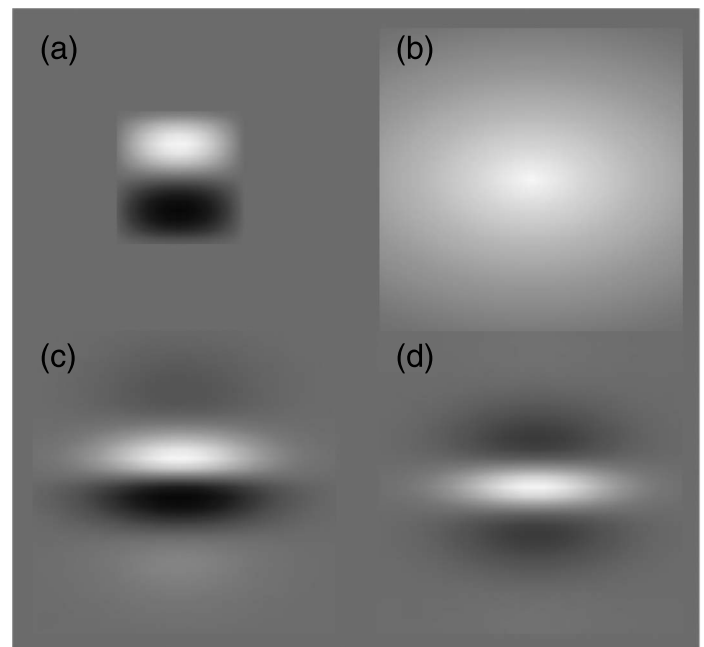

Figure 2. Stimulus and model elements. (a) A micro-pattern made from a single square cycle of a sine-wave grating (2.5 c/deg) multiplied by an orthogonal half-cycle of a cosine function. The Michelson contrasts of our stimuli were identical to the Michelson contrasts of these elements. The experiments measured sensitivity to these contrasts. (b) Spatial weighting used to simulate retinal inhomogeneity across the stimulus region. This is derived from experiments that have reported a sensitivity loss of 0.3 dB per cycle in the horizontal meridian and 0.5 dB per cycle in the vertical meridian (Pointer & Hess, 1989). (c, d) Sine and cosine phase log Gabor filter elements used in the filter models (spatial frequency bandwidth = 1.6 octaves and orientation bandwidth = ±25° at half height). Note that panels a, c, and d are to the same scale. The "attenuation field" in panel b is to a much smaller scale.

failed to achieve its aim and behavior will be indistinguishable from probability summation (Figures 1a and 1e).

The second advantage of the Battenberg is that the stimuli used here are almost immune from the deleterious effects of retinal inhomogeneity (Pointer & Hess, 1989) because of their signal dispersion (Appendix B). This improves on previous work by providing an effective way of studying area summation of contrast in the central visual field, which is where most visual processing is performed for everyday visual tasks.

For clarity, I now review the five models in Figure 1 in the contexts of grating and Battenberg stimuli. Strictly speaking, the stimulus type makes no difference for summation predictions in any of the models. The predictions are derived from the simple consideration of what happens when the number of signals is doubled, regardless of their spatial dispersion. Because the models in Figures 1a–1c have no flexibility, they *must* predict the same behavior for each stimulus type. The crucial point for the combination model (Figure 1e), as discussed
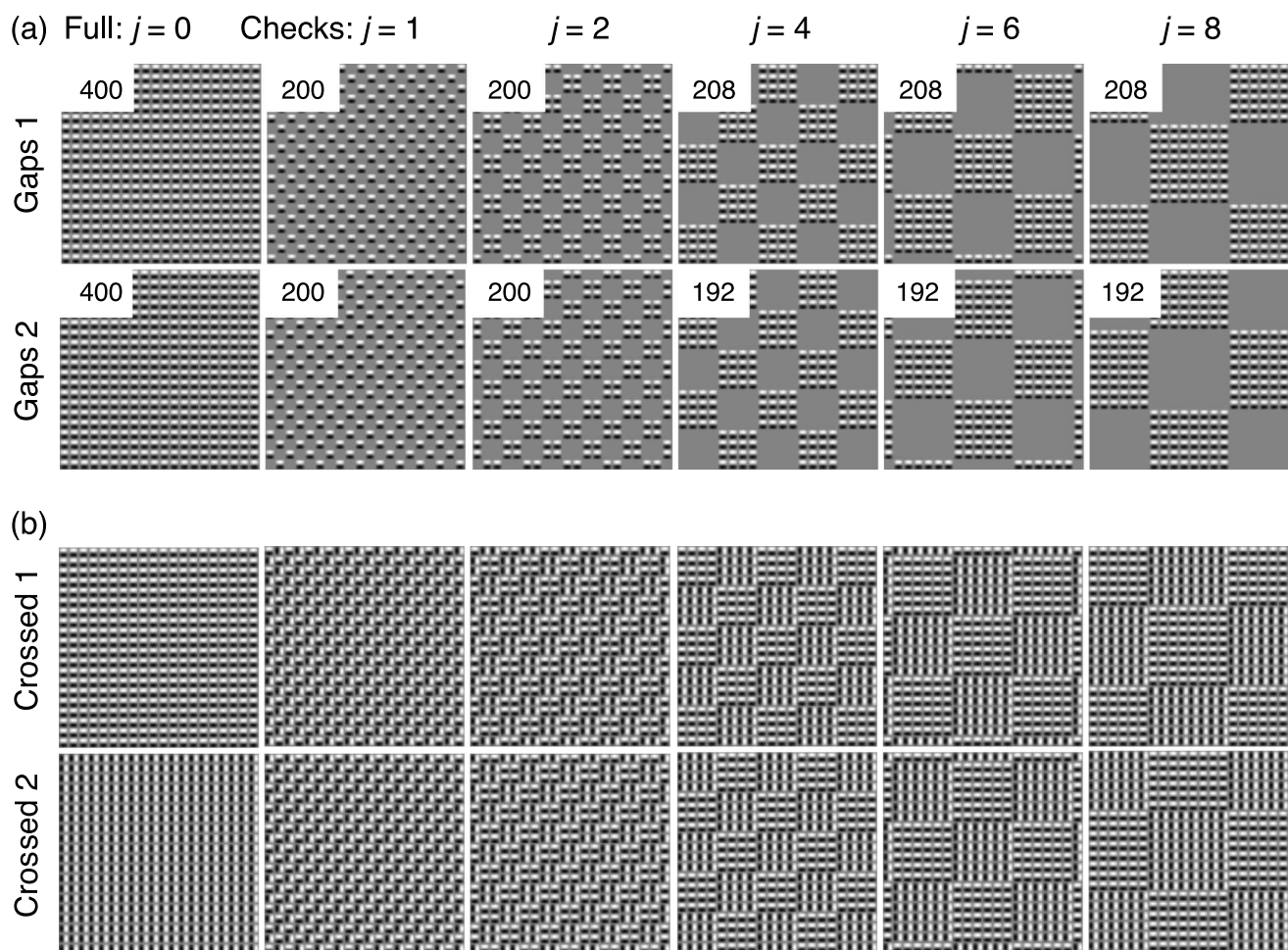
Figure 3. Battenberg stimuli used in the two experiments made from the micro-patterns in Figure 2a. (The stimuli are named after a distinctive cake that was made for the wedding between Prince Louis of Battenberg and Queen Victoria's granddaughter. The cake contains large yellow and pink checks of sponge wrapped in a marzipan casing and is available at supermarkets and corner shops throughout the UK.) (a) The stimuli used in the "gaps" experiment, indexed by j. The numerical insets indicate the number of micro-patterns. For $j > 0$, the number of micro-patterns is approximately constant. Note that the full stimuli ($j = 0$) contain the same number of micro-patterns as the sum of the complementary pairs (Gap 1 and Gap 2) of each of the other patterns. (b) The stimuli used in the "crossed" experiment. The only difference between the two experiments was that in the "crossed" experiment, the blank regions of the stimuli from the "gaps" experiment were filled with micro-patterns with orthogonal orientation. Note that the stimuli ($j = 0$ to 8) in this experiment have identical contrast energy to each other.

above, is the *assumption* that the model is able to operate in selective pooling mode for gratings (Summers & Meese, 2007) but not for Battenbergs. Finally, note that if the ideal summation model in Figure 1d were also unable to operate in selective pooling mode, it would revert to the linear summation model of Figure 1b.

Methods

All observers were given several sessions of practice before formal data collection began and wore their normal optical correction. Contrast detection thresholds (75% correct, estimated by probit analysis) were measured using interleaved staircases and a two-interval forced-choice (2IFC) technique where observers had to indicate which of two temporal intervals contained a target. Auditory feedback was provided to indicate correctness of response. The stimulus duration was 100 ms and the duration between the 2IFC intervals was 400 ms. For each observer, data were averaged from between 4 and 8 runs, each based on 100 trials per condition (i.e., average threshold estimates were based on 400 to 800 trials). When the standard error determined by probit analysis for a single threshold estimate was >3 dB, the data were discarded and the conditions were re-run. In each experimental session, stimuli were interleaved trial by trial from both rows of either Figure 3a or Figure 3b and from either the odd or even numbered columns to produce manageable session

lengths of about 20 minutes (6 conditions). Sessions were alternated between the odd and even columns.

Stimuli (Figure 3) were displayed on luminance calibrated raster monitors (120 Hz; Eizo F553M and Eizo 6600M) using a Cambridge Research Systems VSG2/5 and average mean luminance of 58 cd/m^2. The experiments were controlled by a PC. Observers were the author (TSM) and four undergraduate optometry students (OS, KM, MN, and IM) who completed the study as part of their course requirement. The experiments were performed in a darkened room and with the aid of a chin and headrest at a viewing distance of 72 cm.

Results

Spatial summation: The "gaps" experiment

Figure 4 shows contrast detection thresholds for the stimuli in Figure 3a. These have been normalized to indicate summation ratios between the full stimulus ($j = 0$)
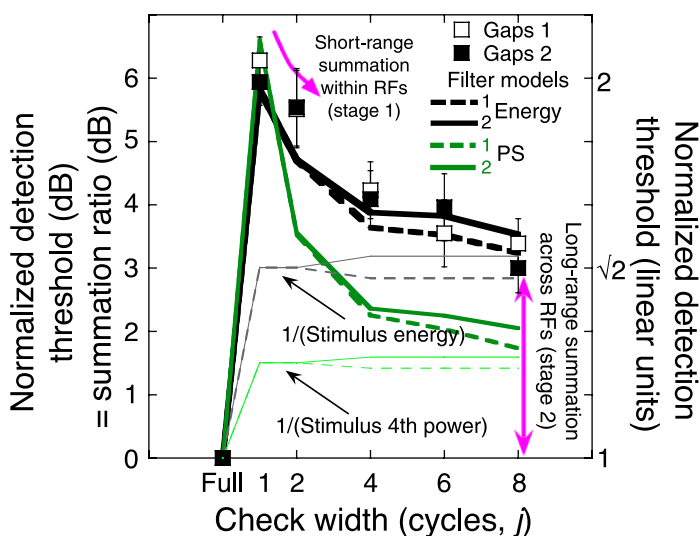

Figure 4. Contrast detection thresholds for the "gaps" experiment. Results are averaged across five observers (error bars are ±1*SE* of the means in dB). Detection thresholds are normalized to those for the "full" stimulus (Figure 3a, far left). This gives the SR for each of the check patterns verses the "full" stimulus. The thick curves are predictions for the filter models (based on Figure 1a and mandatory pooling in Figure 1e) for the stimuli in the top and bottom rows of Figure 3a (dashed and solid curves, respectively). The thresholds predicted by image contrast measures for energy (second power) and fourth power are shown by the thin curves. Note that the slight differences between the curves for the gap 1 and gap 2 stimulus series derive from the slightly different numbers of micro-patterns contained in the stimuli (see Figure 3a). The pink arrows highlight the effects of short- and long-range summation at Stages 1 and 2 in the main (three-stage) filter model (Figure 6). RF: receptive fields (i.e., filter elements).
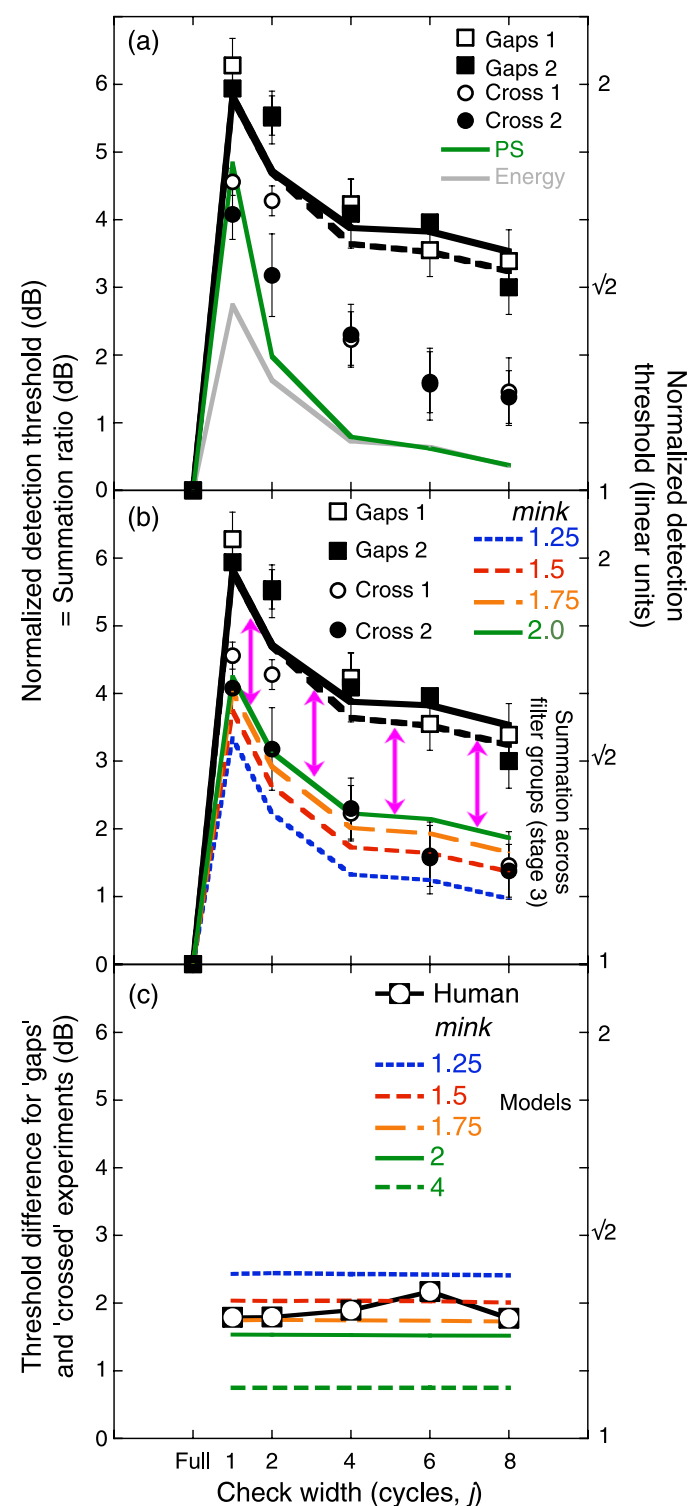
and each of the patchy stimuli, thereby illustrating the benefit of filling the gaps in the patchy stimuli with additional micro-patterns. For example, when this was done for the first pair of "check" stimuli ($j = 1$; threshold ~6 dB), detection thresholds halved ($j = 0$; threshold = 0 dB), consistent with full linear summation (Figure 1b). However, this does not imply that linear summation extends over the entire signal region; it could be that behavior is determined by summation only between neighboring micro-patterns. The spatial extent of summation can be assessed by grouping the micro-patterns in blocks of increasing size, as in the stimulus sequence in Figure 3a. By the time the central check region (j) is 6 or 8 micro-patterns square (Figure 4, far right), the benefit of the extra micro-patterns in the "full" stimulus has declined but is still ≥3 dB. Thus, between $j = 1$ and $j = 8$, summation falls from perfectly linear to quadratic. This transition is discussed later, but assuming that aggregation is contiguous and because sensitivity to the full stimulus is at least 3 dB ($\sqrt{2}$) greater than it is to each of the other stimuli in Figure 3a, the implication is that quadratic summation extends over at least twice as many micro-patterns (Figure 1e) as the width (or height) of the largest cluster in the stimuli on the far right ($j = 8$). That is, at least 16 ($2j$) micro-patterns, or 16 grating cycles: much more extensive than the orthodox view of contrast detection (Anderson & Burr, 1991; Bonneh & Sagi, 1998; Carney et al., 2000; Foley et al., 2007; Meese et al., 2005; Robson & Graham, 1981; Rovamo, Luntinen, & Näsänen, 1993), although within the range indicated by cortical physiology (Pollen et al., 2002; Sclar, Maunsell, & Lennie, 1990; Von der Heydt, Peterhans, & Dursteler, 1992).

To provide formal assessment of the results, I developed a computational model including retinal inhomogeneity (Pointer & Hess, 1989) (Figure 2b) and the combined architectures in Figures 1b and 1e, where short-range linear summation of contrast (Figure 1b) occurs within oriented filter elements (Figures 2c and 2d) followed by squaring and a stage of long-range linear summation across the entire signal region and across two phases of filter (Figure 1c) (Appendix C; see also Figure 6). Predictions for the two stimulus series (Figure 4, thick black curves, with no free parameters) are very good—much better than for the competing PS model (Figure 4, medium green curves), which was implemented using the same filtering and the same retinal inhomogeneity but followed by Minkowski summation with *mink* = 4 (Appendix D). Note that the combination of spatial filtering and boundary effects (within the stimulus) mean that both filter models predict more summation than expected from second (energy)- or fourth-power metrics applied to the raw contrast images (Appendix E) (compare thin and thick curves in Figure 4). Note also that for summation studies such as this one—where the analysis is restricted to just a single criterion level of performance—the energy model is identical to Minkowski summation with *mink* = 2 (see The energy model is equivalent to Minkowski

summation (with *mink* = 2) at a single level of perfor- mance section).

Spatial segmentation: The "crossed" experiment

The "gaps" experiment reveals long-range summation, similar to an energy detection process. But is this process

Figure 5. Results from the "crossed" experiment. (a) Contrast detection thresholds for the "crossed" experiment (circles) with those replotted from the "gaps" experiment (squares). The lower two curves indicate predictions for the "crossed" experiment for each of the two filter models assuming indiscriminate summation over area and orientation. (b) The same as in panel a but for Minkowski summation across orthogonal filters following orientation selective long-range summation (i.e., predictions for the three-stage model; Figure 6). The RMS error of the model predictions with the single free parameter, *mink* = 1.75, is 0.5 dB. The pink arrows highlight the effects of cross-group summation at Stage 3 in the three-stage model (Figure 6). (c) Mean differences (in dB) between the results and model predictions for the two experiments. Note that predictions for *mink* = 4 are excluded from panel b for clarity.

completely general, aggregating contrast over the entire image, or is it selective for particular local features such as orientation? In general, we might reason that if long-range summation is indiscriminate then, to a first approximation, sensitivity should remain the same (0 dB) when half of the micro-patterns in the full stimulus are rotated by 90°, as for the stimuli in Figure 3b. Indeed, the contrast energy is identical for each of the stimuli in this figure. However, spatial filtering complicates matters a little because of short-range summation effects at the boundaries of the micro-patterns, and this means that filter models are needed to make detailed predictions. These are shown in Figure 5a for PS (green curve) and energy (gray curve) (see Summation across orientation section). Neither of these anticipated the results of the "crossed" experiment (circles in Figure 5a), where performance was consistently worse than they predicted. This implies that, contrary to the orthodox view (Meese & Williams, 2000; Robson & Graham, 1981; Watson & Ahumada, 2005), vision does not use a global summation process (neither PS nor energy) that pools indiscriminately over space and orientation.

In fact, performance in the "crossed" experiment (circles) varied with stimulus configuration in a similar way to the first experiment (squares), although overall sensitivity was slightly higher. This was quite well predicted by Minkowski summation of signal responses from orthogonally oriented filters following long-range spatial summation within each orientation band (the three-stage model; Minkowski summation between orthogonal energy mechanisms: The three-stage model section; Figure 6). The average difference between the thresholds in the two experiments (for *j* > 0 in Figure 3) was 1.88 dB (Figure 5c). This represents the average summation between the orthogonal signals and is very similar to previous estimates for superimposed stimulus pairs where components have differed widely in orientation (Carney et al., 2000; Georgeson & Shackleton, 1994) and spatial frequency (Carney et al., 2000; Graham & Nachmias, 1971). Plotting the results this way suggests a Minkowski

Stage 1
Spatial filtering:
Short-range
spatial summation

Stage 2
Selectivity:
Segmentation and
long-range
spatial summation

Stage 3
Assimilation:
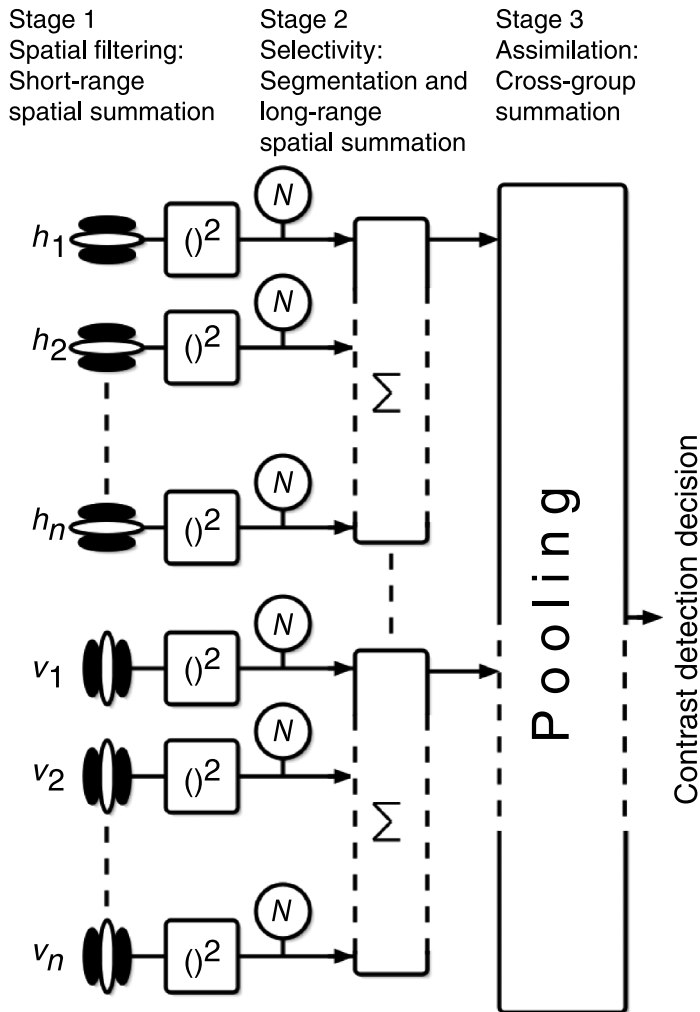Cross-group
summation

Figure 6. A new model of contrast summation and detection involving a three-stage hierarchy. Stage 1 involves linear spatial filtering, which performs mandatory short-range spatial summation of signal contrast within each filter element (receptive field) (see Figure 1b). Noise at this stage is insignificant (and not shown) relative to the performance limiting noise at the next stage. For simplicity, filter elements are shown for only one phase (see A deterministic implementation of the combination model section). At Stage 2, signals are summed over area following nonlinear (square-law) transduction of the contrast response. Area summation takes place within each of one or more groups of filter elements, permitting representations of multiple textures or contours (here, a pair of orthogonal orientations). The figure depicts a flexible long-range summation mechanism for each group, although selective pooling might be achieved using multiple hard-wired mechanisms instead (see text for details). Stage 3 pools across the filter groups from Stage 2 and the output forms the decision variable. The only free parameter in the model is *mink*, which sets the strength of cross-orientation summation at Stage 3. Note that retinal inhomogeneity is omitted from the figure for simplicity but is placed at the far left in the model.

exponent (*mink*) of about 1.75 (thin curves, Figure 5c) for summation across orthogonal filter orientations.

Discussion

The principal result is that Battenberg stimuli produce quadratic summation over a much greater area than might have been expected from earlier experiments using gratings. Looking back to Figure 1, we see that three of the five models are able to achieve this: energy summation (Figure 1c), ideal summation (Figure 1d), and the combination model (Figure 1e). However, only the combination model is able to also achieve the result with gratings (Meese & Summers, 2007; Summers & Meese, 2007).

Overall, the experiments point to three stages of summation for simple luminance contrast textures. These are (1) short-range linear summation within oriented filter elements (classical receptive fields) that segment the textures, (2) long-range linear summation across filter elements (following response nonlinearity) for coherent textures, and (3) a final stage of pooling across orthogonal filter groups that delivers the psychophysical decision variable. This hierarchy of neuronal convergence is schematized in Figure 6 and the details of each stage are discussed below.

The properties of short-range spatial summation (Stage 1)

Linear summation of luminance contrast suggests a convolution process that can be thought of as linear spatial filtering. This is a well-established property of spatial vision and the model filters here were identical to those used in a closely related model of a conventional area summation experiment in which the target area grew in proportion to the square of the stimulus diameter (Meese & Summers, 2007). Their spatial frequency and orientation bandwidths (1.6 octaves; ±25°) are similar to other psychophysical estimates (Foley et al., 2007; Kersten, 1984; Watson & Ahumada, 2005) and are consistent with cells in primary visual cortex (De Valois, Albrecht, & Thorell, 1982), but how critical are those parameters here? When the model bandwidths are made broader (and the receptive fields become smaller), the spatial filtering becomes less relevant (not shown) and the model predictions approach those made by the power transforms on the raw luminance images in Figure 4. However, the large differences between the data and those predictions, towards the left of Figure 4, imply that the filters do impose their presence and are inconsistent with such broad bandwidths. The filter model also fails when the bandwidths are made much narrower (and the number of

lobes in the receptive field increases) because this raises the right hand limb of the predictions (Figure 4) above the data (not shown). Thus, the spatial filters assumed here have both external validity and consistency with the results. Nevertheless, the precise parameter values are not critical; allowing the filter parameters to be free in the model fitting produced a slight change in their values (1.4 octaves; ±20°) and a slight improvement in the fit. This detailed analysis of spatial filter bandwidths as well as their relation to the model's transducer exponent is presented in Appendix G.

Long-range summation (Stage 2)

The "crossed" experiment suggests that long-range summation is selective for orientation, quite different from the usual interpretation of PS. Furthermore, the strength of long-range summation is greater than that which is usually associated with PS (Appendix H). All this points to deterministic physiological convergence such as that depicted by Stage 2 in Figure 6. Note that linear summation takes place at both Stage 1 and Stage 2 in the model, but the benefit of summation is less at the second stage owing to the intervening square-law contrast nonlinearity. Thus, as the clusters of micro-patterns fall outside the short range of individual filter elements (Stage 1), summation declines from about a perfect factor of 2 (Figure 1b; Figure 4, $j = 1$) to about a factor of $\sqrt{2}$, consistent with long-range detection by the energy model (Figure 1c; Figure 4, $j = 8$).

Selective pooling (Stage 2)

The details of selective pooling were not seen for the experiments here because the results were entirely consistent with simple mandatory pooling over all of the co-oriented filter elements at Stage 1. However, for the model to generalize beyond these experiments, greater flexibility is needed. For example, previous studies have investigated the effects of varying grating diameter on contrast sensitivity and found fourth-root summation (SR = 1.5 dB), consistent with spatial PS (see Introduction section). But this raises a problem because the PS model is rejected by the analysis here. Conveniently, the new model (Figure 6) has the flexibility to accommodate the fourth-root results by selective pooling for various n (Meese & Summers, 2007) (Figure 1e). In general, this is a good strategy for detecting a range of stimulus sizes. For example, when the target is small, not only is spatial resolution preserved (i.e., the observer is able to access individual filter elements to help identify individual blades of grass) but the selectivity prevents the aggregation of noise from irrelevant filter elements that would degrade the signal-to-noise ratio. In contrast, when the target is large, the detection process benefits from the aggregation of local signals.

For simplicity, Figure 6 depicts selective pooling by a mechanism at Stage 2 that is able to select its range of inputs. However, an equivalent arrangement involves multiple hard-wired mechanisms at this stage, each summing over different ranges (various n), and the appropriate selection by the observer of the relevant summing mechanism. Each of these methods is easily achieved with the assumption of labeled lines (Watson & Robson, 1981). However, the observer does not always have the information necessary to make the appropriate selection—as in the case where stimulus conditions are interleaved from trial to trial—but this appears to have little affect on the form of area summation (Meese et al., 2005). One explanation for this involves nonlinear pooling (e.g., either fourth-root summation (Tyler & Chen, 2000) or a MAX operation (Summers & Meese, 2007)) over multiple long-range mechanisms that sample the continuum of n. This type of scheme can be arranged to produce the summation behavior of selective pooling, even when the observer does not have complete knowledge of the stimulus (Tyler & Chen, 2000) (see also Appendix A).

Further properties of long-range spatial summation (Stage 2)

As stated in the Results section, quadratic summation extends over at least 16 micro-patterns, implying that long-range summation also extends over this range. However, I present this figure with some caution. For simplicity in the model, long-range summation was performed over the entire stimulus region. This worked well (Figures 4 and 5), but the experiments here did not determine the upper range for long-range summation nor the number of long-range mechanisms involved. It is plausible that several long-range mechanisms are scattered over the visual field and that performance depends on PS (or other nonlinear combination) between them (e.g., at Stage 3) (Syväjärvi, Näsänen, & Rovamo, 1999). In that case, a system involving mechanisms with a somewhat shorter range of long-range summation at Stage 2 might be supported by the results.

Another issue is the shape of the summation region. For simplicity, the model sums evenly over the two spatial image dimensions, but the results do not rule out the possibility that long-range summation is elongated along the direction of the contours (Meese & Hess, 2007), orthogonal to the contours, or a combination of the two. Thus, in principle, the 16 micro-patterns referred to above might be arranged in rows (snakes), columns (ladders), or two-dimensional arrays containing 16×16 micro-patterns.

An important question raised by the work here is as follows. Why can the visual system perform selective summation for conventional gratings but not for Battenbergs (Figure 7)? It seems likely that this has to do with the peculiar shape of the signal regions in Battenbergs. For example, one possibility is that the visual system is limited to summing over smooth contiguous (amoeboid)
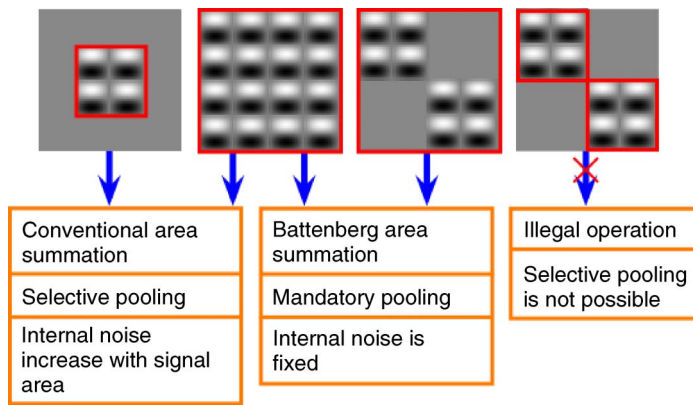
Figure 7. Summation regions (red squares) for the combination model. The internal noise is proportional to the square root of the areas enclosed by the red squares. For conventional area summation experiments where the area of the signal increases with stimulus size, internal noise also increases with signal area. The combination of nonlinear contrast transduction (C^p) and noise summation results in a fourth-root summation rule when $p = 2$. For Battenberg stimuli, summation cannot be restricted to the signal area but performance does benefit from summing over the entire stimulus region (from Equation 2). Because noise is not a factor for Battenberg summation, the level of long-range summation is affected only by nonlinear contrast transduction and follows a square root rule for $p = 2$. For simplicity of presentation, the operations are shown over smaller stimulus regions than those thought to be used in the experiments.

regions of the retina. This would happen if the summing templates were pre-wired, for example (see also the previous subsection and Appendix A). With this single restriction and a need to maximize the SNR (Equation 2), the system would operate in selective pooling mode for gratings (Meese & Summers, 2007; Summers & Meese, 2007), but mandatory pooling model for Battenbergs (Figure 1e), just as the results require. Clearly, this is an important topic for future research.

Finally, a comparison of the results from the two experiments (Figure 5) indicates that orthogonal orientations did not benefit from the long-range summation process (see also Appendix G), suggesting processes of texture segmentation and grouping (Bergen & Adelson, 1988; Graham & Sutter, 1998; Grossberg & Mingolla, 1985; Victor & Conte, 2005). This is achieved in the model (Figure 6) by performing long-range summation only within orientation bands, but more flexible or dynamic arrangements (e.g., Gestalt-type grouping rules) are possible. For example, it is plausible that long-range summation extends over smooth variations of local features (the Gestalt law of "good continuation") to perform contour and form integration (Field, Hayes, & Hess, 1993; Kingdom & Prins, 2009; Wilson & Wilkinson, 1998; Wilson, Wilkinson, & Habak, 1998) and more general texture processing (Grossberg & Mingolla, 1985; Motoyoshi & Kingdom, 2004; Motoyoshi & Nishida,

2004), such as that involved in pictorial depth cues (Li & Zaidi, 2000; Meese & Holmes, 2004).

Further detailed investigation is required to more fully understand each of the factors above.

Summation across textures and filter groups (Stage 3)

The results here show that (i) contrast detection improves when orthogonal micro-patterns are placed in the gaps of the stimuli used in the first experiment (Figure 5) but (ii) the level of improvement is less than that achieved by filling the gaps with micro-patterns of the same orientation (Figures 4 and 5). The first result implies that some form of pooling must exist across orthogonal orientations, and the second implies that it is not simply a consequence of indiscriminate long-range summation at Stage 2 (see also Appendix G). Thus, Stage 3 pools across the different groups of filter elements that emerge at Stage 2 (in this case, horizontal and vertical), but how should this final stage of summation be interpreted?

One possibility is that Stage 3 involves a process of PS. However, the estimate of the summation exponent at Stage 3 ($mink \approx 1.75$; Figure 5c) is much too low for this. Most contemporary models of PS would put this around $mink = 4$ (Tyler & Chen, 2000) (see also Appendix H). Even older models that assume the summation exponent can be estimated directly from the slope of the psychometric function (Quick, 1974; Sachs, Nachmias, & Robson, 1971) would fail. For example, for TSM I have estimated the Weibull slope of the psychometric function for Battenbergs to be $\beta = 4.06$, no different from those for gratings (Mayer & Tyler, 1986; Meese & Williams, 2000) and much higher than the $mink = \beta = 1.75$, that is needed.

Thus, the work here not only challenges the orthodox position on PS across area but also PS across feature dimensions such as orientation. Perhaps this is surprising because the levels of summation between the orthogonal elements here (Figure 5c) are very similar to those found in other studies using more conventional grating stimuli: they are all close to a fourth-root prediction (where SR = 1.5 dB), consistent with PS. However, the canonical model for PS involves a linear transducer (Tyler & Chen, 2000). When this is replaced with the squaring transducer implied by the "gaps" experiment, PS drops to about SR = 0.75 dB owing to the cascade of nonlinearities—markedly less summation than is found in the experiments (compare data and lower dashed green line in Figure 5c). Thus, whatever the process at Stage 3, its limit on summation is less severe than PS. Motoyoshi and Nishida (2004) came to a similar conclusion in a study on the segregation of suprathreshold textures.

An alternative possibility is that there is higher-order physiological summation, where $mink = 1.75$ (Figure 5c) represents further nonlinear contrast transduction prior to linear summation at Stage 3. However, there are problems

with this since the model's limiting noise is at Stage 2, placing it before the additional transducer. According to Birdsall's theorem (Klein & Levi, 2009; Lasley & Cohn, 1981), early limiting noise will linearize the effects of subsequent nonlinearities on the d' psychometric function and computer simulations have shown that this diminishes the transducer's effect on summation (Meese & Summers, 2009).

Yet another possibility is that observers performed ideal summation (Figure 1d) across the noisy outputs of the Stage 2 filter groups. This predicts *mink* = 2, close to the estimate from the results (Figure 5c). Why the estimate of *mink* here is a little less than this (1.75) is not clear, but it might be expected if a further source of additive (though not dominant) noise were to be injected at a later stage. Furthermore, whether pooling at Stage 3 is an explicit part of a higher-order visual code (e.g., for complex "gingham" textures) or represents a more general-purpose decision strategy by the observer is also unclear. Nonetheless, whatever the details of Stage 3, the experiments and analyses here suggest a strict sequence to the summation stages (Figure 6) owing to the relative sizes of the effects. Short-range summation (Stage 1 filtering; ~6 dB) precedes long-range summation (Stage 2 aggregation; ~3 dB), and this precedes summation across filter groups (Stage 3 pooling; ~1.9 dB).

Energy models

Energy models have been championed with much conviction in the past (Bergen & Adelson, 1988; Manahilov et al., 2001; Morrone & Burr, 1988; Rovamo et al., 1993; Watson & Ahumada, 2005; Watson et al., 1983), although they have received little direct empirical support when testing over extended regions in the central visual field at threshold (Campbell & Green, 1965; Howell & Hess, 1978; Meese & Holmes, 2004; Meese, Hess, & Williams, 2005; Meese & Williams, 2000; Kukkonen et al., 1993; Watson et al., 1983). Indeed, it is clear from Figure 4 that contrast sensitivity is not characterized by a direct measure of overall stimulus energy. However, the results do provide good evidence for spatially extensive energy detection following oriented spatial filtering. As alluded above, previous failures to reveal this process can be explained by the accompanying spatial aggregation of noise that caused energy detection to masquerade as PS (Figures 1a and 1e).

Nonlinear contrast transduction and Birdsall's theorem

A serious objection to accelerating transducer models (including the energy model here) is that the slope of the empirical psychometric function is too steep when contrast sensitivity is measured in the presence of external noise (Lu & Dosher, 2008; although see also Kersten, 1984). According to Birdsall's theorem, if the external noise is the performance-limiting noise, then this should "linearize" the

transducer and the psychometric function should be shallow (a d' slope of unity) (Klein & Levi, 2009; Lasley & Cohn, 1981). So how could the accelerating contrast nonlinearity—central to the model here—seemingly outmaneuver Birdsall's theorem? There are several possibilities (Klein & Levi, 2009; Lu & Dosher, 2008), but one is "distraction." The psychophysical performance of a distracted observer is similar to an uncertain observer—each produce steep psychometric functions (Kontsevich & Tyler, 1999). Thus, the linearizing effects of Birdsall's theorem would not be observed if the external noise pattern also served to distract the observer's attention away from the relevant target mechanisms. In other words, the usual objection to nonlinear contrast transduction need not apply if the observer is distracted by external noise. See also Burgess and Colborne (1988) and more recent papers by Lu and Dosher (2008) and Klein and Levi (2009).

Spatial summation in noise

One experimental approach related to that here has been to try and clamp the total noise level by swamping it with high contrast external noise. This was done by Kersten (1984) who found little or no evidence for summation beyond a single grating cycle in dynamic large-field one-dimensional noise. Why long-range summation was not found in that study is unclear, but it is possible that Kersten's external noise interfered with Stage 2 in the model (Figure 6), providing counter-evidence for the presence of coherent textures and disabling the pooling process. Other factors that are probably also involved in noise-masking studies (including Kersten's), although have been often overlooked, are pedestal masking (Legge & Foley, 1980; Meese, 2004), dilution masking (Meese & Summers, 2007), surround suppression (Meese, 2004; Meese, Challinor, Summers, & Baker, 2009; Meese et al., 2005), and retinal inhomogeneity (Pointer & Hess, 1989; Foley et al., 2007). Furthermore, other groups have found different results from Kersten. Syväjärvi et al. (1999) used two-dimensional static white noise and found that area summation was very similar with and without external noise. Thus, a detailed picture of the relation between external noise and area summation is yet to be elucidated.

Second-order contrast detection

Although our interest throughout has been with first-order contrast detection (this is what we measured), our Battenberg stimuli are contrast modulated (second-order) stimuli when $j > 0$ (Figure 3). One view of second-order spatial vision is that second-order mechanisms pool information from first-order mechanisms (Henning, Hertz, & Broadbent, 1975). In fact, the contrast mechanisms that we propose can be thought of exactly this way: they are second-order mechanisms sensitive to the DC (0 c/deg)

component of contrast modulation. But could it be that other second-order mechanisms were used to detect the contrast *boundaries* in our Battenbergs? We cannot rule this out, although the low sensitivity to second-order modulation found in other experiments makes this seem unlikely (Schofield & Georgeson, 1999, 2003). Furthermore, if it were the contrast boundaries that were detected in our Battenbergs, this could not apply to the full stimulus ($j = 0$) where there are no contrast boundaries. Thus, the high level of summation that we find in our "gaps" experiment would still need to be explained.

Summation above threshold and contrast gain control

Despite the initial motivation of the present work regarding aggregation of visual texture (see Introduction section), the evidence here for extensive area summation of contrast poses a problem when the contrast is raised above detection threshold. For example, while it makes good ecological sense to sum broadly when signals are weak, a large sports field should not appear to have higher contrast than a small front lawn. In fact, a similar problem has been encountered in binocular vision. Summation of contrast between the eyes is typically substantial at threshold (>3 dB) (Meese, Georgeson, & Baker, 2006), but under normal viewing conditions, the world does not appear to be of higher contrast when viewed with two eyes instead of one. This ocularity invariance (Meese et al., 2006) is achieved by suppressive gain control within and between the eyes (Baker, Meese, & Georgeson, 2007; Ding & Sperling, 2006; Meese et al., 2006). An analogous process (area invariance) involving suppression across the visual field is presumably also involved in area summation (Meese et al., 2005; Meese & Summers, 2007; Sclar et al., 1990), providing a plausible solution (Meese & Summers, 2007) to the problem identified above.

Conclusions

The *Battenberg* stimuli developed here (Figure 3) provide a new method by which neuronal convergence can be assessed in human vision. Experiments using these stimuli shed a very different light on the processes of early spatial vision compared with the orthodox interpretation established in the early 1980s (Graham & Nachmias, 1971; Graham, Robson, & Nachmias, 1978; Robson & Graham, 1981; Sachs et al., 1971). In those studies, the performance benefit achieved by either increasing the area of a grating or adding further gratings at different orientations or spatial frequencies was attributed to a single unselective (dumb) process of PS. The work presented here suggests a more strategic arrangement: the long-range process of area

summation is more potent than once thought, consistent with a contrast transducer exponent of 2 and energy detection within—but not between—spatial textures. Nonetheless, when the aggregation of internal noise covaries with stimulus size—as I propose it does in conventional studies of area summation—the degree of area summation drops to more modest levels (Figure 1e), typical of those usually attributed to PS. The mechanistic hierarchy proposed here (Figure 6) offers valuable insights into the missing link between early sensory processes of contrast vision and later stages in which substantial neuronal convergence is needed to represent larger shapes, forms, structures, and surfaces. Therefore, understanding the details of stimulus selection by the long-range summation at Stage 2 is a top priority for future research.

Appendix A

Mathematical derivations of summation ratios for the models in Figure 1

With reference to Figure 1, let the standard deviation of each of the independent noise sources be $\sigma = 1$, and let the signal strengths (Michelson contrasts) on each signal line be C. I consider summation ratios (SR) where the sensitivity of each of the input lines is equal and where the number of signals is doubled from $n/2$ to n for any even n. I denote the contrasts in these cases as C_{half} and C_{full}, respectively, and refer to the conditions as the "half" and "full" conditions. The signal-to-noise ratio (SNR) at a given output is $signal/\sigma_{tot}$, where $signal$ is the response to $n/2$ or n signals as appropriate and σ_{tot} is the standard deviation of the noise on that output line. Without losing generality we can set the criterion SNR required for detection to unity. Thus, the summation ratio (SR) is given by: C_{half}/C_{full}, for SNR = 1. Note also, because noise variances sum, the standard deviation of the sum of n noise sources is $\sqrt{(n)}$.

For the ideal summation model and the combination model (when in selective mode), I first assume that the observer knows what signals are being presented and uses this information to sum over the appropriate inputs. This requirement is then relaxed.

Throughout the Appendices, contrast gain parameters are ignored for simplicity. This is safe since the expressions for contrast response are ultimately normalized (i.e., they are used to calculate summation ratios) and the effects of response gain cancel.

Probability summation (Figure 1a)

If the assumptions of high-threshold theory hold (negligible false-positive responses in 2IFC and a hard

detection threshold, only above which the system is in a "detect" state) and the noise has a Weibull distribution, then the psychometric function (percent correct as a function of contrast) is a Weibull function and its slope is given by the Weibull slope parameter β. For probability summation (PS), this is equal to the exponent (*mink*) in Minkowski Summation (Quick, 1974) (Equation 1 main body). Typical empirical estimates give $\beta \approx 4$ in contrast detection studies (e.g., see Meese & Williams, 2000), which results in a fourth-root rule where SR = $2^{\sim 1/4} \approx 1.2$, or 1.5 dB. This is also equivalent to the vector summation model of Quick (1974).

However, the assumptions behind a strict derivation of Minkowski summation (Equation 1 in main body) from PS have long been falsified (Nachmias, 1981), and contemporary treatments of PS use a MAX operator within a two-interval forced-choice (2IFC) signal detection framework. There is no simple derivation of PS within this framework and the reader is referred to Pelli (1985) and Tyler and Chen (2000) for details. However, for a linear transducer and reasonable assumptions about uncertainty and knowledge of the signal, Tyler and Chen showed that PS is well approximated by Minkowski summation with a Minkowski exponent of ≈ 4. Thus, contemporary treatment of PS is also equivalent to a two-signal summation ratio (SR) of ≈ 1.2, or ≈ 1.5 dB.

Note that in each of the schemes above, contrast transduction (i.e., the growth of the contrast response with signal contrast) is assumed to be linear, albeit with the involvement of subsequent threshold nonlinearity in the first derivation. If, on the other hand, contrast transduction is an accelerating nonlinearity (as evidence suggests: Meese & Summers, 2009), then the SR for PS will decrease with an increase in the contrast-response exponent in each scheme.

Mandatory linear summation (Figure 1b)

For the full condition we have: SNR = $nC_{full}/\sqrt{(n)} = 1$, which rearranges to give: $C_{full} = 1/\sqrt{(n)}$. For the half condition we have: $nC_{half}/2\sqrt{(n)} = 1$, which rearranges to give: $C_{half} = 2/\sqrt{(n)}$. Thus, SR = $C_{half}/C_{full} = 2 = 6$ dB.

Energy model (Figure 1c)

For the full condition, we have: SNR = $nC_{full}^2/\sqrt{(n)} = 1$, which rearranges to give: $C_{full} = 1/n^{1/4}$. For the half condition, we have: $nC_{half}^2/2\sqrt{(n)} = 1$, which rearranges to give: $C_{half} = \sqrt{(2)}/n^{1/4}$. Thus, SR = $C_{half}/C_{full} = \sqrt{2} = 3$ dB.

Ideal summation (Figure 1d)

For the full condition, we have: SNR = $nC_{full}/\sqrt{(n)} = 1$, which rearranges to give: $C_{full} = 1/\sqrt{(n)}$. For the half condition, we have: $nC_{half}/2\sqrt{(n/2)} = 1$, which rearranges to give: $C_{half} = \sqrt{(2)}/\sqrt{(n)}$. Thus, SR = $C_{half}/C_{full} = \sqrt{(2)} = 3$ dB.

This is the summation expected for an ideal observer and is identical to that predicted by the energy model. Strictly speaking, the ideal observer knows the stimulus exactly on each trial and detects it using a linear matched filter, a strategy depicted by Figure 1d and described above. However, as Tyler and Chen (2000) point out (pp. 3130–3131), the same level of summation (though not the same overall sensitivity) might be expected in less specific situations. Suppose that the observer has several mechanisms that sample the size (diameter) dimension of the target gratings but does not know the size of the target to be presented on each trial. In that case, if each mechanism is weighted by the reciprocal of its expected noise level then fourth-root (Minkowski) summation over the set of mechanisms will also produce the behavior of ideal summation (Tyler & Chen, 2000). Thus, the same summation ratio (3 dB) might be expected in the ideal framework (Figure 1d) regardless of whether the experimental trials are blocked or interleaved across stimulus conditions.

Combination model (Figure 1e)

When the model is in mandatory pooling mode it sums over all of its available inputs and is identical to the energy model. Thus, SR = $\sqrt{2} = 3$ dB (see above). When the model is in selective pooling mode, it operates in a similar way to the ideal summation model above. However, the model is not ideal (in the formal sense) because of the squaring nonlinearity. For the full condition, we have: SNR = $nC_{full}^2/\sqrt{(n)} = 1$, which rearranges to give: $C_{full} = 1/n^{1/4}$. For the half condition, we have: $nC_{half}^2/2\sqrt{(n/2)} = 1$, which rearranges to give: $C_{half} = 2^{1/4}/n^{1/4}$. Thus, SR = $C_{half}/C_{full} = 2^{1/4} = 1.2 = 1.5$ dB. This is identical to the PS model with *mink* = 4. Following similar analysis to Tyler and Chen above, it can be shown that this level of summation is to be expected (under reasonable assumptions) for both blocked and interleaved experimental designs (Summers & Meese, 2007).

Appendix B

Battenberg stimuli combat the effects of retinal inhomogeneity

Figure B1 shows the results (squares) and filter model predictions (thick black curves) from the "gaps" experiment from Figure 5. The thinner (red) curves are for the same model (see main body and Appendix C) but with retinal inhomogeneity removed. The thinnest (blue) curves are for the model with retinal inhomogeneity reinstated, but with twice the severity as in the main model (i.e.,
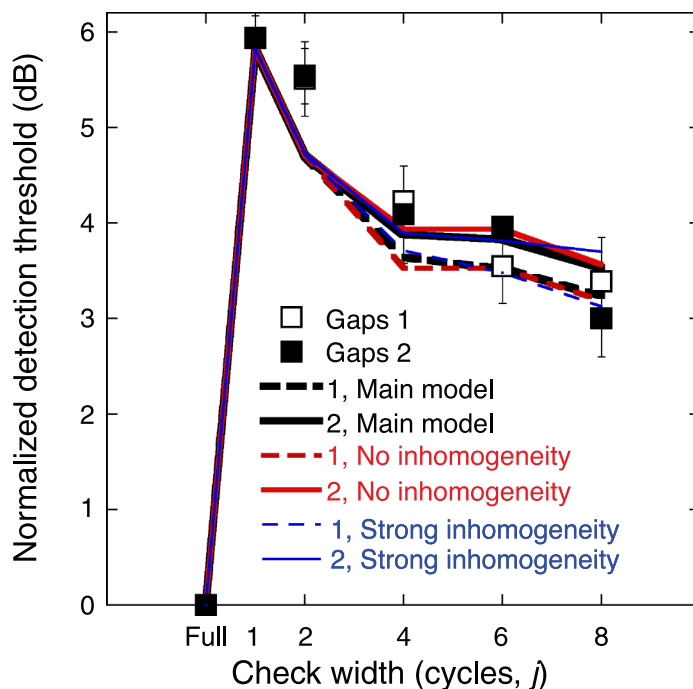
Figure B1. The effects of retinal inhomogeneity on performance of the filter models used here are negligible. See text for details.

sensitivity losses of 0.6 dB and 1.0 dB per cycle in the horizontal and vertical meridians respectively). Each of these adjustments had a negligible affect on the behavior of the model, indicating that (for the range of conditions used here) the stimulus design provides an effective counter-measure to the deleterious effects of retinal inhomogeneity (for further details, see Appendix C).

Appendix C

Main model implementation

A deterministic implementation of the combination model

Here I describe the "combination model" (Figure 1e) in conjunction with spatial filtering and retinal inhomogeneity. The basic model architecture and filtering is the same as that used by Meese and Summers (2007), although here I use a transducer exponent of 2 instead of 2.4. The exposition here is slightly different from that used in the supplementary material of Meese and Summers but is formally equivalent (with the exception of the value of the transducer exponent). For the stimuli here, I assume that the model operates in mandatory summation mode and is therefore equivalent to the energy model. (If, instead, the observer were able to sum selectively over only those stimulus locations that contained target micro-patterns,

then fourth-root summation would be expected and the model predictions would be inconsistent with the experimental results.)

Images had a contrast of 100% and were sampled with a resolution of 10 pixels per carrier cycle (though this was not critical) and multiplied by the attenuation surface shown in Figure 2b to simulate the effects of retinal inhomogeneity. This surface was derived from the empirical results of Pointer and Hess (1989) who found a sensitivity loss of 0.3 dB per carrier cycle in the horizontal meridian (x coordinate) and 0.5 dB per cycle in the vertical meridian (y coordinate). The attenuated image was then filtered by a pair of quadrature[1] log-Gabor filters (see Appendix C), with spatial frequency bandwidth of 1.6 octaves (full width at half-height) and orientation bandwidth of ±25° (half-width at half-height). The filters were matched to the spatial frequency (2.5 c/deg) and orientation (horizontal) of the micro-patterns in the "gaps" experiment and their outputs (2D arrays: Hsfilt and Hcfilt) were full-wave rectified. With this formulation, the basic filter responses did not represent a response to particular stimulus contrast, but the spatial distribution of responses of the filter elements (convolution kernels) across space for a particular stimulus. The contrast response was then derived by multiplying this pattern of filter responses by the Michelson contrast (0:1) of the stimulus. (This describes Stage 1 of the model in Figure 6.)

Each pixel value was raised to a power of 2.0 to represent the accelerating contrast-response nonlinearity of the mechanism at each pixel location. Unit-variance, zero-mean, Gaussian noise was added to each mechanism (pixel) and linear summation was then performed across the square stimulus region, which was identical for each stimulus, and across the quadrature filters. This gave a deterministic noise level of $\sqrt{(2n)}$ for each image, where n is the number of mechanisms (pixels) in the stimulus. Thus, the signal-to-noise ratio (SNR) for the summation process is given by:

$$\text{SNR} = \frac{\sum_{i=1:n} \left(|C_{\text{stim}} \times \text{Hsfilt}_i|^2 + |C_{\text{stim}} \times \text{cHfilt}_i|^2 \right)}{\sqrt{2n}},$$
(C1)

where C_{stim} is the Michelson contrast of the stimulus (in the range 0 to 1), and Hsfilt_i and Hcfilt_i are the sine- and cosine-phase filter responses at the ith of n pixels in the image, scaled to a response range of −1 to 1 for all stimuli. Assuming a criterion signal-to-noise ratio (SNR) of unity at detection threshold (this is arbitrary), it follows that:

$$1 = \frac{C_{\text{thresh}}^2 \sum_{i=1:n} \left(|\text{Hsfilt}_i|^2 + |\text{Hcfilt}_i|^2 \right)}{\sqrt{2n}},$$
(C2)

and solving for C_{thresh} gives:

$$C_{\text{thresh}} = \left[\frac{\sum_{i=1:n} \left(|\text{Hsfilt}_i|^2 + |\text{Hcfilt}_i|^2 \right)}{\sqrt{2n}} \right]^{-1/2}. \quad \text{(C3)}$$

As filter gain parameters have been omitted for simplicity, the contrast units of C_{thresh} are arbitrary. These arbitrary units cancel when the model predictions are normalized by the response to the full stimulus ($j = 0$ in Figure 3) to calculate a summation ratio. Note also that for the stimuli used here, the noise level in the model was constant and therefore irrelevant to the calculation of the model summation ratios. This describes the output of the summation box for the horizontal filter at Stage 2 in Figure 6.

The energy model is equivalent to Minkowski summation (with mink = 2) at a single level of performance

The equation for Minkowski summation over $i = 1$ to n linear contrast-response mechanisms r_i is:

$$\text{SNR} = \left[\sum_{i=1:n} |r_i|^{mink} \right]^{1/mink}. \quad \text{(C4)}$$

Let the SNR = 1 at a criterion level of performance (e.g., 75% correct). Then Equation C4 rearranges to give:

$$\sum_{i=1:n} |r_i|^{mink} = 1. \quad \text{(C5)}$$

The formulation of the energy model here is:

$$\text{SNR} = \frac{\sum_{i=1:n} |r_i|^2}{\sqrt{n\sigma^2}}, \quad \text{(C6)}$$

where σ is the standard deviation of independent zero-mean Gaussian noise added to the output of each contrast-response mechanism after squaring. At the same criterion level of performance as above, Equation C6 rearranges to give:

$$\sum_{i=1:n} |r_i|^2 = k, \quad \text{(C7)}$$

where k is the standard deviation of the total noise. This is absolved by an implicit contrast gain parameter in each model. Thus, there is an equivalence between energy summation (Equation C7) and Minkowski summation (Equation C5) when $mink = 2$. Note that this holds only for a single level of performance owing to the outer exponent ($1/mink$) in Equation C4. Put another way, the slope of the psychometric function (performance as a function of signal strength) is different for the two formulations. See also Meese and Summers (2009).

Log-Gabor filters

The choice of filter type in the model here is not particularly important. However, the log-Gabor filter has the convenient property that it has zero response to mean luminance for all filter phases (unlike a regular Gabor filter) and also has a modulation transfer function that resembles those of cortical cells. The two-dimensional modulation transfer function for the filters here is described by the following equation:

$$\text{logGab2D}(f, \theta) = \text{logGab1D}(f, \theta) \times \text{orthFunc}(f, \theta), \quad \text{(C8)}$$

where (f, θ) are polar coordinates in the Fourier plane (spatial frequency [in c/deg] and orientation [in degrees]). The first term on the right is defined as:

$$\text{logGab1D}(f, \theta) = \exp\left(\frac{-\left\{ \log_2\left(\frac{f|\cos(\theta - \theta_0)|}{f_0} \right) \right\}^2}{2(0.424\omega)^2} \right), \quad \text{(C9)}$$

where f_0 and θ_0 are the preferred spatial frequency and orientation of the filter and ω is the filter's spatial frequency bandwidth (full-width at half-height, in octaves). This is a Gaussian function on a log spatial frequency axis.

The second function is defined as follows:

$$\text{orthFunc}(f, \theta) = \exp\left(\frac{-\{f\sin(\theta - \theta_0)\}^2}{2\eta^2} \right), \quad \text{(C10)}$$

where

$$\eta = f_0 \sin(h) \sqrt{-\left(2.\ln\left\{ \frac{0.5}{\text{logGab1D}(f_0, \theta_0 + h)} \right\} \right)^{-1}}, \quad \text{(C11)}$$

where h is the filter's orientation bandwidth (half-width at half-height, in degrees).

Thus, the filters have modulation transfer functions that are the product of two one-dimensional functions in Fourier space. One is defined along a radial spatial frequency axis and would be a Gaussian shape if this axis

were logarithmic. The other is a Gaussian at right angles to this. The positions and dimensions of these filters (in the frequency domain) are defined entirely by four of the parameters above: f_0, θ_0, ω, and h. The phase of the filters is set directly in the phase spectrum of the Fourier domain.

Note that the filters here are different from the *polar separable* log Gabor filters that have been described elsewhere: http://www.csse.uwa.edu.au/~pk/research/matlabfns/PhaseCongruency/Docs/convexpl.html.

Appendix D

Probability summation and Minkowski summation

Minkowski summation with *mink* = 4 was used as a model of spatial probability summation (PS). Following the retinal inhomogeneity and spatial filtering described in Appendix C, this gives:

$$1 = \sum_{i=1:n} \left(|C_{\text{thresh}} \times \text{sfilt}_i|^{mink} + |C_{\text{thresh}} \times \text{cfilt}_i|^{mink} \right),$$

(D1)

at detection threshold. Solving for C_{thresh} we have:

$$C_{\text{thresh}} = \left[\sum_{i=1:n} \left(|\text{sfilt}_i|^{mink} + |\text{cfilt}_i|^{mink} \right) \right]^{-1/mink}.$$

(D2)

Appendix E

Second- and fourth-power metrics applied to the image

Second- and fourth-power metrics (p = 2 and 4, respectively) are calculated by summing over the i = 1 to m image pixels thus:

$$\text{Metric}(p) = \sum_{i=1:m} \left| \frac{L_i - L_0}{L_0} \right|^p,$$

(E1)

where L_i is local pixel intensity and L_0 is the mean pixel intensity across the image. For convenience, these metrics are normalized such that Metric(p) = 1, at detection threshold. This is equivalent to:

$$1 = C_{\text{thresh}}^p \sum_{i=1:m} |\text{image}_i|^p,$$

(E2)

where C_{thresh} is the Michelson contrast at detection threshold, and image is the luminance profile of the stimulus scaled to the range: −1:1. Solving for C_{thresh} gives:

$$C_{\text{thresh}} = \left(\sum_{i=1:m} |\text{image}_i|^p \right)^{-1/p}.$$

(E3)

Thus, Equation E3 solves for detection threshold in Figure 4 and is the normalized reciprocal of Metric(p). It is equivalent to spatial Minkowski summation across raw local contrasts, where p = *mink*.

Appendix F

A model with two filter orientations

Summation across orientation

Following retinal inhomogeneity, the images in Figure 3b were filtered by the horizontal filters described above and similar vertical filters (i.e., filters with preferred orientations of 90° and 0°). The 2D arrays of responses of the vertical filters to stimuli with a Michelson contrast of 100% are denoted Vsfilt and Vcfilt. These were combined with horizontal filter responses using Minkowski summation. As the summation region is constant, the noise level is also constant. This is absolved by the implicit gain parameters and for our purposes here can be safely dropped (though see below). Thus, following the developments above, we have:

$$C_{\text{thresh}} = \left[\sum_{i=1:n} \left(|\text{Hsfilt}_i|^{mink} + |\text{Hcfilt}_i|^{mink} \right. \right.$$
$$\left. \left. + |\text{Vsfilt}_i|^{mink} + |\text{Vcfilt}_i|^{mink} \right) \right]^{-1/mink}.$$

(F1)

This is the arrangement used for the two lower curves (green and gray) in Figure 5a, where *mink* = 2 for energy (see The energy model is equivalent to Minkowski summation (with *mink* = 2) at a single level of performance section) and *mink* = 4 for PS (Appendix A).

Minkowski summation between orthogonal energy mechanisms: The three-stage model

Strictly speaking, when the noise is placed before the summation process within each orientation band (as here; Figure 6), then it should not be dropped from expressions that combine signal and noise across multiple filters.

However, since some interpretations of Minkowski summation take the effects of noise into account, we drop the noise terms in the first instance, and return to them when interpreting the results and the meaning of *mink* in the Discussion section of the main report. Using the nomenclature above, the system response (resp) for Minkowski summation between vertical and horizontal energy filters following long-range spatial summation is given by:

$$
\text{resp} = \left[\left(C^2 \sum_{i=1:n} \left(|\text{Hsfilt}_i|^2 + |\text{Hcfilt}_i|^2 \right) \right)^{mink} \right.
$$
$$
\left. + \left(C^2 \sum_{i=1:n} \left(|\text{Vsfilt}_i|^2 + |\text{Vcfilt}_i|^2 \right) \right)^{mink} \right]^{1/mink}.
$$
(F2)

Setting resp = 1 at detection threshold, and solving for C_{thresh} we have:

$$
C_{\text{thresh}} = \left[\left(\sum_{i=1:n} \left(|\text{Hsfilt}_i|^2 + |\text{Hcfilt}_i|^2 \right) \right)^{mink} \right.
$$
$$
\left. + \left(\sum_{i=1:n} \left(|\text{Vsfilt}_i|^2 + |\text{Vcfilt}_i|^2 \right) \right)^{mink} \right]^{-1/(2mink)}.
$$
(F3)

This is the model used in Figures 5b and 5c, following retinal inhomogeneity and for various values of *mink*, which is the only free parameter. This describes the output of Stage 3 in Figure 6, where the model is operating in mandatory summation mode for each of the filter orientations.

Note the different placements of the parameter *mink* in Equations F1 and F3. In Equation F1, *mink* operates on the contrast response of each filter element, whereas in Equation F3, the exponent applied to each filter element is set to 2 (consistent with energy detection), and *mink* operates on the responses at each orientation after long-range summation within each of those orientation bands.

Appendix G

Narrower bandwidths and higher exponents

Alternative models

The three-stage model in the main report here (Equation F3) performed well using spatial filters with spatial frequency and orientation bandwidths of 1.6 octaves and ±25°, respectively (Table G1, Line 13). These bandwidths were not chosen to optimize the fit but were the same as those used in a previous study (Meese &

Summers, 2007). In fact, the best fit (Table G1, Line 14) was found using slightly narrower bandwidths of 1.4 octaves and ±20° (see Figure G2a), which also correspond exactly with the median estimates from single-cell physiology (De Valois et al., 1982; De Valois, Yund, & Hepler, 1982). Decreasing the filter bandwidths further causes the right hand limb of the model predictions to rise above the human data (not shown). However, increasing the contrast-response exponent can overcome this mismatch. Thus, to evaluate the effects of filter bandwidths for the conclusions here, I compared the results from both experiments with the more conservative two-stage model described by Equation F1 for various bandwidths and exponent values (*mink*). (This model involves only one stage of indiscriminate pooling beyond the initial filtering.) To illustrate the point, I began by arbitrarily halving the bandwidths (to 0.8 octaves and ±12.5°) and optimizing the exponent by fitting to the data from both experiments. This gave an exponent value of 2.9. With this arrangement, the model performed well and the results are shown in Figure G2b and Table G1 (Line 1).

I then performed a more thorough test by varying spatial frequency and orientation bandwidths in steps of 0.1 and ±1°, respectively, and the exponent (*mink*) in steps of 0.1. Local minima (nearest neighbor comparisons) in the three-dimensional error surface (expressed as the RMS error in dB) are shown in Table G1 (Lines 2–9). The best exponent always had a value of *mink* = 2.9. Essentially, Table G1 (Lines 2–9) describes a contour through spatial frequency/orientation space where the quality of the fit is almost constant. Put another way, Table G1 implies that a fairly extensive receptive field is needed to produce the level of spatial summation found in the experiments when only a single stage of pooling is used after spatial filtering (i.e., Equation G13 instead of Equation G15). This can be achieved by making the filter element long and thin (narrow orientation bandwidth), short and fat (narrow spatial frequency bandwidth), or anything along a continuum of combinations between the two. The nonlinear exponent must also be set higher than in the energy model but lower than for fourth-root summation (i.e., *mink* = 2.9).

The best alternative filter bandwidths are too narrow

The analysis above prompts the question of whether any of the filters in Table G1 (Lines 1 to 9) are reasonable estimates of psychophysical filter bandwidth. One of the most extensive studies relevant to this question is that performed by Foley et al. (2007). They measured detection thresholds for Gabor-type targets for a wide range of stimulus heights, widths, and phases. From their extensive analysis of the area summation functions, I have deduced the following absolute lower limits for filter bandwidths: 1.5 octaves for spatial frequency and approximately ±16° for orientation (Table G1, Line 10). However, most of their data suggested broader bandwidths

Line	Filter derivation and extra parameters	SF B/W (octaves)	Orient B/W (±deg)	Transducer exponent (*mink* in Equation F1 and fixed at 2 in Equation F3)	RMS error (dB)
1	Equation F1 and bandwidths = half those in main report	**0.8**	**12.5**	2.9	*0.383*
2	Optimized fit using Equation F1	0.4	23	2.9	*0.350*
3	"	0.5	15	2.9	*0.338*
4	"	0.6	12	2.9	*0.336*
5	"	0.7	11	2.9	*0.337*
6	"	0.8	10	2.9	*0.337*
7	"	1.0	9	2.9	*0.339*
8	"	1.5	8	2.9	*0.342*
9	"	5.9	7	2.9	*0.351*
10	Foley et al. narrow	**1.5**	**16**	—	—
11	Foley et al. broad	**4.9**	**30**	—	—
12	Foley et al. mid (Equation F3; *mink* = **1.75**)	**2.1**	**22**	**2**	*0.498*
13	Main model in main report (Equation F3; *mink* = 1.75)	**1.6**	**25**	**2**	*0.499*
14	Optimized fit using Equation F3 (*mink* = **1.75**)	1.4	20	**2**	*0.388*

Table G1. Filter bandwidths and figures of merit (RMS error in dB, set in italics) for comparisons between models and results from both experiments. Entries set in bold are fixed parameters. Entries in plain text (not bold or italics) are free parameters. Line 1: The bandwidths used in Figure G2b. They are half those used in the model in the main report. Lines 2–9: Bandwidths that produced local minima (nearest neighbor comparisons) in the RMS error surface comparing model (Equation F1) and data. Lines 10–12: Bandwidths estimated from Foley et al.'s (2007) study on area summation of contrast. Line 13: The default bandwidths used in the model (Equation F3) in the main part of the report. Line 14: The best fitting bandwidths for the three-stage model (Equation F3). Performance for this implementation of the model is shown in Figure G2a (B/W: bandwidth; Orient: orientation; RMS: root mean square; SF: spatial frequency).
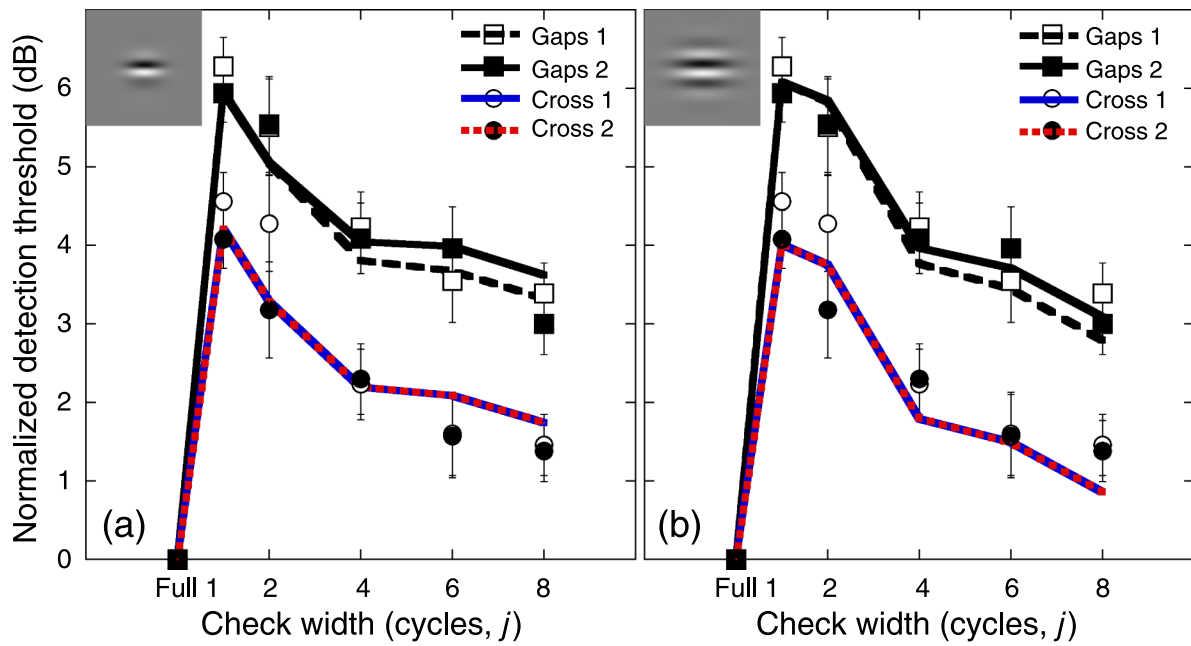

Figure G2. Alternative model fits. (a) The same model as in the main part of the report (Equation S15) but using filter bandwidths of 1.4 octaves and ±20° (see inset). These are the filters that produced the optimum fit to the gaps experiment with no other free parameters (Table G1, Line 14). (b) Example of using Equation G13 with bandwidths half of those in the main report. They are 0.8 octaves and ±12.5° (see inset) and *mink* = 2.9 (Table G1, Line 1). However, this filter element is too large (e.g., it has too many lobes) to be consistent with most other psychophysical estimates.

than these (see Table G1; Lines 10–12). In general, the estimates from Foley et al. are much broader than those derived in the alternative analysis above. In other words, if the stimuli here were detected by the large receptive fields implied in Table G1 (Lines 1–9), then those same mechanisms should have also shown up in Foley et al.'s summation experiment. That is, performance should have improved with Gabor target area more rapidly than was found in their experiment (over the initial range at least). This makes a model involving the combination of Equation F1, the filters in Table G1 (Lines 1 to 9) and an exponent of $mink = 2.9$ an unlikely candidate.

The bandwidths used in the main model here are about right

In contrast to above, the model bandwidths used in the main part of the report are within the bounds of those derived by Foley et al. (2007) and are fairly typical of those implied by single-cell physiology. Furthermore, the main model behavior here (Figure 5b) was almost identical (not shown) to that found when the filters were switched to the intermediate estimates (Table G1, Line 12) from the Foley et al. study. Finally, as noted above, the bandwidths that achieved the optimum fits for the "gaps" experiment in the three-stage model (Table G1, Line 14) were similar to those used in the main report and matched De Valois et al.'s estimates (De Valois, Albrecht, et al., 1982; De Valois, Yund, et al., 1982) from single-cell physiology exactly. These are within the range of orientation bandwidths derived from Foley et al. but just outside their range of spatial frequency bandwidths (Table G1, Lines 10 and 11).

Appendix H

PS and super PS

As outlined in Appendix A, typical interpretations of probability summation (PS) predict fourth-root summation—an SR of 1.5 dB. However, there are theoretical situations (Tyler & Chen, 2000) in which PS can achieve the same levels of summation as that predicted by Minkowski summation with $mink \approx 2$. This theoretical super PS occurs when the target region is doubled so as to fill the attention window—in the situation here, the set of Stage 1 filter elements monitored by the observer. In principle, this could be responsible for the long-range summation in the "gaps" experiment where the observer's attention window might have been matched to the full stimulus. However, studies that have measured the slope of the psychometric function (Meese & Summers, 2009; Meese & Williams, 2000; Robson & Graham, 1981) have consistently found it to be far too steep to be consistent with super PS, which requires a

linear transducer and d' psychometric slope of unity (equivalent to a Weibull slope parameter $\beta \sim 1.3$). A steep empirical slope might be attributed to uncertainty (Pelli, 1985) rather than a nonlinear transducer, but in either case, super PS would be abolished (Meese & Summers, 2009; Tyler & Chen, 2000). Alternatively, the steep slope of the psychometric function is consistent with a second-power (i.e., square-law) contrast transducer (plus a little uncertainty; Meese & Summers, 2009) such as that proposed here and elsewhere (Graham & Sutter, 1998; Klein & Levi, 2009; Lu & Dosher, 2008; Manahilov, Simpson, & McUlloch, 2001). In sum, it seems unlikely that the summation results in the "gaps" experiment here can be attributed to PS of the super PS variety or any other.

Another place where super PS should be considered is across orthogonal filters at Stage 3 of the model (Figure 6). However, the blocking of trials across the two experiments means that the uncertainty conditions that are needed for this interpretation (Meese & Summers, 2009; Tyler & Chen, 2000) are unlikely to have been met (i.e., it is unlikely that the observer monitored the additional mechanisms needed for the second experiment during the first). Thus, super PS also seems an unlikely candidate for summation at Stage 3 in the model.

Acknowledgments

Commercial relationships: none.
Corresponding author: Tim S. Meese.
Email: t.s.meese@aston.ac.uk.
Address: Aston Triangle, Birmingham, B4 7ET, UK.

Footnote

[1]The use of quadrature filters was not critical; very similar results are found using only sine or cosine phase filters.

References

Anderson, S. J., & Burr, D. C. (1991). Spatial summation properties of directionally selective mechanisms in human vision. *Journal of the Optical Society of America A, 8,* 1330–1339.

Baker, D. H., Meese, T. S., & Georgeson, M. A. (2007). Binocular interaction: Contrast matching and contrast discrimination are predicted by the same model. *Spatial Vision, 20,* 397–413.

Bergen, J. R., & Adelson, E. H. (1988). Early vision and texture perception. *Nature, 333,* 363–364.

Bonneh, Y., & Sagi, D. (1998). Effects of spatial configuration on contrast detection. *Vision Research, 38,* 3541–3553.

Burgess, A. E., & Colborne, B. (1988). Visual signal detection. IV. Observer inconsistency. *Journal of the Optical Society of America A, 5,* 617–627.

Campbell, F. W., & Green, D. G. (1965). Monocular versus binocular visual acuity. *Nature, 208,* 191–192.

Carney, T., Tyler, C. W., Watson, A. B., Makous, W., Beutter, B., Chen, C., Norcia, A. M., et al. (2000). Modelfest: Year one results and plans for future years. In B. E. Rogowitz & T. N. Pappas (Eds.), *Human Vision and Electronic Imaging: V. Proceedings of SPIE* (vol. 3959, pp. 140–151).

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

De Valois, R. L., Yund, E. W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research, 22,* 531–544.

Ding, J., & Sperling, G. (2006). A gain-control theory of binocular combination. *Proceedings of the National Academy of Sciences of the United States of America, 103,* 1141–1146.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual-system—Evidence for a local association field. *Vision Research, 33,* 173–193.

Foley, J. M., Varadharajan, S., Koh, C. C., & Farias, C. Q. (2007). Detection of Gabor patterns of different sizes, shapes, phases and eccentricities. *Vision Research, 47,* 85–107.

Georgeson, M. A., & Shackleton, T. M. (1994). Perceived contrast of gratings and plaids—Nonlinear summation across oriented filters. *Vision Research, 34,* 1061–1075.

Goris, R. L. T., Wagemans, J., & Wichmann, F. A. (2008). Modelling contrast discrimination data suggest both the pedestal effect and stochastic resonance to be caused by the same mechanism. *Journal of Vision, 8*(15):17, 1–21, http://www.journalofvision.org/content/8/15/17, doi:10.1167/8.15.17. [PubMed] [Article]

Graham, N., & Sutter, A. (1998). Spatial summation in simple (Fourier) and complex (non-Fourier) texture channels. *Vision Research, 38,* 231–257.

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

Graham, N., & Nachmias, J. (1971). Detection of grating patterns containing two spatial frequencies—A comparison of single-channel and multiple-channels models. *Vision Research, 11,* 251–259.

Graham, N., Robson, J. G., & Nachmias, J. (1978). Grating summation in fovea and periphery. *Vision Research, 18,* 815–825.

Green, M. G., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Robert E. Krieger Publishing Company.

Grossberg, S., & Mingolla, E. (1985). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception & Psychophysics, 38,* 141–171.

Henning, G. B., Hertz, B. G., & Broadbent, D. E. (1975). Some experiments bearing on the hypothesis that the visual system analyses spatial patterns in independent bands of spatial frequency. *Vision Research, 15,* 887–897.

Howell, E. R., & Hess, R. F. (1978). The functional area for summation to threshold for sinusoidal gratings. *Vision Research, 18,* 369–374.

Kersten, D. (1984). Spatial summation in visual noise. *Vision Research, 24,* 1977–1990.

Kingdom, F. A. A., & Keeble, D. R. T. (1996) A linear systems approach to the detection of both abrupt and smooth spatial variations in orientation defined textures. *Vision Research, 36,* 409–420.

Kingdom, F. A. A., & Prins, N. (2009) Texture-surround suppression of contour-shape coding in human vision. *NeuroReport, 20,* 5–8.

Klein, S. A., & Levi, D. M. (2009). Stochastic model for detection of signals in noise. *Journal of the Optical Society of America A, 26,* B110–B126.

Kontsevich, L. L., & Tyler, C. W. (1999). Distraction of attention and the slope of the psychometric function. *Journal of the Optical Society of America A, 16,* 217–222.

Kukkonen, H., Rovamo, J., Tiippana, K., & Näsänen, R. (1993). Michelson contrast, RMS contrast and energy of various spatial stimuli at threshold. *Vision Research, 33,* 1431–1436.

Lasley, D. J., & Cohn, T. E. (1981). Why luminance discrimination may be better than detection. *Vision Research, 21,* 273–278.

Legge, G., & Foley, J. (1980). Contrast masking in human-vision. *Journal of the Optical Society of America, 70,* 1458–1471.

Lennie, P. (1998). Single units and visual cortical organization. *Perception, 27,* 889–935.

Li, A., & Zaidi, Q. (2000). Perception of three-dimensional shape from texture is based on patterns of oriented energy. *Vision Research, 40,* 217–242.

Lu, Z. L., & Dosher, B. A. (2008). Characterizing observers using external noise and observer models: Assessing internal representations with external noise. *Psychological Review, 115,* 44–81.

Mayer, M. J., & Tyler, C. W. (1986). Invariance of the slope of the psychometric function with spatial summation. *Journal of the Optical Society of America A, 3,* 1166–1172.

Manahilov, V., Simpson, W. A., & McCulloch, D. L. (2001). Spatial summation of peripheral Gabor patches. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 18,* 273–282.

Meese, T. S. (2004). Area summation and masking. *Journal of Vision, 4*(10):8, 930–943, http://www.journalofvision.org/content/4/10/8, doi:10.1167/4.10.8. [PubMed] [Article]

Meese, T. S. (in preparation). Cross-eyed and spaced-out: Contrast is summed across eyes and area along the entire dipper function. *Journal of Vision.*

Meese, T. S., & Summers, R. J. (2007). Area summation in human vision at and above detection threshold. *Proceedings of the Royal Society of London B, Biological Sciences, 274,* 2891–2900.

Meese, T. S., Georgeson, M. A., & Baker, D. H. (2006). Binocular contrast vision at and above threshold. *Journal of Vision, 6*(11):7 1224–1243, http://www.journalofvision.org/content/6/11/7, doi:10.1167/6.11.7. [PubMed] [Article]

Meese, T. S., Hess, R. F., & Williams, C. B. (2005). Size matters, but not for everyone: Individual differences for contrast discrimination. *Journal of Vision, 5*(11):2, 928–947, http://www.journalofvision.org/content/5/11/2, doi:10.1167/5.11.2. [PubMed] [Article]

Meese, T. S., & Holmes, D. J. (2004). Performance data indicate summation for pictorial depth-cues in slanted surfaces. *Spatial Vision, 17,* 127–151.

Meese, T. S., & Summers, R. J. (2009). Neuronal convergence in early contrast vision: Binocular summation is followed by response nonlinearity and area summation. *Journal of Vision, 9*(4):7, 1–16, http://www.journalofvision.org/content/9/4/7, doi:10.1167/9.4.7. [PubMed] [Article]

Meese, T. S., & Williams, C. B. (2000). Probability summation for multiple patches of luminance modulation. *Vision Research, 40,* 2101–2113.

Meese, T. S., & Hess, R. F. (2007). Anisotropy for spatial summation of elongated patches of grating: A tale of two tails. *Vision Research, 47,* 1880–1892.

Meese, T. S., Challinor, K. L., Summers, R. J., & Baker, D. H. (2009). Suppression pathways saturate with contrast for parallel surrounds but not for superimposed cross-oriented masks. *Vision Research, 49,* 2927–2935.

Morrone, M., & Burr, D. (1988). Feature detection in human-vision—A phase-dependent energy-model. *Proceedings of the Royal Society of London B, Biological Sciences, 235,* 221–245.

Motoyoshi, I., & Kingdom, F. A. A. (2004). Differential roles of contrast polarity reveal two streams of second-order visual processing. *Vision Research, 47,* 2047–1054.

Motoyoshi, I., & Nishida, S. (2004). Cross-orientation summation in texture segregation. *Vision Research, 44,* 2567–2576.

Motoyoshi, I., Nishida, S., Sharan, L., & Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature, 447,* 206–209.

Nachmias, J. (1981). On the psychometric function for contrast detection. *Vision Research, 21,* 215–223.

Näsänen, R., Kukkonen, H., & Rovamo, J. (1994). Relationship between spatial integration and spatial spread of contrast energy in detection. *Vision Research, 34,* 949–954.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 2,* 1508–1532.

Pointer, J. S., & Hess, R. F. (1989). The contrast sensitivity gradient across the human visual-field—with emphasis on the low spatial-frequency range. *Vision Research, 29,* 1133–1151.

Pollen, D. A., Przybyszewski, A. W., Rubin, M. A., & Foote, W. (2002). Spatial receptive field organization of macaque V4 neurons. *Cerebral Cortex, 12,* 601–616.

Quick, R. F. (1974). A vector-magnitude model of contrast detection. *Biological Cybernetics, 16,* 65–67.

Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual-field. *Vision Research, 21,* 409–418.

Rovamo, J., Luntinen, O., & Näsänen, R. (1993). Modeling the dependence of contrast sensitivity on grating area and spatial-frequency. *Vision Research, 33,* 2773–2788.

Sachs, M. B., Nachmias, J., & Robson, J. G. (1971). Spatial-frequency channels in human vision. *Journal of the Optical Society of America, 61,* 1176.

Schofield, A. J., & Georgeson, M. A. (1999). Sensitivity to modulations of luminance and contrast in visual white noise: Separate mechanisms with similar behaviour. *Vision Research, 39,* 2697–2716.

Schofield, A. J., & Georgeson, M. A. (2003) Sensitivity to contrast modulation: The spatial frequency dependence of second-order vision. *Vision Research, 43,* 243–259.

Sclar, G., Maunsell, J. H., & Lennie, P. (1990). Coding of image contrast in central visual pathways of the macaque monkey. *Vision Research, 30,* 1–10.

Summers, R. J., & Meese, T. S. (2007). Area summation is linear but the contrast transducer is nonlinear: Models of summation and uncertainty and evidence from the psychometric function [ECVP abstract supplement]. *Perception, 36,* 5.

Syväjärvi, A., Näsänen, R., & Rovamo, J. (1999). Spatial integration of signal information in Gabor stimuli. *Ophthalmic & Physiological Optics: The Journal of the British College of Ophthalmic Opticians (Optometrists), 19,* 242–252.

Tyler, C. W., & Chen, C. C. (2000). Signal detection theory in the 2AFC paradigm: Attention, channel uncertainty and probability summation. *Vision Research, 40,* 3121–3144.

Victor, J. D., & Conte, M. M. (2005) Local processes and spatial pooling in texture and symmetry detection. *Vision Research, 45,* 1063–1073.

Von der Heydt, R., Peterhans, E., & Dursteler, M. R. (1992). Periodic-pattern-selective cells in monkey visual cortex. *Journal of Neuroscience, 12,* 1416–1434.

Watson, A., & Robson, J. (1981). Discrimination at threshold: Labelled detectors in human vision. *Vision Research, 21,* 1115–1122.

Watson, A. B. (1979). Probability summation over time. *Vision Research, 19,* 515–522.

Watson, A. B., & Ahumada, A. J., Jr. (2005). A standard model for foveal detection of spatial contrast. *Journal of Vision, 5*(9):6, 717–740, http://www.journalofvision.org/content/5/9/6, doi:10.1167/5.9.6. [PubMed] [Article]

Watson, A. B., Barlow, H. B., & Robson, J. G. (1983). What does the eye see best? *Nature, 312,* 419–422.

Wilson, H. R., & Wilkinson, F. (1998) Detection of global structure in Glass patterns: Implications for form vision. *Vision Research, 38,* 2933–2947.

Wilson, H. R., Wilkinson, F., & Habak, C. (1998) Detection and recognition of radial frequency patterns. *Vision Research, 38,* 3555–3568.