i-PERCEPTION a Pion publication

i-Perception (2013) volume 4, pages 1–16

dx.doi.org/10.1068/i0556 ISSN 2041-6695 perceptionweb.com/i-perception

A common rule for integration and suppression of luminance contrast across eyes, space, time, and pattern

Tim S. Meese
School of Life and Health Sciences, Aston University, Aston Triangle, Birmingham B4 7ET, UK;
e-mail: t.s.meese@aston.ac.uk

Daniel H. Baker
School of Life and Health Sciences, Aston University, Aston Triangle, Birmingham B4 7ET, UK;
e-mail: d.h.baker1@aston.ac.uk

Received 5 August 2012, in revised form 30 November 2012; published online 2 January 2013.

Abstract. Visual perception begins by dissecting the retinal image into millions of small patches for local analyses by local receptive fields. However, image structures extend well beyond these receptive fields and so further processes must be involved in sewing the image fragments back together to derive representations of higher order (more global) structures. To investigate the integration process, we also need to understand the opposite process of suppression. To investigate both processes together, we measured triplets of dipper functions for targets and pedestals involving interdigitated stimulus pairs (*A*, *B*). Previous work has shown that summation and suppression operate over the full contrast range for the domains of ocularity and space. Here, we extend that work to include orientation and time domains. Temporal stimuli were 15-Hz counter-phase sine-wave gratings, where *A* and *B* were the positive and negative phases of the oscillation, respectively. For orientation, we used orthogonally oriented contrast patches (*A*, *B*) whose sum was an isotropic difference of Gaussians. Results from all four domains could be understood within a common framework in which summation operates separately within the numerator and denominator of a contrast gain control equation. This simple arrangement of summation and counter-suppression achieves integration of various stimulus attributes without distorting the underlying contrast code.

Keywords: vision, masking, dipper function, contrast integration, spatial, temporal, binocular, orientation.

1 Introduction

It is well established that higher order vision involves substantial neuronal convergence from the preliminary analyses by tiny receptive fields to more global analyses that are selective for either larger or more complex structures, or both. Detailed neuronal models of these processes are beginning to emerge (e.g. Serre, Wolf, Bileschi, Riesenhuber, and Poggio, 2007). Furthermore, there is growing psychophysical evidence for the various processes of integration including work from Morgan and Hotopf (1989), Field, Hayes, and Hess (1993), Moulden (1994), Wilson and Wilkinson (1998), Olzak and Thomas (1999), Levi and Klein (2000), Parkes, Lund, Angelucci, Solomon, and Morgan (2001), Jones, Anderson, and Murphy (2003), Motoyoshi and Nishida (2004), Dickinson and Badcock (2007), Sassi, Vancleef, Machilsen, Panis, and Wagemans (2010), and many others.

However, none of the studies above dealt with image contrast—the fundamental coding dimension of the primary visual cortex. Recent work in our laboratory has begun to address this at threshold in the spatial domain using spatially modulated carriers (e.g. "Battenbergs:" Meese, 2010; and "Swiss cheese:" Meese & Baker, 2011a; Meese & Summers, 2007, 2009), stimuli that were designed to encourage the observer to integrate over a fixed neuronal manifold while the experimenter varied the extent of the target. The aim of this approach was to clamp the level of internal noise so as to achieve a clean measure of the integration process. In conjunction with extensive modelling, these experiments provided good evidence for narrowband spatial filtering (Meese, 2010), followed by a square-law contrast transducer (Meese, 2010; Meese & Summers, 2009), additive noise and linear summation (integration) of image contrast (Meese, 2010; Meese & Baker, 2011a) that extended over nine or more stimulus cycles (Baker & Meese, 2011). Various models of spatial probability summation

(i.e. "signal selection") were always rejected in preference for the linear integration model described above (Meese, 2010, Baker & Meese, 2011; Meese & Summers, 2007, 2009, 2012). However, this raises a potential problem: if the internal response grows with the number of grating cycles—thereby improving sensitivity—there is a danger that perceived contrast will also vary with area. Because that would misrepresent the physical world, it is clearly an undesirable property of any vision system. Fortunately, this does not happen (much) in practice (Cannon & Fullenkamp, 1993; Meese, Hess, & Williams, 2005), but why?

Experiments that have measured contrast increment sensitivity at and above detection threshold (so-called "dipper functions") suggest that the problem outlined above is overcome using an elaborate network of contrast interactions (Meese & Baker, 2011a; Meese & Summers, 2007) involving summation and counter-suppression: the gain control *giveth with one hand and taketh away with the other* (Baker, Meese, & Georgeson, 2013). Put simply, the contrast code is normalized (Albrecht & Geisler, 1993; Heeger, 1992) by the population of filter elements from which the contrast signal has been integrated. This is not the zero-sum game it might seem. The idea is that although suppression is global and consistent across all of the integrators in the population, the extent of integration is not. Thus, some integrators will lose more from the global suppression than they gain from their own more restricted integration. This provides the potential for a population code along the dimension of interest—in this case, retinal image size (see Meese & Baker, 2011a, for details).

The general stimulus design that was central to the conclusions above was as follows. One stimulus (A) has elements that are interdigitated with another (B) along the dimension of interest, where contrast sensitivities to A and B are very similar. While observers are more sensitive to $A + B$ (hereafter AB) than A (or B) alone around detection threshold, sensitivity to A is the same as it is to AB above threshold (i.e. when the targets are placed on matched pedestals). It was once thought this meant that when shifting from threshold to suprathreshold, the visual system changed its mode of operation (Legge & Foley, 1980), but it is now clear that this pattern of behaviour can be understood within the single framework described above (Meese & Baker, 2011a; Meese & Summers, 2007). A critical comparison in reaching this conclusion was between the A on $\boldsymbol{AB}$ dipper (for clarity, we set the pedestal stimulus in bold) and the AB on $\boldsymbol{AB}$ dipper. This showed that performance was substantially better for the AB signal than the A signal alone *along the entire dipper function*. This indicates that area summation of contrast takes place at all contrasts, but that its effects can be obscured in the results derived from some experimental designs (e.g. A on $\boldsymbol{A}$ versus AB on $\boldsymbol{AB}$). In our model, the obscuration derives from the counter-suppression from the contrast gain control.

The phenomena above are not limited to the area/spatial dimension. Similar contrast interactions between the eyes have been found in binocular vision (Legge, 1984), producing ocularity invariance (Baker, Meese, & Georgeson, 2007)—the phenomenon that the contrast of the world looks very similar with two eyes as with one, in spite of binocular summation of signals across the eyes (Baker et al., 2007; Ding & Sperling, 2006; Meese & Baker, 2011a; Meese, Georgeson, & Baker, 2006; Moradi & Heeger, 2009). Here, we asked whether our findings can be extended beyond the spatial and ocular dimensions studied so far by investigating the orientation (pattern) domain and the temporal domain.

2 Methods

2.1 Equipment and viewing conditions

Stimuli were displayed on a Nokia Multigraph 445x monitor, which had a mean luminance of 60 cd/m^2. The monitor was gamma-corrected using standard techniques and ran at a frame rate of 120 Hz. Viewing distance was 119 cm, for which 48 monitor pixels subtended 1° of visual arc. We used a ViSaGe system (Cambridge Research Systems, Ltd., Kent, UK) to display the stimuli with pseudo-14-bit luminance resolution.

2.2 Observers

Each of the two main experiments was performed by three observers: ASB, DHB, and LP. The supplementary experiment reported in Appendix B was completed by ASB, DHB, and TSM. All observers performed the experiments wearing their normal optical correction where appropriate. DHB and TSM were authors. ASB and LP were unaware of the purpose of the experiment, but were psychophysically well practiced.

2.3 Stimuli

We first describe our stimulus construction in the orientation domain. Our *AB* stimulus was a difference of Gaussians (DoG) stimulus, constructed to have zero mean luminance (Figure 1a, left). Thus, the standard deviations of the positive and negative Gaussian functions were 0.21° and 0.31° respectively, and their amplitudes were unity and 0.44, respectively. The peak spatial frequency was 1 c/deg. We filtered our DoG stimulus with digital filters that had flat spatial frequency spectra, cosine phase spectra, and orientation spectra defined by a raised sine function of polar angle, with a period of two cycles over 360° (Figure 1a, middle). These filters were centred on orientations of ±45°, producing left oblique and right oblique (Figure 1a, far right) stimuli. Summing these two component stimuli (*A, B*) reproduced the original DoG stimulus. For these stimuli we produced a triplet of conditions as follows: (i) The target was *A* (or *B*) detected on a pedestal that was ***A*** (or ***B***) (Figure 1b, left)). (ii) The target was *AB* detected on a pedestal that was ***AB*** (Figure 1b, middle). (iii) The target was *A* (or *B*) detected on a pedestal that was ***AB*** (Figure 1b, right). Occasionally, we find it convenient to refer to *A* or *B* as a *single* stimulus (target or pedestal, as appropriate) and *AB* is a *dual* stimulus.

Stimulus duration was 100 ms and there was a 400-ms gap between the two intervals in our two-interval, forced-choice procedure. Contrast is expressed in dB where *contrast in dB* $= 20\log_{10}$(*contrast in %*) and *contrast in %* $= (L_{max} - L_{min})/(L_{max} + L_{min})$, where L is luminance. This is a convenient way of expressing log contrast, where 0 dB = a Michelson contrast of 1% and increments of 6 dB represent (approximately) a doubling of Michelson contrast.

Our temporal domain stimulus design followed the same principles as above. The spatial pattern was a sine-wave grating with a spatial frequency of 1 c/deg. It was modulated by a circular window

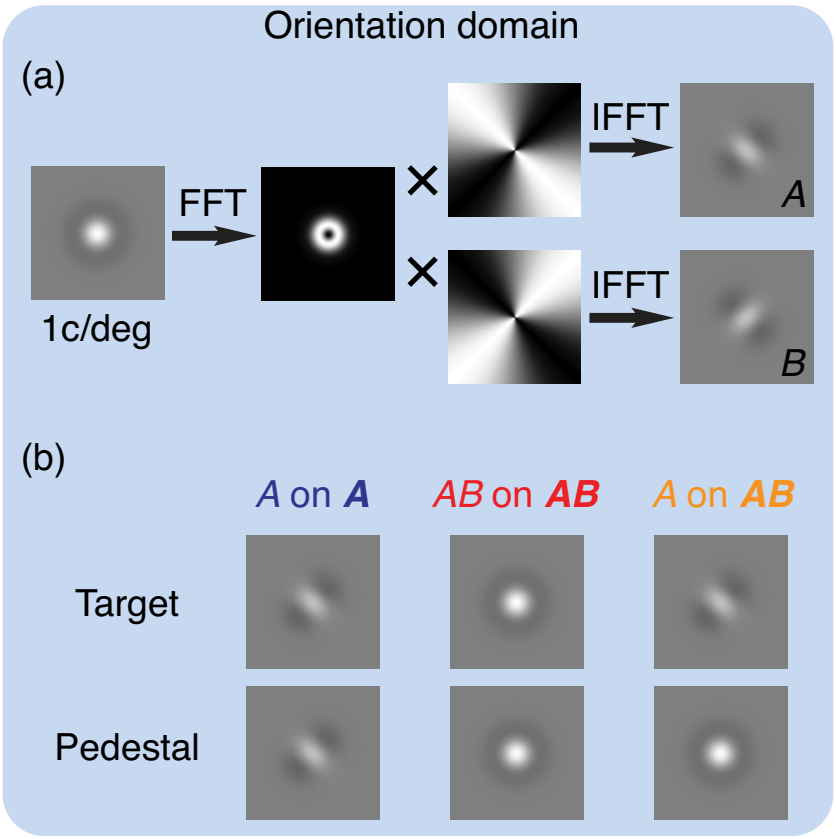

Figure 1. Stimulus construction and conditions for the orientation domain. (a) The non-oriented DoG stimulus (left) was the sum of the left and right oblique oriented components shown at the far right. The oriented components were constructed by filtering the Fourier transform of the DoG stimulus (second image from left) with digital filters, the spectra of which are shown in the third column of the panel. Panel (b) shows the triplet of stimulus conditions for the orientation domain. From left to right these are single on single, dual on dual, and single on dual. The targets and pedestals are shown spatially displaced for clarity; in the experiments they were spatially superimposed. In addition, there was a *B* on ***B*** condition and a *B* on ***AB*** condition. Results from these conditions were always averaged with *A* on ***A*** and *A* on ***AB*** conditions, respectively.

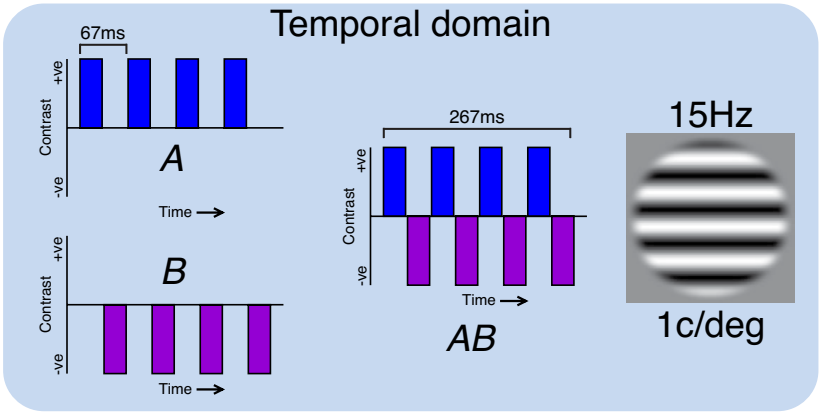

Figure 2. Stimulus construction in the temporal domain. The design principles are the same as those in the orientation domain (Figure 1).

with a central plateau width of 4 cycles and a blurred boundary with a width of 0.5 cycles (Figure 2, upper right inset). For the *AB* stimulus, the temporal modulation was four cycles of 15-Hz square-wave flicker. The *A* stimulus contained only the positive parts of the temporal modulation, whereas the *B* stimulus contained only the negative parts (i.e. they were positive and negative half-wave rectified temporal signals; see Figure 2). A triplet of conditions was constructed in exactly the same way as in the orientation domain.

2.4 Stimulus designs for comparison with previous studies

Our case will be strengthened by comparisons against two of the related studies mentioned in Section 1. For completeness, we summarize the relevant stimulus designs in those studies here. In the area/spatial domain Meese and Summers (2007) generated *A* and *B* components by modulating a sine-wave carrier with a raised plaid envelope. For the *A* stimulus (Figure 3, top), the plaid components were in cosine phase with the centre of the display. For the *B* stimulus (not shown), they were in anti-cosine phase. Thus, when added together (*AB*) this reproduced the original carrier grating (Figure 3, middle). The detection of *A* (or *B*) on *AB* produced the third (novel) condition within the triplet.

In the ocular domain, component stimuli (*A, B*) were monocular patches of sine-wave grating presented to either the left eye or the right eye. The compound stimulus (*AB*) was the same patch, presented to both eyes. The third and novel member of the triplet involved presenting the pedestal patch to both eyes and the target patch to just one (Meese et al., 2006).

2.5 Procedure

Both experiments were blocked by pedestal contrast, and subdivided into smaller sub-blocks by condition (*A* on ***A***, *AB* on ***AB***, or *A* on ***AB***). The order of contrast and conditions was randomized within each stimulus domain, and the conditions for each stimulus domain were repeated four times by each observer. For the orientation conditions, the single target was the left-oblique stimulus (*A*) for two

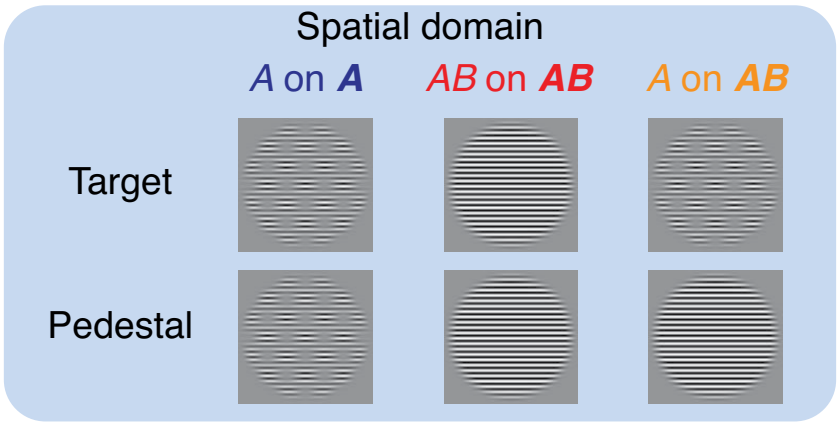

Figure 3. Triplet of stimulus conditions in the area/spatial domain studies by Meese and Summers (2007).

repetitions and the right-oblique stimulus (*B*) for the other two. With this arrangement, the observer was not uncertain about the target orientation. In the temporal domain, the flicker was sufficiently rapid to render the two target phases indistinguishable and so trials for *A* and *B* targets were interleaved.

Target contrasts were determined by 3-down-1-up staircases with a step size of 3 dB, in a two-interval forced choice (2IFC) design. For each pedestal contrast, a single experimental session consisted of a pair of randomly interleaved staircases. These either tracked the same stimulus configuration, or one tracked *A* on ***A*** and the other tracked *B* on ***B***, as described above. One 2IFC interval contained the combined pedestal and target, the other interval contained the pedestal only. Observers reported the interval they believed contained the target by pressing one of two buttons on a mouse. They received auditory feedback indicating the correctness of their response. We fitted the staircase data from each repetition with a cumulative log-Gaussian psychometric function (using Probit analysis) to estimate threshold at the 75% correct point. We then averaged across repetitions to give a mean threshold for each observer, and averaged across observers to produce the grand means shown in Figure 4.

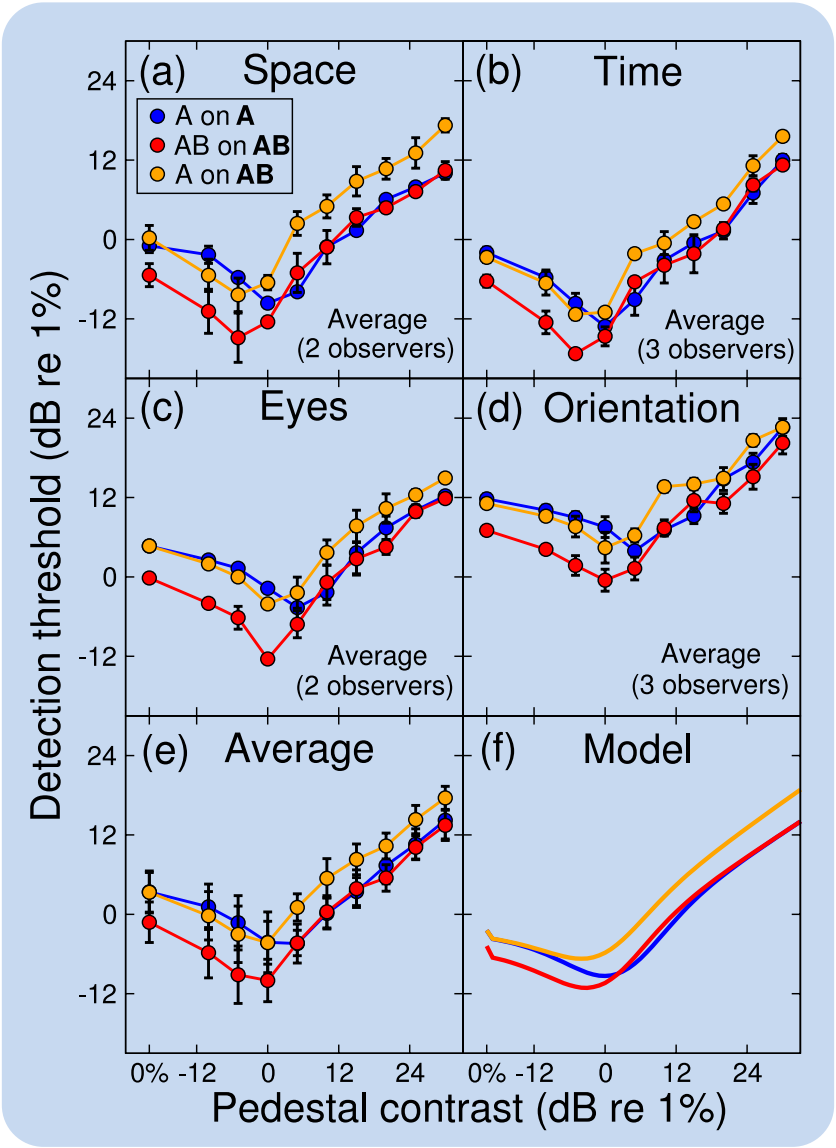

Figure 4. Results replotted from previous studies in (a) the area/spatial domain (Meese & Summers, 2007) and (c) the ocular domain (Meese et al., 2006) and those from the (b) temporal and (d) orientation domains studied here. Results are averaged across two observers (panels a and c) and three observers (panels b and d). The average of the four domains is shown in panel (e) and the predictions of the generic model of contrast integration (with the four free parameters set to the default values described in the text) are shown in panel (f).

3 Results

Preliminary analysis confirmed that sensitivities to *A* and *B* targets were very similar and results from these conditions were averaged in each stimulus domain (for simplicity, we refer to the average as *A* or "single"'). Figure 4 shows dipper functions for the temporal domain (panel b) and orientation domain (panel d) along with those from previous studies of the area/spatial domain (panel a) and ocular domain (panel c). The pattern of results in all cases is identical. Sensitivity to the single increment (*A*; blue) was much less than for the dual increment (*AB*) at detection threshold. When the experiment was extended above threshold, the two dipper functions converged, indicating that the benefit of the extra signal in the dual increment was lost. However, the results from the single increment (*A*) on dual pedestal (***AB***) (orange) show that the benefit is in fact preserved along the entire dipper function—it only appears to be lost in the *A* on ***A*** versus *AB* on ***AB*** comparisons because the signal benefit is offset by counter-suppression from the extra pedestal component (Meese et al., 2006).

To formalize the relationships in our data, we used a simple functional model (derived from Meese & Summers, 2007; see also Baker et al., 2013) where the overall contrast response ("*resp*") was given by

$$resp = \frac{A^p + B^p}{z + A^q + B^q} \quad , \tag{1}$$

where *A* and *B* are the Michelson contrasts of the *A* and *B* components, respectively, the exponents *p* and *q* were set to standard values of 2.4 and 2, respectively (from Legge & Foley, 1980), and the constant *z* was set to 2 for the purpose of illustration (e.g. Figure 4f). We assumed that the observer compared the response of this model equation across the two stimulus intervals and chose the interval that produced the greater response. The contrast detection threshold (or discrimination threshold, when the pedestal contrast was above 0%) was derived by solving for the lowest target contrast (i.e. the contrast increment to *A*, *B*, or *AB*) that produced a response-difference across the two intervals that equalled a criterion level of response *k*. In the model here, this was set to 0.2 for the purposes of illustration (e.g. Figure 4f). The values of *z* and *k* influence the location of the dip and overall sensitivity (see Appendix A for further details). The parameters *p* and *q* control the depth of the dip and the angle of the dipper handle. There were no other free parameters to adjust and no parameters controlled the quantitative relation between the triplet of dipper functions. The general success of the model is shown by a qualitative comparison between the average of the results from the four domains in Figure 4(e) and the model predictions in Figure 4(f) (see also Appendix A).

To highlight the signal integration properties of the visual system, we plotted the difference in the thresholds for the single (*A*) on dual (***AB***) stimuli (orange symbols in Figure 4) and the dual (*AB*) on dual (***AB***) stimuli (red symbols in Figure 4). The empirical results were handled in the same way as the model. Note that for the dB axes used here, a difference of 6 dB represents a contrast summation factor of two.

The model predictions (pink dashed curves) and the relevant data are shown for each observer (coloured symbols) and their average (black curves) in Figure 5. Although there is variability amongst observers, the model does a fairly good job in predicting the high levels of average summation (black curves) that were found across the entire contrast range (we will consider differences across stimulus conditions and observers in Section 4). Because of the general success of this model (Equation (1)) across the four stimulus domains (Figure 5) here and elsewhere (Baker et al., 2013), we refer to this as the *generic contrast integration model*.

4 Discussion

We have used the triplet of stimulus conditions introduced by Meese et al. (2006) and Meese and Summers (2007) to investigate signal integration in the temporal and orientation domains. We found that the pattern of results in each of these domains was the same as that found previously in the ocular domain (Meese et al., 2006) and the area/spatial domain (Meese & Summers, 2007).

Detailed inspection of Figure 5 reveals slightly more summation in the area/spatial domain than in the other three (i.e. there is a greater tendency for the individual results and their mean [black curve] to sit above the model prediction [pink curve] in Figure 5a compared with Figures 5b–d). As elaborated in Appendix A, this is probably because of the overlap between the *A* and *B* stimuli in the (spatial) dimension of interest for this stimulus (see Figure 3). The simple functional model used here is insensitive to

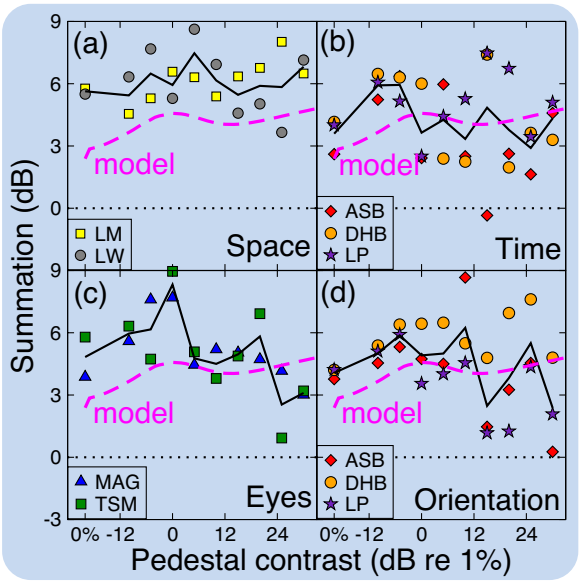

Figure 5. Summation results and model predictions for each of the four stimulus domains in Figure 4. Summation was derived by plotting the dB difference (equivalent to 20 times the log of the ratio) of the results from the *A* (or *B*) on ***AB*** condition and the *AB* on ***AB*** condition.

this aspect of the stimulus since its *A* and *B* terms are independent. However, the filter-based model of Meese and Summers (2007) was sensitive to this nuance and successfully accommodated the extra level of summation as well as predicting the correct qualitative pattern of results for the triplet. However, the stimuli used in the orientation domain also involved overlap along the domain of interest, so why did we not see the extra summation in that condition? It is difficult to be sure, but quite possibly, LP and ASB adopted slightly different strategies from DHB. As outlined in Appendix A, our model assumes that the observer is unable to select the output from the A component in the *A* on ***AB*** condition, but performs blanket integration over the entire stimulus (as implied by Equation (1)). Meese and Summers (2007) found support for this in the spatial domain from an identification experiment. At contrast detection threshold (both with and without an ***AB*** pedestal), observers were unable to identify whether the target was *A* or *AB,* suggesting that they were unable to achieve the benefit available by restricting signal integration to the target region (see Figure A1b in Appendix A). Blanket integration over the stimulus region results in a theoretical phenomenon that we have referred to elsewhere as *dilution masking* (Meese & Baker, 2011a; Meese & Summers, 2007; Baker et al., 2013; see also Appendix A). This arises from excitatory integration of inappropriate pedestal contrast and enhances the predicted level of summation in the model (for the *A* on ***AB*** versus *AB* on ***AB*** comparison). Dilution masking is distinct from more well-known forms of masking such as within-channel masking (Legge & Foley, 1980) and cross-channel masking (Foley, 1994; see Baker et al., 2013, for further comment). However, if LP and ASB were able to avoid this inappropriate integration by isolating the excitatory *A* response in the time and orientation domains (for some trials or pedestal contrasts at least), then this would diminish the level of summation, consistent with their results (see Figure 5).

The model used here is a simplification of the more elaborate schemes used elsewhere that were developed to account for quantitative details of other results (e.g. Meese & Baker, 2011a; Meese et al., 2006; Meese & Summers, 2007). However, all embody the same central concept: signal integration is offset by counter-suppression from the contrast gain control. The benefits of this general arrangement were discussed at length by Meese and Baker (2011a) (see also Baker et al., 2013); here we limit ourselves to a brief discussion of why the visual system might want to perform signal integration in each of the four domains that we have studied.

Objects, surfaces, and textures extend well beyond the spatial footprint of individual receptive fields that are typical in the primary visual cortex. Therefore, their explicit representation must involve integration through neural convergence, consistent with our model and the results from Meese and Summers (2007) in the spatial domain (Figure 5a). A similar argument applies to the orientation domain (Figure 5d): objects and textures contain multiple orientations and these need to be sewn together, not only along contours (e.g. the *association field* of Field et al., 1993) and within textures

(Motoyoshi & Nishida, 2004), but also across orientation at a single location, to represent conjunctions and orientation broadband pattern features such as blobs and plaids (Georgeson, 1992; Georgeson & Meese, 1997; Meese & Georgeson, 1996; Olzak & Thomas, 1999; Peirce, 2007; Peirce & Taylor, 2006). We speculate that a similar process also occurs for spatial frequency. However, the variation in sensitivity to different spatial frequencies (illustrated by the contrast sensitivity function) has (so far) made it difficult for us to derive a triplet of spatial frequency conditions that is suitable for testing this hypothesis.

Generally, during normal viewing, although the images in the two eyes are not identical (owing to the effects of depth and vantage point) they are very similar, and binocular combination is needed to achieve single vision without disturbing perception of contrast when one eye is closed (Baker et al., 2007; Ding & Sperling, 2006). This ocularity invariance is a property of our model and consistent with the experimental results of Meese et al. (2006) (Figure 5c) and Baker et al. (2007).

Why the visual system applies the same process in the temporal domain is less clear, but detailed temporal resolution is known to be lost for pulses of the time course studied here (Hess & Maehara, 2011) so the loss of resolution implied by the temporal integration is not so surprising. We suggest that the temporal integration that we report (Figure 5b) might be involved in the process of tracking temporal continuity, though more work is clearly needed to address this.

4.1 An alternative explanation of the triplet of pedestal masking

Another way in which signal integration can be achieved is through summation within a single filter element that is sufficiently broadly tuned to be sensitive to both the *A* and *B* components. Such an arrangement predicts a triplet of pedestal masking functions, similar to those found here (see Figure A1d in Appendix A). However, there is little or no evidence for such broad tuning at this low level in the hierarchy in the spatial domain (Baker & Meese, 2011a) or in the orientation domain (see Meese & Baker, 2011b, for a review), where receptive fields in the primary visual cortex (filter elements) have a small number of lobes (i.e. are selective to very few stimulus cycles) and are orientation tuned (DeValois & DeValois, 1990). Similarly, although binocularity is a distinct feature of visual cortex, the initial stages up to the primary visual cortex are known to be monocular (DeValois & DeValois, 1990). A counterargument in the temporal domain is less direct, because the temporal impulse response function is known to have a (small) negative lobe (Georgeson, 1987; Watson & Nachmias, 1977) and might be expected to achieve (some of) the temporal integration reported here (i.e. the sign of the contrast reversed *B* component would be effectively re-reversed by the negative part of the impulse response, allowing it to be summed constructively with the *A* component within the filter). One way to test this is to measure masking functions for an *A* on ***B*** configuration (e.g. in the ocular domain this is known as dichoptic masking). If the *A* and *B* components fall within a single filter element, then the masking function should look very similar to an *A* on ***A*** dipper function. However, Baker et al. (2013) found that this was not the case. In all four stimulus domains, *A* on ***B*** masking was much more severe than *A* on ***A*** masking, just like the well-known result for dichoptic masking (Baker & Meese, 2007; Legge, 1979; Meese et al., 2006). Unusual non-monotonic aspects of the empirical psychometric functions were also found for *A* on ***B*** masking and shown to be inconsistent with a within-filter-element account, but were predicted by the generic contrast integration model considered here, and there (i.e. Equation (1)). There is a problem, however. The configuration of the temporal stimulus used by Baker et al. (2013) was different from that used here. In Baker et al. (2013), the *A* and *B* components occupied interdigitated time slots, as they do here, but were not subject to contrast reversal (see Figure 2). Therefore, we also measured contrast masking functions for *A* on ***A*** and *A* on ***B*** configurations of the temporal configuration used here (see Appendix B). We found that *A* on ***B*** masking was stronger than *A* on ***A*** masking, confirming our earlier interpretation involving Equation (1).

5 Conclusions

To achieve higher order representations of objects, surfaces, and patterns, the visual system must integrate over the basic image properties that are analysed in the primary visual cortex (e.g. orientation and position). It must also combine information across the eyes and keep track of image detail over time. We propose that all of these can be achieved without distorting the underlying contrast code by a simple equation involving contrast summation and counter-suppression from contrast gain control. We refer to this model as the generic contrast integration model.

Acknowledgments. This work was supported by grants from the Engineering and Physical Sciences Research Council (UK, EP/H000038/1) and the Biotechnology and Biological Sciences Research Council (UK, BB/H00159X/1).

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. doi:10.1017/S0952523800010336

Baker, D. H., & Meese, T. S. (2007). Binocular interactions: Dichoptic masking is not a single process. *Vision Research, 47,* 3096–3107. doi:10.1016/j.visres.2007.08.013

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. doi:10.1167/11.14.14

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. doi:10.1163/156856807781503622

Baker, D. H., Meese, T. S., & Georgeson, M. A. (2013). Paradoxical psychometric functions ("swan functions") are explained by dilution masking in four stimulus dimensions. *i-Perception, 4,* 17–35. doi:10.1068/i0552

Cannon, M. W., & Fullenkamp, S. C. (1993). Spatial interactions in apparent contrast: Individual differences in enhancement and suppression effects. *Vision Research, 33,* 1685–1695. doi:10.1016/0042-6989(93)90034-T

DeValois, R. L., & DeValois, K. K. (1990). *Spatial vision*. New York, Oxford University Press.

Dickinson, J. E., & Badcock, D. R. (2007). Selectivity for coherence in polar orientation in human form vision. *Vision Research, 47,* 3078–3087. doi:10.1016/j.visres.2007.08.016

Ding, J., & Sperling, G. (2006). A gain-control theory of binocular combination. *Proceedings of the National Academy of Sciences USA, 103,* 1141–1146. doi:10.1073/pnas.0509629103

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. doi:10.1016/0042-6989(93)90156-Q

Foley, J. M. (1994). Human luminance pattern-vision mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America* A, *11,* 1710–1719. doi:10.1364/JOSAA.11.001710

Georgeson, M. A. (1987). Temporal properties of spatial contrast vision. *Vision Research, 27,* 765–780. doi:10.1016/0042-6989(87)90074-5

Georgeson, M. A. (1992). Human vision combines oriented filters to compute edges. *Proceedings of the Royal Society* B, *249,* 235–245. doi:10.1098/rspb.1992.0110

Georgeson, M. A., & Meese, T. S. (1997). Perception of stationary plaids: The role of spatial filters in edge analysis. *Vision Research, 37,* 3255–3271. doi:10.1016/S0042-6989(97)00124-7

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9,* 181–197. doi:10.1017/S0952523800009640

Hess, R. F., & Maehara, G. (2011). Does cognitive perception have access to brief temporal events? *iPerception, 2,* 142–149. doi:10.1068/i0418

Jones, D. G., Anderson, N. D., & Murphy, K. M. (2003). Orientation discrimination in visual noise using global and local stimuli. *Vision Research, 43,* 1225–1235. doi:10.1016/S0042-6989(03)00095-6

Legge, G. E. (1979). Spatial frequency masking in human vision: Binocular interactions. *Journal of the Optical Society of America* A, *69,* 838–847. doi:10.1364/JOSA.69.000838

Legge, G. E. (1984). Binocular contrast summation II. Quadratic summation. *Vision Research, 24,* 385–394. doi:10.1016/0042-6989(84)90064-6

Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America* A, *70,* 1458–1471. doi:10.1364/JOSA.70.001458

Levi, D. M., & Klein, S. A. (2000). Seeing circles: What limits shape perception? *Vision Research, 40,* 2329–2339. doi:10.1016/S0042-6989(00)00092-4

Meese, T. S. (2004). Area summation and masking. *Journal of Vision, 4,* 930–943. doi:10.1167/4.10.8

Meese, T. S. (2010). Spatially extensive summation of contrast energy is revealed by contrast detection of micro-pattern textures. *Journal of Vision, 10,* 14. doi:10.1167/10.8.14

Meese, T. S., & Baker, D. H. (2011a). Contrast summation across eye and space is revealed along the entire dipper function by a "Swiss cheese" stimulus. *Journal of Vision, 11,* 23. doi:10.1167/11.1.23

Meese, T. S., & Baker, D. H. (2011b). A reevaluation of achromatic spatio-temporal vision: Nonoriented filters are monocular, they adapt, and can be used for decision making at high flicker speeds. *iPerception, 2,* 159–182. doi:10.1068/i0416

Meese, T. S., & Georgeson, M. A. (1996). The tilt aftereffect in plaids and gratings: Channel codes, local signs and "patchwise" transforms. *Vision Research, 36,* 1421–1437. doi:10.1016/0042-6989(95)00212-X

Meese, T. S., Georgeson, M. A., & Baker, D. H. (2006). Binocular contrast vision at and above threshold. *Journal of Vision, 6,* 1224–1243. doi:10.1167/6.11.7

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,* 928–947. doi:10.1167/5.11.2

Meese, T. S., & Summers, R. J. (2007). Area summation in human vision at and above detection threshold. *Proceedings of the Royal Society* B*, 274,* 2891–2900. doi:10.1098/rspb.2007.0957

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,* 7. doi:10.1167/9.4.7

Meese, T. S., & Summers, R. J. (2012). Theory and data for area summation of contrast with and without uncertainty: Evidence for a noisy energy model. *Journal of Vision,* (under review). doi:10.1167/12.11.9

Moradi, F., & Heeger, D. J. (2009). Inter-ocular contrast normalization in human visual cortex. *Journal of Vision, 9,* 13. doi:10.1167/9.3.13

Morgan, M. J., & Hotopf, W. H. (1989). Perceived diagonals in grids and lattices. *Vision Research, 29,* 1005–1015. doi:10.1016/0042-6989(89)90115-6

Motoyoshi, I., & Nishida, S. Y. (2004). Cross-orientation summation in texture segregation. *Vision Research, 44,* 2567–2576. doi:10.1016/j.visres.2004.05.024

Moulden, B. (1994). Collator units: Second-stage orientational filters. *Ciba Foundation Symposium, 184,* 170–184.

Olzak, L. A., & Thomas, J. P. (1999). Neural recoding in human pattern vision: Model and mechanisms. *Vision Research, 39,* 231–256. doi:10.1016/S0042-6989(98)00122-9

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4,* 739–744. doi:10.1038/89532

Peirce, J. W. (2007). The potential importance of saturating and supersaturating contrast response functions in visual cortex. *Journal of Vision, 7,* 13. doi:10.1167/7.6.13

Peirce, J. W., & Taylor, L. J. (2006). Selective mechanisms for complex visual patterns revealed by adaptation. *Neuroscience, 141,* 15–18. doi:10.1016/j.neuroscience.2006.04.039

Sassi, M., Vancleef, K., Machilsen, B., Panis, S., & Wagemans, J. (2010). Identification of everyday objects on the basis of Gaborized outline versions. *iPerception, 1,* 121–142. doi:10.1068/i0384

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29,* 411–426. doi:10.1109/TPAMI.2007.56

Stevenson, S. B., & Cormack, L. K. (2000). A contrast paradox in stereopsis, motion detection and vernier acuity. *Vision Research, 40,* 2881–2884. doi:10.1016/S0042-6989(00)00164-4

Watson, A. B., & Nachmias, J. (1977). Patterns of temporal interaction in the detection of gratings. *Vision Research, 17,* 893–902. doi:10.1016/S0042-6989(00)00164-4

Wilson, H. R., & Wilkinson, F. (1998). Detection of global structure in Glass patterns: Implications for form vision. *Vision Research, 38,* 2933–2947. doi:10.1016/S0042-6989(98)00109-6

Yu, C., & Levi, D. M. (2000). Surround modulation in human vision unmasked by masking experiments. *Nature Neuroscience, 3,* 724–728. doi:10.1038/76687

Appendix A: Model details and behaviours

The model here is of a simple basic construction that belies the complexity of its operation. A deeper understanding of the model's behaviour is best appreciated by working through some variations in model architectures and observer strategies. Following the approach in the main body of the report, we shall consider functional models with two input lines. These might represent inputs from different eyes, different spatial locations, different orientations, different time slots, or other potential dimensions. Some of the points made below have been made previously (Meese, 2004; Meese et al., 2005; Meese & Summers, 2007) but are repeated here (briefly) for completeness within this more comprehensive treatment. We consider the following four model arrangements.

A.1 Model A1: Full summation and full suppression

The first arrangement is derived from the main model in Meese and Summers (2007) and is the model used in the main body of this report (Equation (1)). For convenience, we repeat it here with fixed exponents:

$$resp = \frac{A^{2.4} + B^{2.4}}{z + A^2 + B^2}, \tag{A1}$$

where z is a saturation constant (here set to 2), and A and B refer to the contrasts (pedestal plus target, as appropriate) on the two input lines. Following the language of the main report, when the signal (target) is presented to one or both input lines it is a single or dual increment, respectively.

A.2 Model A2: Selective summation and full suppression

In this model, the observer is able to match the region of excitation to the target, whereas suppression cannot be switched out. Although we have argued that the observer cannot do this for Swiss cheese stimuli (Meese & Summers, 2007), when the target is a small central patch of grating on a much larger patch of pedestal, it seems that something akin to this does take place (Meese, 2004). It might have also taken place for some of our observers in some of the stimulus conditions from the experiments here. This strategy is achieved in the model by adding a "switching" weight to one of the contrast terms on the numerator, thus

$$resp = \frac{A^{2.4} + wB^{2.4}}{z + A^2 + B^2}. \tag{A2}$$

The weight w is set to unity and zero for the single and dual target increments, respectively. By symmetry, we need to consider only the situation in which the single increment is applied to region A. This model is closely related to the so-called "matched model" in Meese & Summers (2007) (see their Figure 3a).

Note that both models A1 and A2 implement lateral suppression by the spatial pooling of the denominator terms in Equations (A1) and (A2).

A.3 Model A3: Late summation and no suppressive pool

This model involves a Legge and Foley (1980) type of arrangement for each of the stimulus regions, and an output that is their sum. This assumes that the response at each region along the dimension of interest is given by a static sigmoidal nonlinearity, and that the overall response is derived by adding up the responses. The model is given by

$$resp = \left(\frac{A^{2.4}}{z + A^2}\right) + \left(\frac{B^{2.4}}{z + B^2}\right). \tag{A3}$$

A.4 Model A4: Early summation

In a fourth model, contrasts from the two input lines are combined before any nonlinearity. This implies a detector with sufficiently broad tuning to be excited by both components. The model is given by

$$resp = \frac{(A + B)^{2.4}}{z + (A + B)^2}. \tag{A4}$$

Table A1. Pedestal (ped) and target contrast assignments for the A and B contrast terms in each interval (test and null) for the four models (Equations (A1)–(A4)).

	Target increment	Pedestal	Interval	***A***	***B***
A on ***A***	Single	Single	Test	ped + target	0
A on ***A***	Single	Single	Null	ped	0
(AB) on (***AB***)	Dual	Dual	Test	ped + target	ped + target
(AB) on (***AB***)	Dual	Dual	Null	ped	ped
A on (***AB***)	Single	Dual	Test	ped + target	ped
A on (***AB***)	Single	Dual	Null	ped	ped

A.5 Contrast assignments

The configurations studied in the experiments here involved detecting (i) A (or B) on $\boldsymbol{A}$ (or $\boldsymbol{B}$), (ii) AB on $\boldsymbol{AB}$, and (iii) A (or B) on $\boldsymbol{AB}$. For brevity, we shall refer to these conditions as single–single, dual–dual, and single–dual, respectively. In the models, the pedestal and target contrasts were assigned as shown in Table A1. However, recall that for model A2, the effective role of w is to set the B contrast term on the numerator to zero when the target is a single increment (i.e. when it is limited to the A region).

A.6 Three models are similar at detection threshold

At detection threshold, the denominators of each of the model equations are dominated by the saturation constant z. This means that the first three models can be approximated by the same expression:

$$resp = \frac{A^{2.4} + B^{2.4}}{z}. \tag{A5}$$

The fourth model is slightly different and predicts rather more summation at threshold (and above) owing to the summation of A and B before exponentiation:

$$resp = \frac{(A + B)^{2.4}}{z}. \tag{A6}$$

A.7 Noise and the decision variable

We make the simplifying assumption that noise is late and additive in all cases. The effects of noise propagating through from earlier stages are considered below.

The decision variable is given by Equation (A6):

$$k = resp_{\text{target}} - resp_{\text{null}}, \tag{A7}$$

where k is proportional to the standard deviation of the noise. To provide convenient placement of the model curves on double-log contrast axes, we set $k = 0.2$.

The model equations (Equations (A1)–(A4) and (A7)) were solved numerically for target contrast to produce the curves in Figure A1.

A.8 Model behaviours

The behaviours of models A1–A4 are shown in Figure A1. Model A1 (Figure A1a) predicts the same form for the triplet of dipper functions as was found in each of the experiments plotted in Figure 4, confirming its success.

The effect of restricting excitatory integration to the target region (model A2) is shown in Figure A1(b). This model adjustment affects only the single–dual condition (dashed orange curve; the red and blue curves are the same in Figures A1a and A1b), where performance in the dipper handle (masking region) is now better than in the single–single condition (dotted blue curve). This implies a facilitatory effect from the surround in the single–dual configuration (cf. Yu & Levi, 2000), here through the counterintuitive process of suppression. Note also that the model performance for the dual–dual condition (solid red curve) is worse than for the single–dual condition (dashed orange curve). In other words, the addition of signal alone can actually depress performance in this model (see

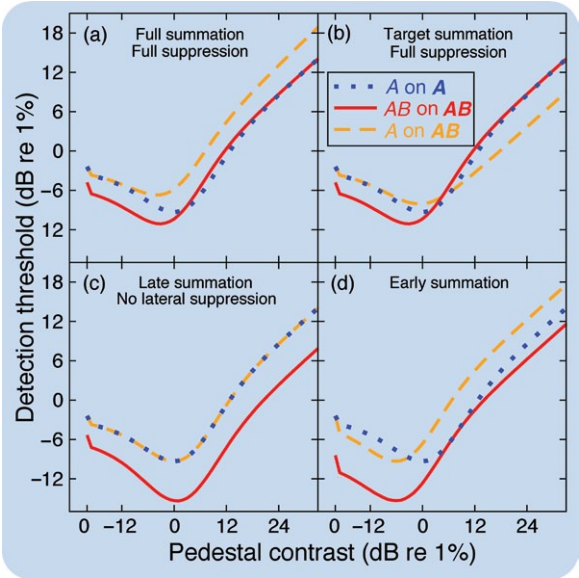

Figure A1. Model behaviours for three stimulus configurations (see Table A1) for each of the four models. (a) In model A1, integration is mandatory on both the numerator and the denominator. (b) In model A2, integration is matched to the signal on the numerator (excitation), but is mandatory on the denominator (suppression). (c) In model A3, there are no interactions at the gain control stage, but the outputs from the two stimulus regions are summed before the decision variable. (d) In model A4, contrasts are combined before the gain control. Note that the dotted (blue) curve is the same in all four panels because the contrast assignments for the single–single condition are the same in all versions of the model (see Table A1 and Equations (A1)–(A4)).

Meese, 2004 for further details). We shall return to this model configuration below when discussing the process of *dilution masking*.

Figure A1(c) shows the results for model A3. Similar to model A1, there is summation across the entire dipper function by comparison between the single–dual (dashed orange curve) and dual–dual (solid red curve) conditions. Thus, models A1 and A3 share similar characteristics in this respect. However, model A3 predicts that the single–single (dotted blue curve) and dual–dual (solid red curve) conditions should *not* converge. In contrast, these empirical functions do converge in each of the four domains plotted in Figure 4 (see also Legge & Foley, 1980 and Legge, 1984).

The predictions of model A4 are shown in Figure A1(d). These broadly resemble the predictions of model A1, except for in two critical ways. First, the level of summation is fixed at a factor of two (6 dB) along the entire dipper function, which is somewhat greater than was typically found (see Figure 5). Second, the dipper functions in the *A* on A and *AB* on ***AB*** conditions do not fully converge at high pedestal contrasts. This is not consistent with our empirical findings (Figure 4), for which the handles do converge. We discuss a third and more critical failing of this model in Appendix B and in Baker et al. (2012).

Thus, of the four models in Figure A1, model A1 provides the best description of the triplet of empirical functions in each of the four stimulus domains (see also Appendix B). In some cases, particularly in the area/spatial domain in Figure 5, the level of summation is slightly greater in the data than predicted by model A1. This owes partly to the fact that the stimulus components (*A, B*) were independent in model A1, but had spatial overlap in the experiment (see Meese & Summers, 2007, for further consideration). Narrowband filtering can also introduce or exacerbate this problem for the same reason: it introduces overlap in the manifold of responses to the *A* and *B* stimuli.

A.9 Lateral suppression

For the dual-pedestal conditions, the region of suppression was assumed to be the same for both single and dual target increments (Table A1). This is reasonable for blanket integration of the (excitatory) numerator (model A1). However, when excitatory integration is matched to the target (model A2), the spatial extent (weight) of suppression represents a potential degree of freedom. This is not explicit in Equation (A2), but could be implemented with a weight term, w_2, in front of the *B* term on the denominator. This is the version of the model that was found to be most appropriate in the studies of

Meese (2004) and Meese et al. (2005), where targets were always central circular patches of grating (of variable size). Thus, it seems that when the target does not have the same diameter as the pedestal and the visual system restricts the region of excitatory integration, there is additional suppression originating from the surround. In fact, we attribute the inconsistencies within these studies (e.g. variations across observers and variations in the level of suppression with the size of the target) to variability in this parameter (w_2). Alternatively, an equivalent arrangement might be constructed where the degree of freedom represents variable noise levels with different extent of integration. Nevertheless, this problematic degree of freedom is avoided in model A1. This did not come about by accident; we anticipated (Meese & Summers, 2007) that our triplet would promote a blanket integration strategy that would sidestep the problematic parameter (w_2) from the earlier studies (Meese, 2004; Meese et al., 2005).

A.10 Early noise

As mentioned earlier, we assumed constant variance and late additive noise (related to k) for convenience. This is reasonable when excitatory integration is over the entire stimulus (model A1) because multiple additive noise sources can be combined to produce an equivalent late noise source. However, in the other cases (models A2 and A3), there is arguably less noise for the single increment condition because the noise from the non-informative regions is switched out. This would have the effect of moving the single–dual (dashed orange) and single–single (dotted blue) curves downwards in Figures A1(b) and (c), and does not change our general conclusions.

A.11 Dilution masking

The different levels of performance for the single–dual (dashed orange curve) conditions across models A1 and A2 indicate that the mandatory excitatory integration (sum of numerator terms) of the non-informative pedestal region causes masking (the dashed orange curve is higher in Figure A1a than in Figure A1b). Meese and Summers (2007) referred to this process as *dilution masking*. It is a quantitative implementation of a process that appears closely related to the qualitative processes described by Stevenson and Cormack (2000) and Parkes et al. (2001). It is distinct from the cross-mechanism suppression of Heeger (1992) and Foley (1994), which does not involve excitatory integration. It is also distinct from the within-channel masking of Legge and Foley (1980), as it does not drive the signal response up the accelerating part of the transducer and therefore cannot produce pedestal facilitation (a dipper region).

Good evidence for dilution masking across eyes has been found by Baker and Meese (2007), who argued that it was responsible for a substantial part of dichoptic masking. They referred to this as the *indirect effect* of dichoptic masking. Further discussion of dilution masking and evidence for it in all the four domains considered here can be found in our companion paper (Baker et al., 2012).

Appendix B: *A* on *A* and *A* on *B* masking functions are different for our temporal stimuli

Baker et al. (2012) showed that the *A* on ***A*** masking functions were different from *A* on ***B*** masking functions for the spatial domain, the orientation domain, and the ocular domain. Here, we confirm that this is also the case for the temporal stimuli used here. Using the same methods as described in the main body of the report, we measured the two relevant masking functions for three observers (DHB, ASB, and TSM) at seven pedestal contrasts. The results for the three observers were averaged and are shown in Figure B1(a). The other two plots are model predictions for where the *A* and *B* components are initially processed independently, using Equation (1) from the main body of the report (Figure B1b), or for when the *A* and *B* components are first summed within the first-stage filter element (Figure B1c), using Equation (A4).

The results resemble the first prediction much more than the second, confirming that the various interactions shown in the main body of the report (Figure 4) do not derive merely from summation within first-stage filter elements (prior to nonlinearities).

Our preferred model (Figure B1b) predicts no region of facilitation in the *A* on ***B*** condition, yet there is some evidence for this in the experimental results, though less than for the *A* on ***A*** condition. Therefore, it is possible that some part of the interaction for the temporal condition derives from signal combination within the initial impulse response function.

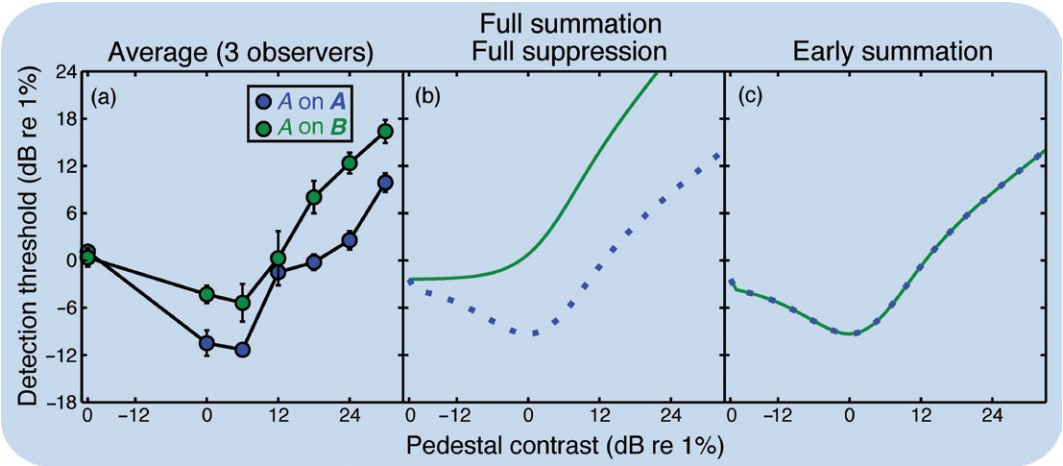

Figure B1. Results (a) and model predictions (b, c) for *A* on ***A*** and *A* on ***B*** masking. Experimental results are the means of three observers. Error bars show ±1 SE. See text for further details.

Tim S. Meese worked as a telecommunications engineer for five years before studying Psychology and Computer Science at the University of Newcastle-Upon-Tyne, where he graduated in 1989. He did his PhD at the University of Bristol and is now Professor of Vision Science at Aston University. His main research interests are in binocular and spatial vision and depth perception. He has been on the executive committee of the Applied Vision Association for more than 15 years and is now one of the chief editors of *Perception* and *i-Perception*. Here he is about to enter a lead mine.

Daniel H. Baker studied Psychology at the University of Nottingham from 2000 to 2003. He then worked in industry for a year before beginning a PhD at Aston University under the supervision of Tim Meese. Daniel held a postdoctoral position at the University of Southampton (2007–09), and is currently a research fellow at Aston University. His main research interests are spatial vision, binocular vision (including binocular rivalry), and motion perception. For more information, visit http://www1.aston.ac.uk/lhs/staff/az-index/daniel-baker/.

a Pion publication