

Contrast integration over area is extensive: A three-stage model of spatial summation

Daniel H. Baker

School of Life and Health Sciences, Aston University,
Birmingham, UK



Tim S. Meese

School of Life and Health Sciences, Aston University,
Birmingham, UK



Classical studies of area summation measure contrast detection thresholds as a function of grating diameter. Unfortunately, (i) this approach is compromised by retinal inhomogeneity and (ii) it potentially confounds summation of signal with summation of internal noise. The *Swiss cheese* stimulus of T. S. Meese and R. J. Summers (2007) and the closely related *Battenberg* stimulus of T. S. Meese (2010) were designed to avoid these problems by keeping target diameter constant and modulating interdigitated checks of first-order carrier contrast within the stimulus region. This approach has revealed a contrast integration process with greater potency than the classical model of spatial probability summation. Here, we used Swiss cheese stimuli to investigate the spatial limits of contrast integration over a range of carrier frequencies (1–16 c/deg) and raised plaid modulator frequencies (0.25–32 cycles/check). Subthreshold summation for interdigitated carrier pairs remained strong (~4 to 6 dB) up to 4 to 8 cycles/check. Our computational analysis of these results implied linear signal combination (following square-law transduction) over either (i) 12 carrier cycles or more or (ii) 1.27 deg or more. Our model has three stages of summation: short-range summation within linear receptive fields, medium-range integration to compute contrast energy for multiple patches of the image, and long-range pooling of the contrast integrators by probability summation. Our analysis legitimizes the inclusion of widespread integration of signal (and noise) within hierarchical image processing models. It also confirms the individual differences in the spatial extent of integration that emerge from our approach.

Keywords: contrast detection threshold, spatial summation, area summation, contrast integration, contrast energy, computational modeling

Citation: Baker, D. H., & Meese, T. S. (2011). Contrast integration over area is extensive: A three-stage model of spatial summation. *Journal of Vision*, 11(14):14, 1–16, <http://www.journalofvision.org/content/11/14/14>, doi:10.1167/11.14.14.

Introduction

Most models of spatial vision attribute area (spatial) summation of contrast to probability summation between independent noisy detecting mechanisms (Anderson & Burr, 1991; Graham, 1989; Robson & Graham, 1981). However, recent psychophysical evidence (Meese, 2010; Meese & Baker, 2011; Meese & Summers, 2007, 2009) is inconsistent with this. A more successful model is one in which contrast integration (linear spatial summation of contrast) follows a nonlinear (square-law) contrast transducer (Foley, Varadharajan, Koh, & Farias, 2007; Meese & Summers, 2007, 2009) and additive noise (Meese, 2010). The cascade of the two quadratic effects produced by square-law transduction (Legge, 1984) and summation of signal and noise (Campbell & Green, 1965; Tyler & Chen, 2000) means that sensitivity improves with the fourth root of area, similar to the conventional probability summation model and several empirical reports (see Meese & Baker, 2011 for a review). However, the behavior of model and humans is very different when contrast area is manipulated in more interesting ways, as we now describe.

Meese and Summers (2007) designed what we have come to call a “Swiss cheese” stimulus (Meese & Baker, 2011). In this type of stimulus, diameter is constant and a sine-wave carrier is modulated by a raised plaid pattern such that it contains interdigitated patches of high-contrast regions (“cheese”) and low/zero-contrast regions (“holes”). Using plaid modulators of opposite phase, a complementary pair of Swiss cheese stimuli can be created (see Figure 1a). For convenience, we sometimes give these the nominal titles of “black” and “white” cheese (or “checks”), by reference to the polarity of the modulator at the center of the display. Of course, when “black” and “white” cheeses are summed, this recreates the original carrier grating (the “full fat cheese” or the “full” stimulus). Thus, contrast detection thresholds of the carriers for “black” and “white” cheeses and their sum can be used to assess area summation of contrast.

Meese and Summers (2007, 2009) did this and found that sensitivity improved substantially when a “black” cheese was added to an original “white” cheese, providing good evidence for the physiological signal combination model described above. Based on the results of an identification experiment, they argued that observers

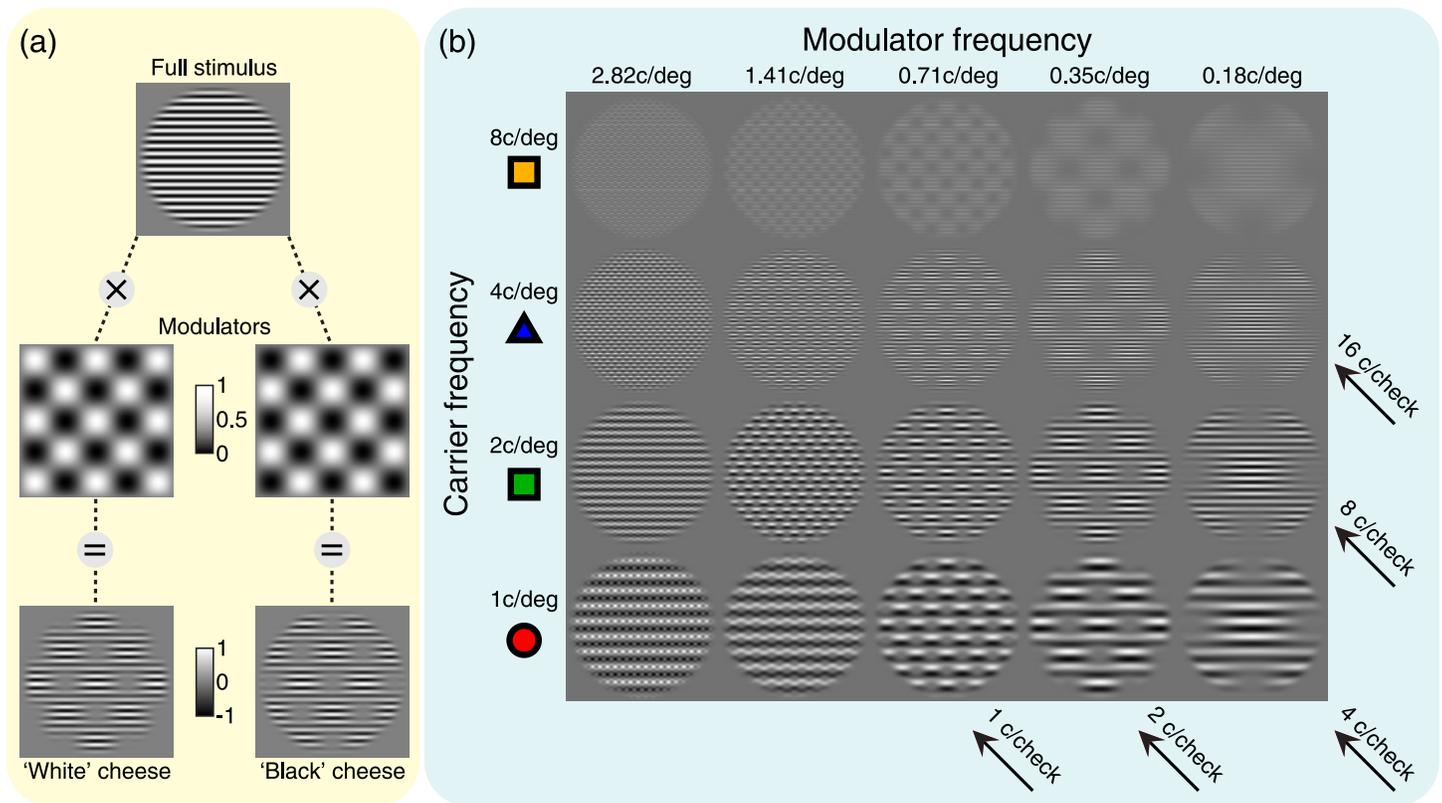


Figure 1. Swiss cheese and full stimuli. (a) Cheeses (check stimuli) were the product of a full stimulus and a raised plaid modulator. The modulator was in \pm cosine phase with the center of the image, producing either “white” or “black” cheeses (left and right columns). The original full stimulus could be reconstructed by summing the two cheeses. (b) High-contrast examples of “white” cheese stimuli for a range of carrier and modulator frequencies. Stimuli for the 16 c/deg carrier were identical to those labeled 8 c/deg but viewed at twice the distance. Symbols on the left correspond to those used in subsequent data figures. Stimuli along a diagonal path through this stimulus space have the same number of cycles per check but differ in carrier and modulator frequencies.

could not construct an accurate template of the Swiss cheese and therefore resorted to uniform integration over much of the stimulus region. This meant that the level of internal noise (assumed to be Gaussian and additive) did not vary with the stimulus condition and that the primary factor affecting the performance benefit from area summation was the square-law signal transduction. For the Swiss cheese stimulus, this benefit is somewhat greater (in model and data) than the factor of $\sqrt{2}$ (3 dB) that might be expected on first approximation (Legge, 1984) because the “black” and “white” cheeses are not independent signals (there is spatial overlap between them). This point was addressed using the closely related Battenberg stimuli of Meese (2010). These are made from clumps of independent micropatterns, which means that complementary stimulus pairs can be constructed without spatial overlap. In summation experiments with these stimuli, performance dropped to the expected factor of $\sqrt{2}$ (3 dB), supporting the model (Meese, 2010).

Meese and Baker (2011) proposed that the process of area summation of contrast is involved in deriving a general-purpose size code for textures and patterns. They suggested that this task is achieved with the help of a

contrast normalization network (Meese & Baker, 2011; Meese & Summers, 2007) that protects the suprathreshold image-contrast code (Albrecht & Geisler, 1991; Heeger, 1992). On this hypothesis, suprathreshold stimulus size is encoded by a population of mechanisms with different sized integration apertures (Meese & Baker, 2011), but perceived contrast is prevented from varying with stimulus size, as is required (Cannon & Fullenkamp, 1991). On the other hand, when the stimulus is around detection threshold, the gain control is (effectively) inoperative and only the performance-enhancing benefits of contrast integration are seen. Hence, sensitivity improves with the integral of contrast over space. We use the term “contrast integration” to refer to this type of signal combination (Syväjärvi, Näsänen, & Rovamo, 1999), distinct from formulations that attribute area summation to probability summation (Meese & Williams, 2000; Robson & Graham, 1981).

Having concluded that contrast integration extends beyond that of a typical three- or four-lobe receptive field in V1 (Meese, 2010), two questions naturally arose. First, over what spatial range does this integration extend (i.e., what is the size of the integration aperture)? Second, is the

process we have identified scale invariant (Banks, Geisler, & Bennett, 1987; Howell & Hess, 1978; Rovamo, Ukkonen, Thompson, & Näsänen, 1994), relating to the number of carrier cycles, or does it relate to retinal image size?

Here, we vary the modulation and carrier frequencies in Swiss cheese stimuli at contrast detection threshold to address the first and second questions, respectively.

Basic intuitions

The intuitions behind our approach can be described as follows. First, the integral of contrast over area (the “contrast area”) in a full stimulus is exactly twice that in a Swiss cheese stimulus. Therefore, if vision performs extensive spatial integration of contrast over the stimulus, then sensitivity to the full stimulus should be twice that to the Swiss cheese (we assume a linear contrast transducer here, merely to simplify the exposition). However, this benefit arises only if the integration aperture extends across at least a pair of checks (i.e., a cheese and hole pair). For example, if the integration aperture is no larger than a single check, there can be little benefit from filling the holes in the Swiss cheese (to make a full stimulus) because this makes little difference to the contrast area within the aperture. Therefore, we reasoned that by manipulating the size of the check region, we could determine the size of the integration aperture by observing the point at which the benefit of filling the holes in the Swiss cheese is lost.

One potentially complicating factor is retinal inhomogeneity. The Swiss cheese stimulus was designed to combat this problem (Meese & Summers, 2007), at least in part. For example, if the check size is small compared to the rate at which sensitivity declines with eccentricity, then each type of cheese (“black” and “white”) are spread fairly equally over the inhomogeneity, with the net result that sensitivity is fairly equal for the two (Meese & Baker, 2011; Meese & Summers, 2007, 2009). However, if the check size is large compared to the rate of decline, then the size of the integration aperture might be underestimated because the loss of sensitivity in the peripheral hole regions would mean that there is little benefit from filling them in. For this reason, we repeated the experiments using a second method (the normalization method), where we adjusted the contrasts in the two different check regions to equate their sensitivity.

Methods

Equipment

A PC was used to control a ViSaGe system (Cambridge Research Systems, Kent, UK) with 14-bit contrast

resolution. Stimuli were displayed on a gamma-corrected Nokia Multigraph 445× monitor with a mean luminance of 60 cd/m² and a refresh rate of 120 Hz.

Observers

Three observers took part in the experiments (DHB, SAW, and TSM). DHB and SAW were emmetropic and TSM wore his normal spectacle correction. All three were psychophysically well experienced. DHB and TSM are authors.

Stimuli

Carrier signals were horizontal sinusoidal gratings with a diameter of 10°, spatially curtailed by a raised cosine envelope (cosine half-period of 1°; full-width, half-height of 9°). Carrier frequencies (at the primary viewing distance of 119 cm) were 1, 2, 4, and 8 c/deg, and all carriers were in sine phase with the center of the display.

Modulators were raised plaids, constructed from two orthogonal grating stimuli ($\pm 45^\circ$) in positive (“white” check) or negative (“black” check) cosine phase with the center of the display (see Figure 1a). The two phases of modulator (“black” and “white”) were applied to the contrast of the carrier to produce interdigitated “Swiss cheese” component stimuli, which could be summed to reproduce the original (unmodulated) carrier (see Meese & Summers, 2007). Modulator frequencies were at and between 0.18 c/deg and 2.82 c/deg, which produced a stimulus set with check sizes from 0.25 to 32 cycles/check (arranged in octave steps). High-contrast examples of “white” cheese stimuli are shown in Figure 1b.

For the 16 c/deg carrier condition, we used the highest spatial frequency stimuli described above and doubled the viewing distance. This doubled the carrier and modulator frequencies and halved the stimulus diameter (in degrees of visual angle).

We measured contrast detection thresholds for the carriers of our stimuli. We express carrier contrast in decibels (dB), calculated as $C_{dB} = 20 \log_{10}(C_{\%})$, where $C_{\%}$ is Michelson contrast in percent, defined as $C_{\%} = 100(L_{max} - L_{min}) / (L_{max} + L_{min})$, where L is luminance. Goodness of fit for models was assessed using the root-mean-square error (RMSe), defined as

$$\text{RMSe} = \sqrt{\frac{\sum_{i=1:n} (\text{model}_i - \text{data}_i)^2}{n}}, \quad (1)$$

where model_i and data_i are the model predictions and empirical data points (in dB) for the i th condition, and n is the number of data points.

Procedure

Observers were seated with their head and chin supported by a rest. Stimuli were viewed binocularly and presented in the center of the display. Observers were instructed to fixate a small black dot, which was presented continuously in the center of the display.

A two-interval forced-choice (2IFC) paradigm was used to estimate contrast detection thresholds for the carriers. Stimuli were presented for 100 ms in one of two intervals (selected at random), each indicated by auditory beeps and separated by an interstimulus interval (ISI) of 400 ms. In the interval that did not contain the stimulus, and in the gaps between intervals and trials, the display remained at a constant mean luminance (apart from the fixation point). Observers indicated which interval they believed contained the target using a two-button mouse and received feedback on the correctness of each response. Carrier contrast was controlled by a pair of 3-down, 1-up staircases with a minimum step size of 3 dB. Each staircase terminated after the lesser of 70 trials or 12 reversals, and the data from the two staircases were pooled for threshold estimation (see below).

Two different experimental procedures were used to measure summation of contrast over area. In the *basic method*, we measured thresholds for the “black” and “white” cheeses (the two components) and the full stimulus (the compound). Summation was estimated by subtracting the threshold for the full stimulus (in dB) from the lowest (best) of the two cheese stimuli (usually the “white” cheese threshold). This is equivalent to calculating the ratio of these two threshold contrasts when expressed in percent. In the *normalization method*, the full stimulus was the sum of “black” and “white” cheeses as before, but to equate sensitivity to the two cheeses, the contrast of the cheese to which the observer was least sensitive was raised (based on the initial estimates of thresholds; e.g., see Baker, Meese, Mansouri, & Hess, 2007). Summation was calculated in the same way as above.

DHB ran each method in separate experiments for all carrier frequencies. In each case, the pairs of “black” and “white” cheese thresholds were measured contemporaneously using an interleaved design. The cheese stimuli were identical across methods and so their thresholds were averaged to achieve the most robust estimates. For SAW, the approach was slightly different. First, the pairs of cheese thresholds were measured in interleaved trials. Then, thresholds for each of the full stimuli (basic and normalized) were measured, also interleaved across trials. SAW performed the experiment at all carrier frequencies except 1 c/deg. TSM performed the experiment in the same way as SAW but for only the 4 c/deg carrier condition.

The raw data were analyzed using probit analysis (Finney, 1971) to estimate threshold at the 75% correct point on the psychometric function. Following our usual

procedure, if the standard error of the probit fit exceeded 3 dB, the data were discarded and the appropriate condition was repeated. Because the normalization method required reliable estimates of the thresholds for the two cheeses and the full stimulus, if any one of these was discarded all three were remeasured. Thresholds and summation ratios were averaged across 4 repetitions of the experiment.

Results and discussion

Preliminary analysis

Figure 2 shows the summation results using the normalization method for each of the 3 observers for the 4 c/deg carrier (see [Methods](#) section). The summation measure indicates the performance benefit (in dB) of filling the holes in the Swiss cheese with contrast. The curves in Figure 2 show the predicted benefit (summation) for a single sine-phase linear filter element (receptive field) in the center of the display for each of two filter bandwidths (see inset and figure caption). (All filters were Cartesian separable two-dimensional log-Gabor filters (see Appendix C of Meese, 2010) with bandwidths reported as full-widths (for spatial frequency) or \pm half-widths (for orientation) at half-height.)

The analyses in Figure 2 show the effects of contrast summation *within* the model filter elements. However, even for the large 8-lobe filter element—which is probably unreasonably large (Foley et al., 2007)—the breadth of summation is nowhere near enough to account for the experimental results: Spatial pooling of some sort must be involved.

Previous detailed analyses (Meese, 2010; Meese & Summers, 2007, 2009) have consistently shown that spatial probability summation across the elements of our standard size filter (spatial frequency bandwidth of 1.6 octaves, orientation bandwidth of $\pm 25^\circ$, 4-lobe sine-phase receptive field) cannot account for our results. This was confirmed again here as follows. Figure 3 shows the results for 3 observers for each method of data collection (standard and normalization methods) and predictions using Minkowski summation across filter outputs with exponents (μ) of 4 and 8, arguably each of which are good approximations to spatial probability summation (see Appendix A). For the standard size filter element (solid curves), each of these formulations typically underestimated summation for all three observers—sometimes quite badly. For the larger filter element (dashed curves), the predictions for $\mu = 4$ were quite good for SAW and DHB but still underestimated summation for TSM (particularly in Figure 3f). Furthermore, as we have shown elsewhere (Meese & Summers, 2009), this formulation is also inconsistent with the steep slope of the

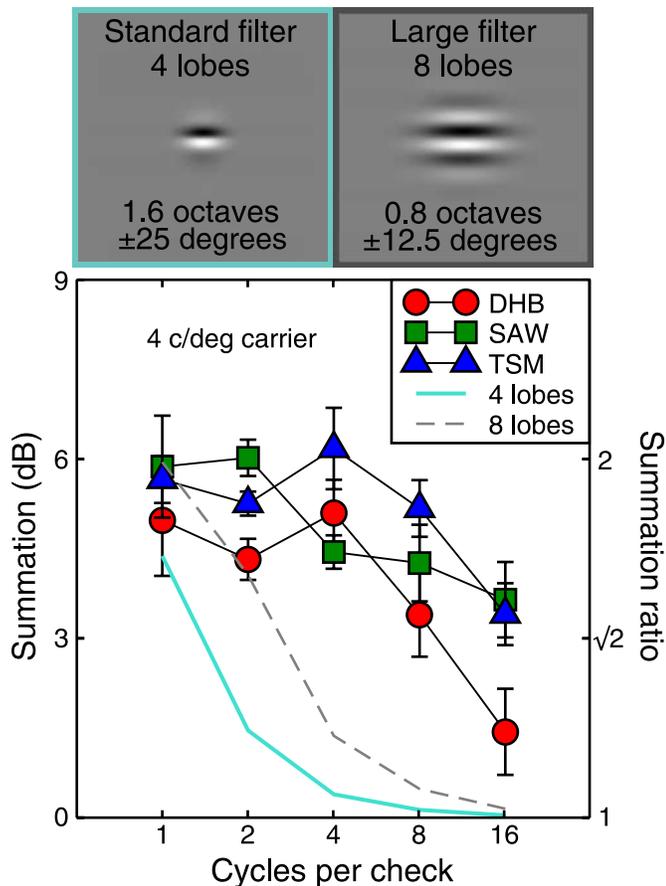


Figure 2. Summation results for three observers using the normalization method. Summation is derived by plotting sensitivity to the full stimuli relative to the cheese stimuli. The cyan and gray curves show the predicted level of summation for a single sine-phase model filter element in the center of the display (there were negligible differences across observers for these predictions; here, we show the average). The standard filter (cyan curve) has an orientation bandwidth of $\pm 25^\circ$ and a spatial frequency bandwidth of 1.6 octaves. Its filter element (receptive field) has 4 lobes. The larger filter (dashed gray curve) has an orientation bandwidth of $\pm 12.5^\circ$ and a spatial frequency bandwidth of 0.8 octave. Its filter element has 8 lobes. Spatial summation occurs within each of these single model filter elements, but neither is sufficient to account for the empirical results.

psychometric function (not shown here). Thus, even when we constructed the model to favor probability summation as best we might (a fairly large filter element with a linear transducer and a fairly low Minkowski summation exponent of 4; see Appendix A), it could not account for all of our results. Therefore, we abandoned the probability summation model and turned our attention (see later) to a model involving linear summation (contrast integration) over area, following spatial filtering and square-law contrast transduction.

Main analysis

Contrast detection thresholds for “black” and “white” cheeses are shown in Figures 4a–4c for three observers. The results are plotted as functions of carrier cycles per modulator check (i.e., check size), so that lower frequency modulators are placed to the right of the plot (consistent with the layout of Figure 1b). As check size increased, thresholds for the “black” and “white” cheeses diverged. Sensitivity to “white” cheese stimuli tended to improve with check size. Presumably, this was because more of the stimulus energy was concentrated in the central part of the visual field where sensitivity was greatest (Foley et al., 2007; Pointer & Hess, 1989; Robson & Graham, 1981). Conversely, for the “black” cheese stimuli, the contrast energy was distributed further away from the central visual field and sensitivity tended to decline with check size. However, over much of the range, the thresholds for the “black” and “white” cheeses were fairly similar, the separation becoming apparent only for the largest one or two check sizes (the points toward the right of each pair of functions). This demonstrates that our Swiss cheese design was effective in combating the undesirable influences of retinal inhomogeneity over much but not all of the range (see Meese & Baker, 2011; Meese & Summers, 2007 for discussion). For the larger check sizes, the relative insensitivity to the “black” cheese stimuli meant that the basic method—where the two cheeses had the same contrasts in the full (compound) stimulus—would be fairly insensitive to the effects of very long-range contrast integration (e.g., ≥ 32 carrier cycles), should it exist. We attempted to compensate for this shortcoming by using the normalization method, where the contrasts of the two cheeses in the full stimulus were adjusted to equate their detectability (see Methods section).

The middle and bottom rows of Figure 4 show summation ratios measured using each of the two experimental methods (basic and normalization). In both cases, summation was substantial (~ 6 dB, or a factor of 2) for modulators that produced ≤ 2 cycles/check, presumably due to summation within linear filter elements (Meese, 2010; Appendix B). As check size increased, the level of summation declined slowly. For the basic method (where component contrasts were always the same), it reached the low levels usually associated with probability summation (1 or 2 dB) by ~ 32 cycles/check. In general, slightly more summation was evident using the normalization method (e.g., it tended not to drop below ~ 3 dB, even at 32 cycles/check). Presumably, this was because the normalization method was effective in overcoming the differences in component sensitivities for the large modulators (see Figures 4a–4c), thus providing a cleaner picture of the area summation process. At first glance (Figures 4g–4i), this might suggest that the range over which summation is more potent than probability summation extends up to 64 cycles (note that the implied

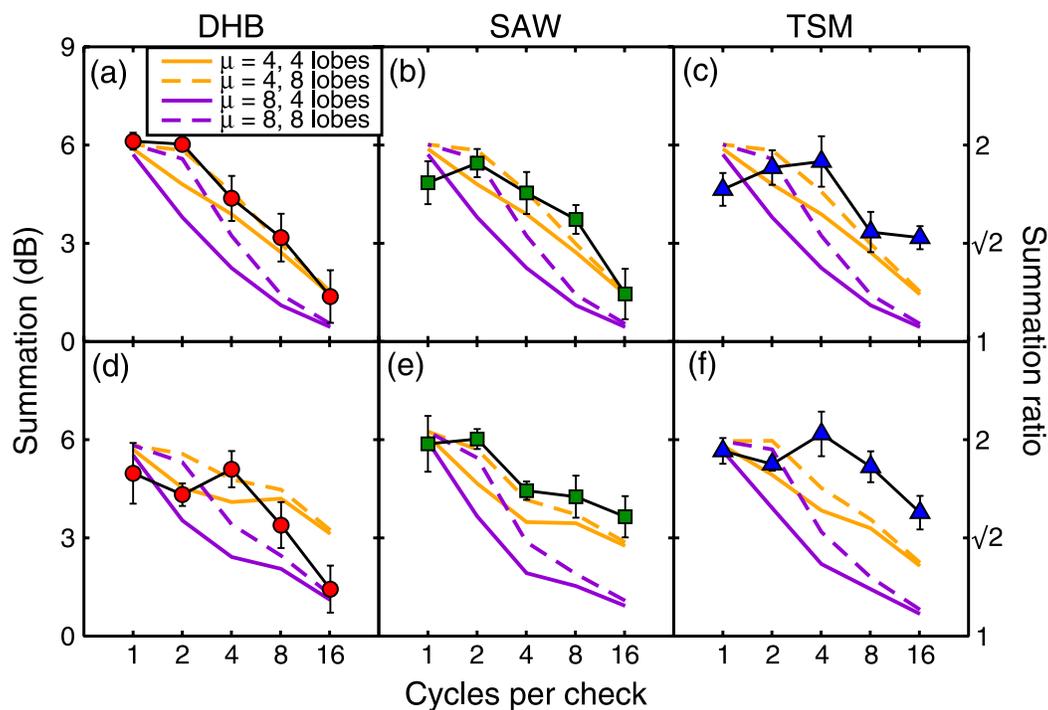


Figure 3. Area summation is more potent than predicted by probability summation. Data are for a 4 c/deg carrier for each of the three observers (different columns) and each of the two experimental methods (the upper and lower rows are for the basic and normalization methods, respectively). Solid curves are model predictions for two different Minkowski exponents ($\mu = 4$; $\mu = 8$; see Appendix A) for the standard size (4-lobe) filter element. The dashed curves are for the larger (8-lobe) filter element.

summation region is twice that of the check size). However, the effects of summation within filter elements for these stimuli means that detailed computational analysis is needed before firm conclusions can be attempted. We do this in the next section.

Another motivation for our study was the question of whether the integration aperture is scale invariant. The tendency for the summation functions in Figure 4 to superimpose might suggest that this is so (i.e., that the integration aperture is a fixed number of carrier cycles). However, since retinal inhomogeneity can contribute to a decline in summation for large apertures (see Introduction section) and since this is scale invariant in our model (Pointer & Hess, 1989; Robson & Graham, 1981), these two effects need to be teased apart. We attempted to do this in the next section.

The range of contrast integration for a scale-invariant model

The summation curves in Figure 4 are spatially extensive, but several factors contribute to their shape including: retinal inhomogeneity, within-filter summation, the size of the long-range integration aperture, probability summation, and the psychophysical method (basic vs. normalization). Our main aim was to establish what our results imply for neuronal convergence (i.e., the size of

the integration aperture), but the complicating factors above mean that this cannot be achieved by direct interpretation of the data. To try and see more deeply into our results, we considered several variants of the filter-based model developed by Meese and Summers (2007, 2009). Our general strategy was to fix what parameters we could—where there is some consensus on their values—leaving us to explore the parameters for which values are unknown (e.g., the size of the integration aperture).

Our models take bitmaps of the stimulus as input and involve several image processing stages in the following sequence:

1. Retinal inhomogeneity (uneven sensitivity) across a two-dimensional array of image pixels;
2. Spatial filtering (using log-Gabor filters) to produce a two-dimensional array of “response” pixels;
3. Pixel-wise nonlinear transduction (squaring);
4. Zero-mean Gaussian noise added to each pixel (conceptual);
5. Integration across pixels (details depend on the model variant);
6. Minkowski pooling (e.g., probability summation) across integrators (details depend on the model variant);
7. Construction of a decision variable.

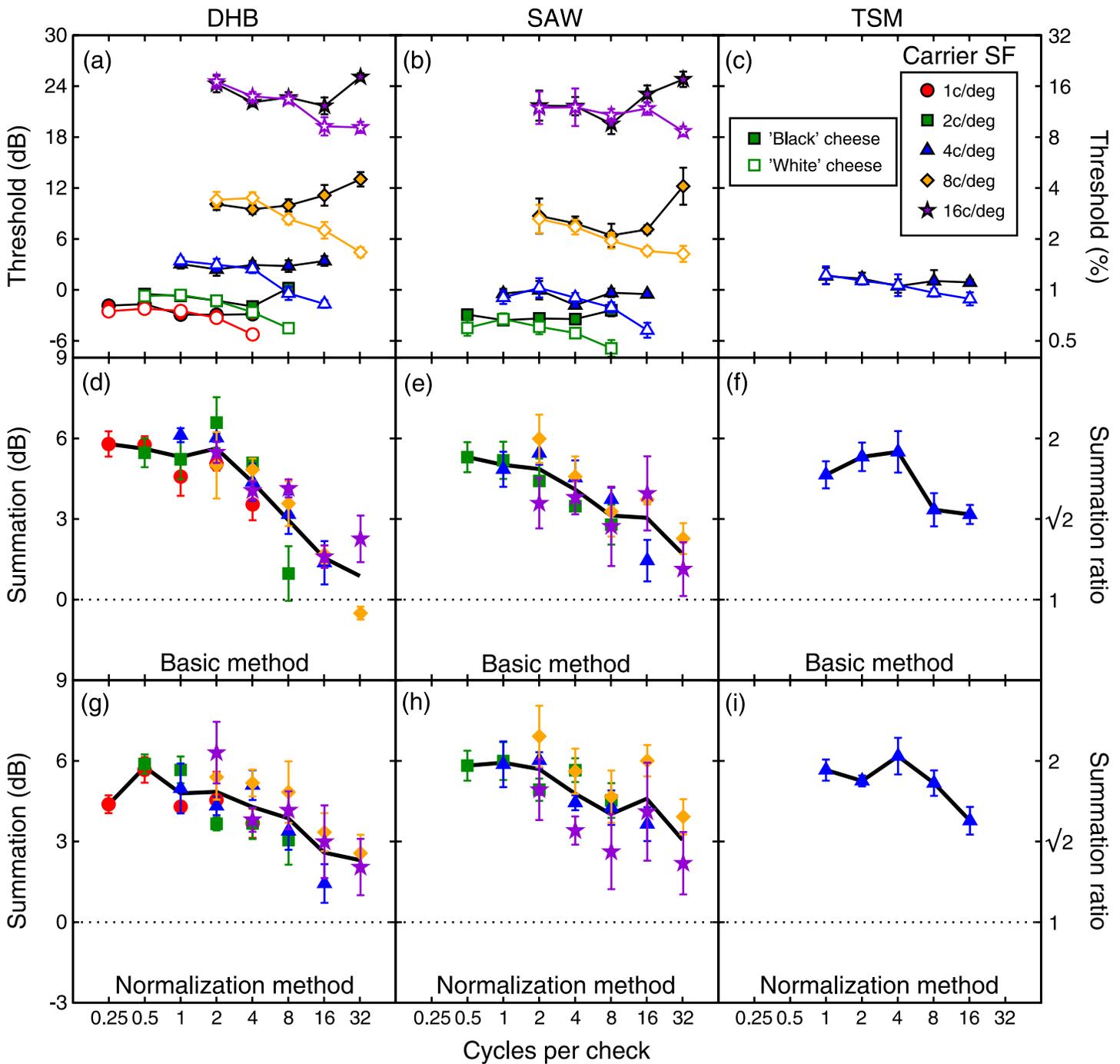


Figure 4. Contrast detection thresholds and summation ratios for Swiss cheese stimuli. (a–c) Detection thresholds for “black” cheese (filled symbols) and “white” cheese (open symbols) for a range of carrier and modulator frequencies for three observers. (d–f) Summation results for the basic method. (g–i) Summation results for the normalization method. In panels (d–i), solid black curves indicate the average summation across carrier frequency. All data points are averaged across four repetitions, and error bars indicate ± 1 SE.

The stimuli were processed at the same spatial resolution as that used in the experiments. Estimates of retinal inhomogeneity were derived from detailed measurements across the visual field. These were performed by Baldwin, Meese, and Baker (2010) at a spatial frequency of 4 c/deg, using the same (or similar) equipment and

observers as in the experiments here, and are summarized in Appendix A. For simplicity, we used a single (average) attenuation surface for all three observers but confirmed that the modeling results were very similar when attenuation surfaces were tailored to individual observers (not shown). We also assumed that the attenuation surface

was scale invariant (Pointer & Hess, 1989; Robson & Graham, 1981). This is to say that retinal sensitivity loss is related to eccentricity in terms of stimulus cycles rather than visual angle. We found that it made no meaningful difference whether inhomogeneity (the attenuation surface) was placed before or after the spatial filtering.

The log-Gabor filters had spatial frequency bandwidths of 1.6 octaves and orientation bandwidths of $\pm 25^\circ$ and were centered on the orientation and spatial frequency of the relevant carrier. This is the “standard” filter element in Figure 2. For simplicity, we used only sine-phase filters in the modeling but confirmed that almost identical results were found when summing over a quadrature pair.

In previous models of this type (e.g., Meese & Summers, 2007), the output of each filter element was squared and added to a stochastic sample of noise at each pixel location. However, since (i) the model integration region was constant for each family of functions (see below) and (ii) we were not interested in the details of the absolute signal-to-noise ratios, it was not necessary to implement the stochastic stage in the model here. Instead, a deterministic output was constructed from the pooled signal responses (see below for details) and compared to a constant (k), related to the standard deviation of the pooled noise sources across the image. Conceptually, the stimulus was deemed to be detected with a probability of 75% when the output equaled k . In practice, rearranging the model equations and solving for contrast analytically allowed us to derive the contrast detection thresholds.

In what we have described so far, k was the only free parameter in the model. Owing to the contrast sensitivity function (Blakemore & Campbell, 1969), it varied with carrier frequency, allowing each pair of model curves (Figures 6a–6c) to be slid freely up and down the ordinate to achieve the best possible fits. Thus, there were 5, 4, and 1 degrees of freedom associated with this parameter for DHB, SAW, and TSM, respectively. However, this sensitivity parameter was of no interest in the study here. For example, it was irrelevant for (i) capturing the splaying shape of the threshold functions (e.g., Figures 5a–5c) and (ii) modeling the summation functions (e.g., Figures 5d–5i), which are *relative* measures of sensitivity.

We considered several ways in which long-range contrast integration might be performed. The simplest was to perform linear summation (of the squared local contrast responses) over the entire stimulus region. This produced a reasonable description of the Swiss cheese thresholds, though the separation of the model functions for the “black” and “white” cheeses was less marked than it was in the data toward the right of the plot (see Figure 5a, which illustrates the fits for DHB). This happens in the model because the large check size (relative to the rate of sensitivity loss with eccentricity) means that much of the “white” cheese stimulus benefits from the highly sensitive central retina whereas very little of the “black” cheese does. However, summation between the “black” and “white” cheeses was substantially overestimated for larger

check sizes (Figures 5d and 5g). This implies that DHB was not able to integrate contrast over the entire stimulus. So what was the upper range for contrast integration?

To estimate this, we ran the model for a range of different sized integration apertures. Apertures were circular, located at the centers of the stimuli, and had hard edges (though Gaussian pooling regions produced similar results). All pixels inside the integration aperture contributed to the model output, and those outside it were discarded. As the diameters of the apertures spanned the same number of cycles for each carrier spatial frequency, this variant of the model is scale invariant.

Figures 6a, 6c, and 6e show the RMS errors of the model predictions as functions of the diameter of the integration aperture (in carrier cycles). The best model performance (i.e., the function minima) for a single aperture involved pooling over about 12 carrier cycles (DHB) or more (SAW, TSM). This is confirmed in Figures 5e and 5h, which show that the summation predictions (no free parameters) were fairly good for DHB with an aperture of this size. However, predictions for the thresholds were typically rather poor (Figures 6a, 6c, and 6e, red curves), particularly for the “black” cheese (Figure 5b, black curves), where the model was much less sensitive than human observers at the larger check sizes. This was because when the check size exceeded the integration aperture, very little “black” check stimulus contrast was available to the model and thresholds rose dramatically. Thus, to summarize, analysis of the Swiss cheese thresholds (Figures 4a–4c) suggested very large integration apertures (Figures 5a, 6a, 6c, and 6e, red curves), whereas the effects of filling the holes in the cheese (the summation results; Figures 4d–4i) suggested somewhat smaller integration apertures (Figures 5e, 5h, 6a, 6c, and 6e). Can this conflict be resolved?

As an alternative detection strategy, we considered an arrangement where contrast integration took place within multiple apertures, followed by further pooling between them. In this model variant, the apertures overlapped such that the center of one sat on the circumference of the next, but the details of the tiling arrangement were not critical. We used Minkowski pooling ($\mu = 4$) across apertures, as described in Appendix A, consistent with contemporary interpretations of probability summation (Tyler & Chen, 2000). MATLAB code for this model can be found in the [Supplementary materials](#).

The effects of varying the integration aperture on goodness of fit for this model variant are shown in Figures 6b, 6d, and 6f. Although there is still a tendency for the threshold analysis and the summation analysis to pull to larger and smaller integration apertures, respectively, the overall improvement in performance, particularly for the cheese thresholds (red), is marked (compare across columns). Unfortunately though, the functions are fairly shallow and a little inconsistent between measures (different solid curves in each panel) and observers (different rows), making a firm conclusion about the size of the

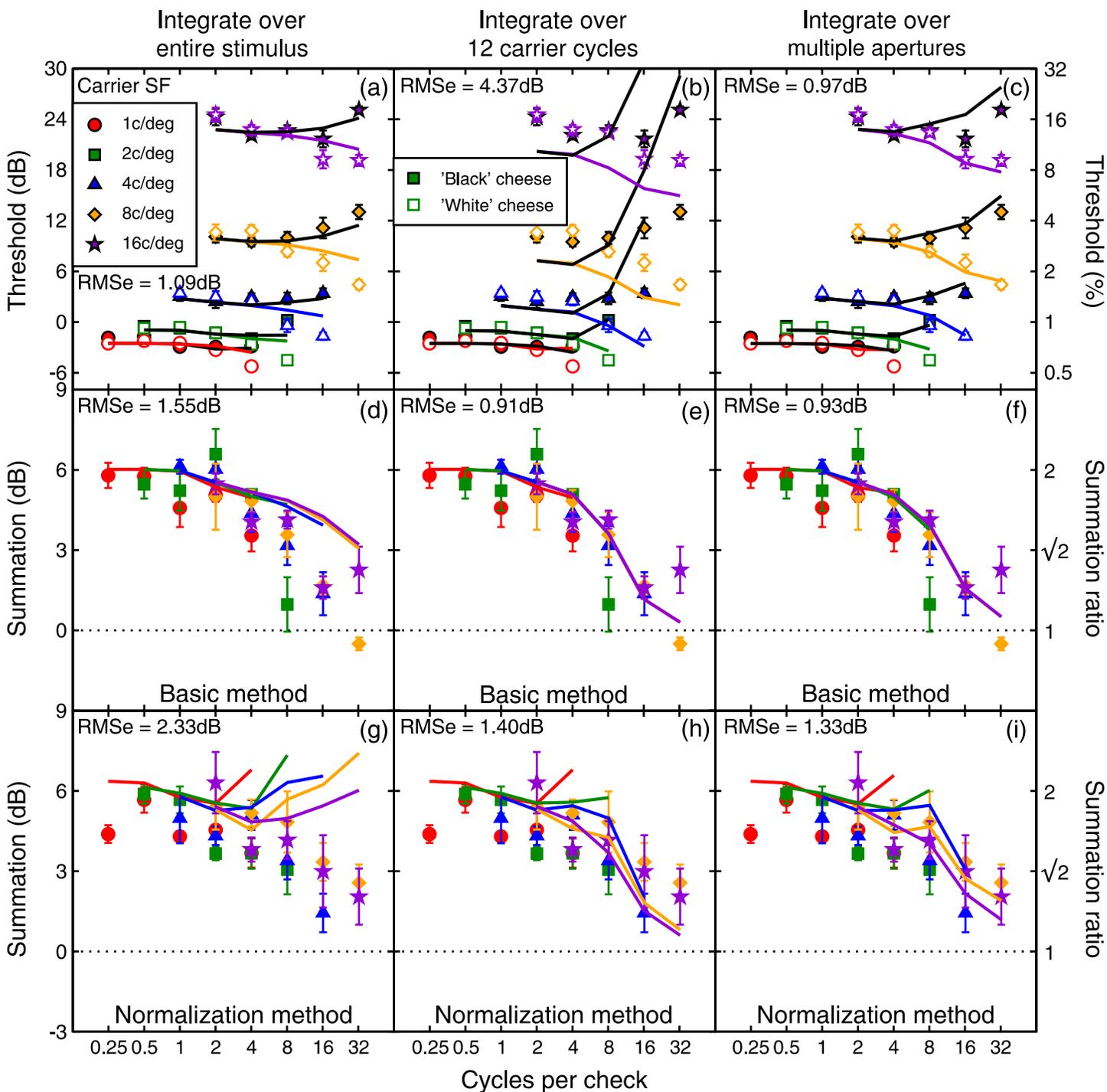


Figure 5. (a–c) Examples of model fits for detection thresholds for “black” and “white” cheeses and (d–i) summation results from the main experiment. Data are for observer DHB and are shown for three model variants (different columns). RMS errors of the fits are given in each panel. Fits to the data of the other observers were qualitatively similar. In the first column, integration extends over the entire stimulus region. In the second column, it is restricted to a circular aperture with a diameter of 12 cycles. In the third column, multiple apertures (mechanisms) like those in the second column tile the image with spatial overlap. Minkowski summation was then performed across the multiple integration apertures.

integration aperture difficult. In an attempt to improve on this, we calculated the RMS error combined across the three measures for each observer (dashed magenta curves in Figure 6). These had minima at 12, 45, and 23 cycles

for DHB, SAW, and TSM, respectively. This analysis confirmed our impression that DHB was the weakest contrast integrator in the study (consistent with other unpublished observations). Nevertheless, all of his results

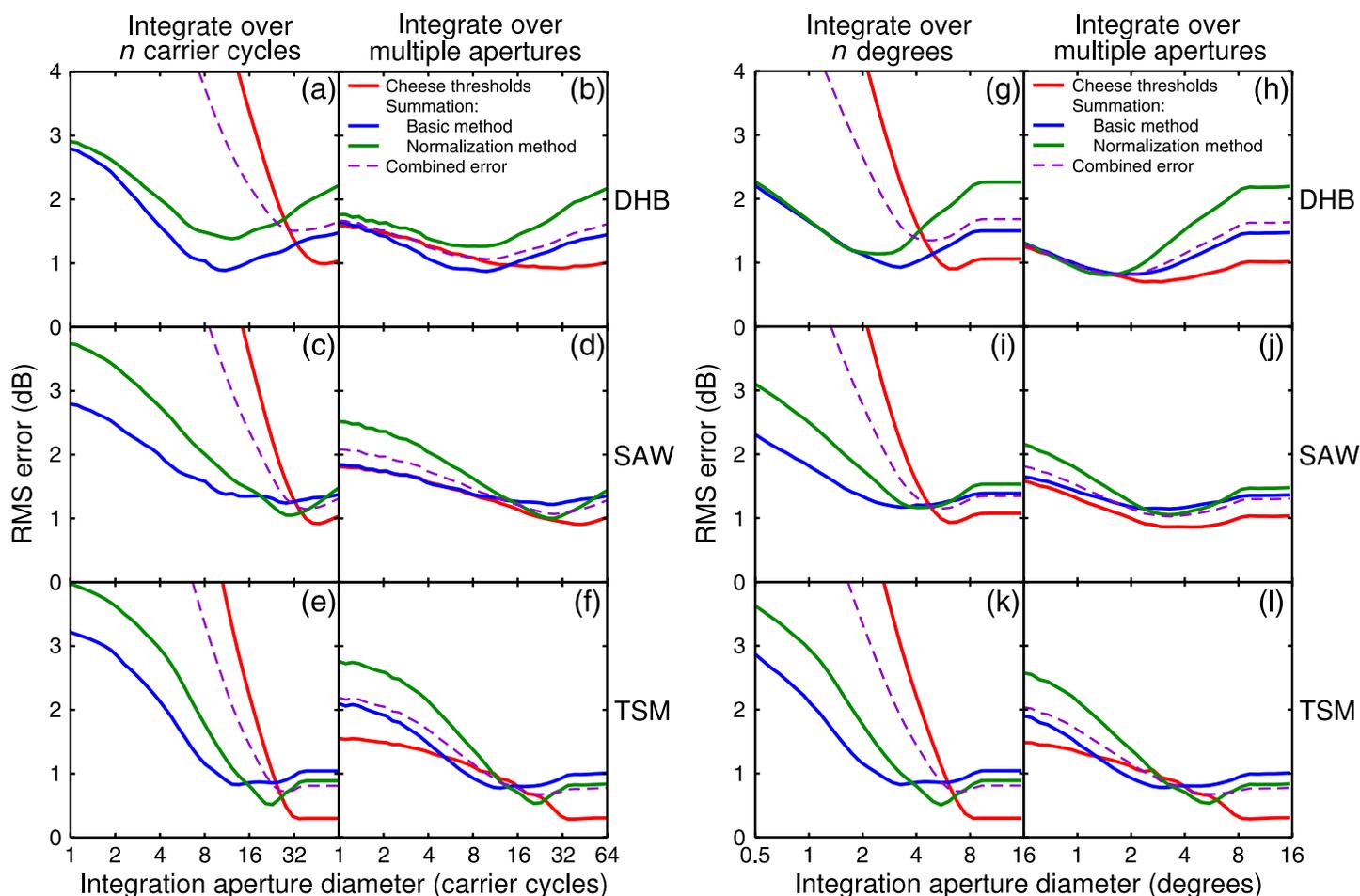


Figure 6. RMS errors of model fitting for two-by-two factorial model variants. One factor was the different scale dependencies (left and right blocks of panels). The other factor was the different integration strategies (different columns). All models involved scale-invariant retinal inhomogeneity, spatial filtering, square-law contrast transduction, and contrast integration. Results are shown for each of the three observers (different rows). In the left block of panels (a–f), integration was scale invariant, occurring over apertures defined in carrier cycles. In the right block of panels (g–l), integration was scale dependent, occurring over apertures defined in degrees of visual angle. Integration (linear spatial summation) occurred within hard-edged apertures with the diameters shown by the x-axis. In the left-hand columns of each block, there was a single, centrally placed aperture. In the right-hand column of each block, there was Minkowski pooling (using a Minkowski exponent of $\mu = 4$) over multiple integration apertures. Errors are shown separately for the cheese thresholds (red), summation using the basic method (blue), and summation using the normalization method (green). The dashed magenta curves show the combined (root mean square) error across all three measures.

were fairly well described by a model using multiple contrast integration apertures with diameters of 12 cycles (Figures 5c, 5f, and 5i): much greater than the one or two cycle limit from the within-filter summation region that is often assumed. For SAW and TSM, it seems likely that their apertures are larger than this (Figures 6d and 6f) but possibly not quite as large as the stimulus (note the minima in the green curves in Figure 6).

The most successful model variant—involving multiple fixed size integration apertures—contained two parameters that could vary (k and the diameter of the integration aperture), plus the fixed parameters that set the filter bandwidths and the Minkowski pooling (μ), which also influenced summation behavior (e.g., see Figure 3). Since k was set separately for each carrier frequency, the model

had 6, 5, and 2 free parameters for DHB, SAW, and TSM, respectively. However, since the k parameters were irrelevant for summation behavior (see above), only one of the free parameters (the diameter of the integration aperture) influenced this independently for each observer.

The range of contrast integration for a scale-dependent model

Inspired by the scale invariance of retinal inhomogeneity (Pointer & Hess, 1989; Robson & Graham, 1981) and the approximate superposition of the summation functions in Figure 4, we assumed scale invariance in the modeling above. That is, the attenuation surfaces and the summation apertures in the models were linked to the

number of stimulus cycles and therefore differed in visual angle across spatial frequency. Although the case for scale-invariant retinal inhomogeneity is quite strong (Pointer & Hess, 1989; Robson & Graham, 1981), it is not clear that the integration aperture is scale invariant (see our comments in the first part of the [Main analysis](#) section) and it is natural to ask how well a model would fare under different assumptions. To examine this, we produced scale-dependent variants of the models, pooling over either a single (Figures 6g, 6i, and 6k) or multiple (Figures 6h, 6j, and 6l) apertures with the same visual angles across spatial frequencies (and therefore scale-dependent numbers of cycles). Note that for TSM, who gathered data at a single carrier frequency, these versions of the model are formally identical to the scale-invariant versions (Figures 6k and 6l are simply relabeled versions of Figures 6e and 6f). However, the model behaviors are different for the other two observers.

This modeling shows that in spite of the approximate superposition of the summation functions when expressed in cycles per check (Figure 4), the scale invariance assumption was not critical for achieving acceptable fits to the full data sets. In fact, the RMS errors achieved in the analysis here are comparable to those involving scale invariance (e.g., compare the low points of the dashed magenta curves across Figures 6b and 6h and Figures 6d and 6j). On this alternative view, we estimated integration apertures with diameters of 1.27, 3.25, and 5.75 deg for DHB, SAW, and TSM, respectively.

General discussion

We measured contrast integration (area summation) for sinusoidal gratings modulated by a raised plaid envelope (Swiss cheese stimulus) for a range of carrier and modulator frequencies. Integration appeared to be scale invariant when plotted as a function of carrier cycles per modulator check, though detailed modeling was not able to reject an alternative interpretation in terms of a scale-dependent integration process using apertures of constant visual angle. Summation was strong for all observers up to 4 cycles/check and declined for larger check sizes. This implied that integration of contrast extended over at least 8 carrier cycles. Computational modeling supported this conclusion, with optimal summation regions having a diameter of 12 or more cycles (varying across observer) but probably less than 64 cycles. Alternatively, the integration region might be 1.27 degrees or more (varying across observers) but probably less than 10 deg (the diameter of our stimuli).

In spite of the uncertainty about the appropriate metric (degrees or cycles) for describing the integration region, the overall message is clear: Contrast integration is more spatially extensive than the receptive field of a single filter

element. This justifies the inclusion of neuronal convergence of the local contrast code in hierarchical models of visual perception beyond the primary visual cortex (where most receptive fields are quite small). The implications of this for suprathreshold vision have been discussed elsewhere (Meese & Baker, 2011; Meese & Summers, 2007).

Criticisms and concerns

One potential concern about our stimuli is that when the number of cycles per check was low, the Michelson contrasts of the “black” and “white” cheeses were attenuated a little by the curtailing effects of the modulator. Since the Michelson contrasts for the full stimuli were not reduced in this way, they would have been higher relative to those of the cheese components, and summation estimates might have been inflated. For our stimuli, this attenuation was only appreciable (>0.5 dB) for stimuli of ≤ 1 cycle/check. Since summation did not begin to decline until after 2 cycles/check (see Figure 4), it is unlikely that contrast attenuation has substantially affected our results. Furthermore, since the modeling used bitmaps of the experimental stimuli as input, any physical shortcomings in the stimuli were also represented in the model.

Another potential concern is that our stimuli were not strictly narrowband. The damping around the stimulus boundary introduced some spectral splatter, though this was only very minor. Perhaps of greater concern is the quad of sidebands introduced by the plaid modulator. Each of these had an amplitude of -12 dB (a factor of 0.25) RE the carrier and the Michelson contrast of the stimulus was given by the sum of the amplitudes of the four sidebands and the carrier. For most of our stimuli, the sidebands were fairly similar to the carrier (in spatial frequency and orientation) and fell well within the passband of our model filter and, presumably, human spatial filters. The exceptions to this were for the stimuli where the modulator frequency was fairly high compared to that of the carrier (the most extreme case is the stimulus in the bottom left of Figure 1b). For these stimuli, the stimulus energy was much more diffuse in the Fourier domain than it was for the full stimuli and could result in high levels of empirical summation for uninteresting reasons (see Appendix B). However, these stimuli (e.g., the leftmost points in Figure 4) were not the ones constraining our estimates of the range of contrast integration. All this goes to say that it is unlikely that our conclusions are compromised by the complicating effects of sidebands in our stimuli.

Another point about our stimuli is that the contrast modulation means that they are second-order stimuli. This raises the possibility that we were tapping into second-order mechanisms rather than first. We have addressed this several times before (Meese, 2010; Meese & Baker, 2011; Meese & Summers, 2007) but make the following

simple point here. In all of the stimuli in this study, we manipulated the contrast of the carrier—a first-order signal. The amplitude of the second-order component in the cheese stimuli covaried with this, but if it were to contribute to detection (i.e., improve sensitivity over that available from first-order mechanisms), this would improve sensitivity to our cheese stimuli relative to our full stimuli. This would mean we have *underestimated* the level of summation. Therefore, even if conventional second-order mechanisms were involved, this does not undermine our point that contrast integration is spatially more extensive than is often supposed. A similar defense can be made regarding any attempts to describe detection of our check stimuli in terms of contrast variance detection (Morgan, Chubb, & Solomon, 2008).

Why can our cascade of pooling stages not be replaced by a single stage?

Our model involves a three-stage hierarchy of spatial summation (Figure 7). There is spatial summation within local filter elements, followed by spatial summation within contrast integrators, followed by spatial summation by a Minkowski stage. So why can these three stages of summation not be replaced with just one or two stages? The answer involves the intermediate nonlinear stages (the fillings) in our triple-decker sandwich of linear summation. Although the filtering stage is linear (Meese, 2010) and the contrast integration is linear, a square-law transducer is placed between the two. The transducer is needed to help account for the shape of the area summation curve in experiments that vary the diameter of a grating (e.g., see Meese & Summers, 2007). It was also needed to account for the transition between short-range (within filter element) summation (6 dB) and long-range (contrast integration) summation (3 dB) found by Meese (2010). It also accounts for much of the dipper region found in contrast discrimination experiments (Legge & Foley, 1980) and the slope of the psychometric function, in conjunction with area and binocular summation (Meese & Summers, 2009). Nonlinearities (here, the transducer) are not commutative, and so this constrains both the linear “bread” and the nonlinear “filling” in the first deck of the sandwich.

A similar argument applies to the next deck. Although we conceptualized Minkowski summation in terms of nonlinear probability summation, it can be construed as a further nonlinear transducer, in this case on the output of each integrator, followed by linear summation across the integrators (see Meese & Baker, 2011 for further comment). The nonlinearity in the Minkowski pooling means that this stage cannot be bundled into the preceding stage. We can appreciate this by recognizing that very widespread integration with a high (Minkowski-type) exponent cannot account for the summation results (Figure 3) and that a single restricted integration aperture cannot account

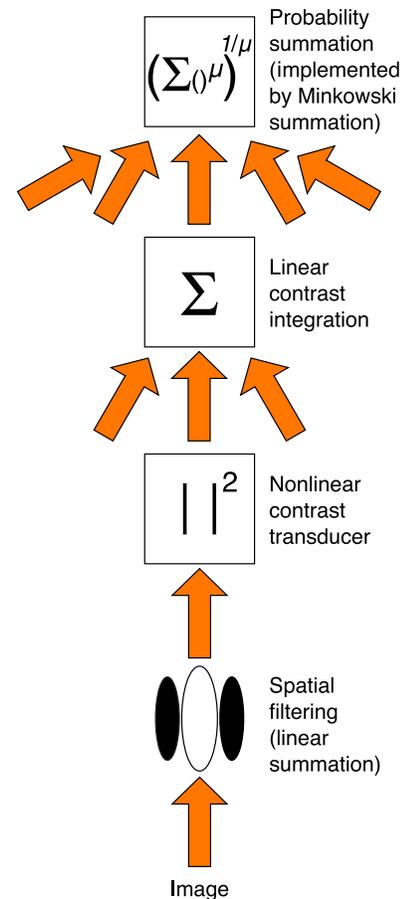


Figure 7. Block schematic of the three stages of spatial summation in the hierarchical pooling model. Each stage of summation (filtering, contrast integration, and Minkowski pooling) operates over progressively larger regions of the retina. In the implementation here, $\mu = 4$, consistent with probability summation.

for the detection thresholds (Figure 5b). Hence, strong summation (transducer exponent = 2) is needed within the aperture to account for the summation results and weaker summation (Minkowski exponent = 4) is needed across them to account for the threshold results.

Contrast integration is a zero-frequency second-order process

Following Henning, Hertz, and Broadbent (1975), second-order vision has enjoyed a well-established history (see Schofield, Rock, Sun, Jiang, & Georgeson, 2010, for recent work and a review). Leaving issues of signal rectification and other forms of transduction aside, most models construct second-order mechanisms by pooling across a spatial array of first-order mechanisms (linear filters) where the spatial frequency of the second-order modulation is derived by the pattern of weights at the pooling stage. The contrast integration that we are proposing here is a simple extension of this idea, where

the pooling weights are uniform across the array. Thus, our contrast integration mechanisms might be viewed as occupying the zero frequency (low-pass) channel of the second-order module.

We also note that our triple-decker sandwich (see previous section) is in a similar spirit to the filter–rectify–filter models that abound studies of second-order vision.

Why is contrast integration spatially limited?

Elsewhere, we have supposed that contrast integration is part of the network of neuronal convergence and contrast gain control that is needed to represent spatially extensive objects, surfaces, and textures (Meese, 2010; Meese & Baker, 2011). However, in principle, such stimuli can extend over the entire retina, so why should we find that the contrast integration process is limited in spatial extent, at least for DHB? One possibility is that more extensive contrast integration involves pooling at higher levels of the hierarchy. This might involve further inefficiencies, attenuators, and nonlinearities that mean the benefits of such pooling are not seen in psychophysical experiments at detection threshold.

Conclusions

This study confirms our previous finding that contrast integration over area is more substantial than widely believed. We have generalized the result across a wide range of carrier and modulator spatial frequencies and

have been able to describe the limit of contrast integration in terms of the number of cycles in the stimulus, which keys nicely with related results concerning retinal inhomogeneity. However, detailed analysis indicated that our data do not reject the possibility that the integration aperture is of fixed spatial extent (or, presumably, anything in between). More generally, our model includes spatial filtering, nonlinear contrast transduction, additive noise, contrast integration (linear summation within an integration aperture), and Minkowski summation (probability summation) across multiple apertures. The computation of contrast energy (within each aperture) probably takes place over at least twelve carrier cycles in most cases. This provides good psychophysical evidence for the neuronal convergence that is needed to represent spatial patterns that extend beyond the footprint of a single receptive field in the primary visual cortex.

Appendix A

Some model details

Minkowski summation

Minkowski pooling has long been associated with various forms of summation for reasons detailed elsewhere (see Quick, 1974; Tyler & Chen, 2000). It is defined as

$$\text{resp} = \left(\sum_{i=1:n} A_i^\mu \right)^{1/\mu}, \quad (\text{A1})$$

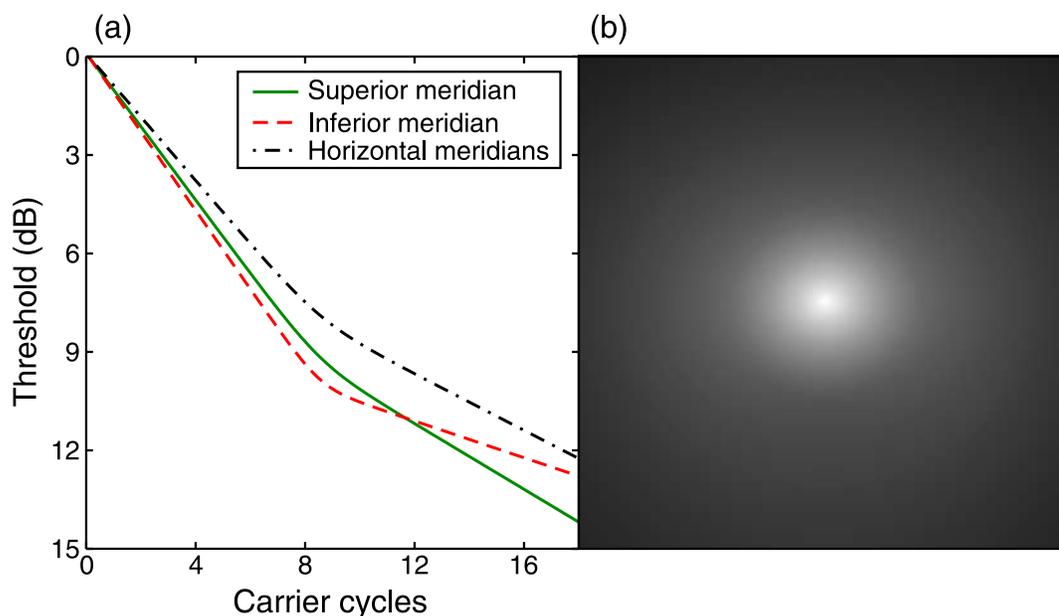


Figure A1. (a) Bilinear functions of the decline in retinal sensitivity with eccentricity. (b) Attenuation surface constructed by radial interpolation between the bilinear functions in (a).

where n is the number of units (filter responses, apertures, or any other appropriate scalar quantities) to be summed, and A_i denotes the response for the i th unit. In the typical formulation for probability summation with linear signal transduction, $\mu \approx 4$. If square-law signal transduction is assumed, then $\mu \approx 8$ (for our purposes here).

Retinal inhomogeneity

Previous work in this series (Meese, 2010; Meese & Summers, 2007, 2009) has assumed a linear decline in log sensitivity with eccentricity. The necessary model parameters were taken from Pointer and Hess (1989), who measured sensitivity across the entire visual field. However, our more recent work (Baldwin et al., 2010) has measured sensitivity in greater detail for just the central visual field (~ 9 deg diameter), where much of our—and other—psychophysical work has concentrated. Based on this work, we modeled retinal inhomogeneity using bilinear functions (on log-linear axes) with parameters derived from four observers, including those who participated in the present study. Figure A1a shows these functions for the different meridians (the same function was used for the two horizontal meridians). From these, we constructed a two-dimensional attenuation surface (Figure A1b) to scale the contrast of our stimuli in the modeling. These data and analyses come from a more detailed study to be presented elsewhere.

Appendix B

Some comments on the sidebands in our stimuli

In general, the spatial frequencies and orientations of the sidebands of our stimuli were not exactly matched to the filter used in the model because that was exactly matched to the spatial frequencies and orientations of the carriers. In fact, this offers another way of thinking about the within filter-element summation effects shown in Figure 2, as we now describe. The Michelson contrast of a check stimulus derives from the sum of the amplitudes of the carrier and each of the four sidebands. Each sideband amplitude is 25% of the carrier amplitude. Therefore, the amplitude of the carrier in the full stimulus (which has no sidebands) is exactly twice that of the carrier in each cheese component stimulus. If the detecting mechanism is a filter that sees *only* the carrier and not the sidebands, then it would produce 6 dB (a factor 2) of summation in our experiments for the trivial reason that the relevant signal amplitude is twice as high in the full stimulus as it is for the cheese components. Toward the far left of Figure 2 (in the main body of the report), the spatial frequencies of the sidebands are half an octave

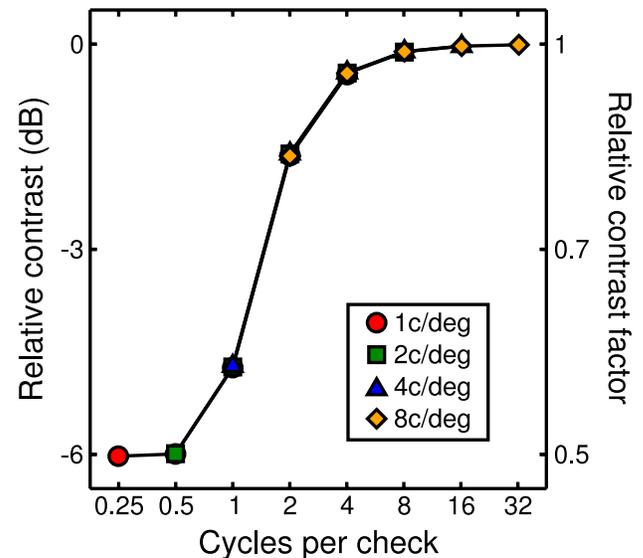


Figure B1. Attenuating effects of our standard filter on the Michelson contrast of our check stimuli. These are a consequence of filtering out the sidebands.

either side of the carrier and differ in the carrier orientation by at least 18° . This causes them to fall outside the passband of the narrowly tuned 8-lobe filter, and summation is ~ 6 dB for the trivial reason described above (gray dashed curve).

We considered this further in Figure B1, which shows the attenuation of Michelson contrast for our “white” cheese stimuli after being filtered by our standard model filter. Toward the left of the plot, the filtering causes a drop of 6 dB relative to the original stimulus. This is because the sidebands have been filtered out. By the time there are ~ 6 cycles/check (consistent with the upper range of 12 cycles of contrast integration for DHB), the filtering has a negligible effect on the Michelson contrast of the image. Therefore, it is unlikely that the complicating effects of the filtering and sidebands have misled our interpretation of the results.

Acknowledgments

This research was supported by a grant from the Engineering and Physical Sciences Research Council (EP/H000038/1). We thank Stuart Wallis for assisting with data collection. We also thank an anonymous reviewer and Keith May for encouraging us to perform the scale-dependent modeling.

Commercial relationships: none.

Corresponding author: Tim S. Meese.

Email: t.s.meese@aston.ac.uk.

Address: School of Life and Health Sciences, Aston University, Birmingham, B4 7ET, UK.

References

- Albrecht, D. G., & Geisler, W. S. (1991). Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Visual Neuroscience*, 7, 531–546. [PubMed]
- 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. [PubMed]
- Baker, D. H., Meese, T. S., Mansouri, B., & Hess, R. F. (2007). Binocular summation of contrast remains intact in strabismic amblyopia. *Investigative Ophthalmology and Visual Science*, 48, 5332–5338. [PubMed]
- Baldwin, A. S., Meese, T. S., & Baker, D. H. (2010). Loss of contrast sensitivity at 4 cycles/deg depends on eccentricity and meridian but not grating orientation for the central 9 deg of the visual field. *Perception*, 39, 1151.
- Banks, M. S., Geisler, W. S., & Bennett, P. J. (1987). The physical limits of grating visibility. *Vision Research*, 27, 1915–1924. [PubMed]
- Blakemore, C., & Campbell, F. W. (1969) On the existence of neurones in human visual system selectively sensitive to orientation and size of retinal images. *The Journal of Physiology*, 203, 237–260. [PubMed]
- Campbell, F. W., & Green, D. G. (1965). Monocular versus binocular acuity. *Nature*, 208, 191–192. [PubMed]
- Cannon, M. W., & Fullenkamp, S. C. (1991). Spatial interactions in apparent contrast: Inhibitory effects among grating patterns of different spatial frequencies, spatial positions and orientations. *Vision Research*, 31, 1985–1998. [PubMed]
- Finney, D. J. (1971). *Probit analysis*. Cambridge, UK: Cambridge University Press.
- Foley, J. M., Varadharajan, S., Koh, C. C., & Farias, M. C. Q. (2007). Detection of Gabor patterns of different sizes, shapes, phases and eccentricities. *Vision Research*, 47, 85–107. [PubMed]
- Graham, N. V. S. (1989). *Visual pattern analysers*. New York: Oxford University Press.
- Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9, 181–197. [PubMed]
- 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. [PubMed]
- Howell, E. R., & Hess, R. F. (1978). The functional area for summation to threshold for sinusoidal gratings. *Vision Research*, 18, 369–374. [PubMed]
- Legge, G. E. (1984). Binocular contrast summation—II. Quadratic summation. *Vision Research*, 24, 385–394. [PubMed]
- Legge, G. E., & Foley, J. M. (1980) Contrast masking in human vision. *Journal of the Optical Society of America*, 70, 1458–1471. [PubMed]
- Meese, T. S. (2010). Spatially extensive summation of contrast energy is revealed by contrast detection of micro-pattern textures. *Journal of Vision*, 10(8):14, 1–21, <http://www.journalofvision.org/content/10/8/14>, doi:10.1167/10.8.14. [PubMed] [Article]
- Meese, T. S., & Baker, D. H. (2011). Contrast summation across eyes and space is revealed along the entire dipper function by a “Swiss cheese” stimulus. *Journal of Vision*, 11(1):23, 1–23, <http://www.journalofvision.org/content/11/1/23>, doi:10.1167/11.1.23. [PubMed] [Article]
- 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. [PubMed]
- 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. [PubMed]
- Morgan, M., Chubb, C., & Solomon, J. A. (2008) A ‘dipper’ function for texture discrimination based on orientation variance. *Journal of Vision*, 8(11):9, 1–8, <http://www.journalofvision.org/content/8/11/9>, doi:10.1167/8.11.9. [PubMed] [Article]
- 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. [PubMed]
- Quick, R. F. (1974). A vector-magnitude model of contrast discrimination. *Kybernetik*, 16, 65–67. [PubMed]
- Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research*, 21, 409–418. [PubMed]
- Rovamo, J., Ukkonen, O., Thompson, C., & Näsänen, R. (1994). Spatial integration of compound gratings with various numbers of orientation components.

- Investigative Ophthalmology & Visual Science*, 35, 2611–2619. [PubMed]
- Schofield, A. J., Rock, P. B., Sun, P., Jiang, X., & Georgeson, M. A. (2010). What is second-order vision for? Discriminating illumination versus material changes. *Journal of Vision*, 10(9):2, 1–18, <http://www.journalofvision.org/content/10/9/2>, doi:10.1167/10.9.2. [PubMed] [Article]
- Syväjärvi, A., Näsänen, R., & Rovamo, J. (1999). Spatial integration of signal information in Gabor stimuli. *Ophthalmic and Physiological Optics*, 19, 242–252. [PubMed]
- 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. [PubMed]